University of Alberta

Internal Analyses of Asymmetric Competitive Market Structure Using Supermarket Aggregate Data

by

Fang Wu

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Marketing

Department of Marketing, Business Economics and Law
School of Business

Fall 2011
Edmonton, Alberta

Examining Committee

Paul R. Messinger, Department of Marketing, Business Economics and Law

Terry Elrod, Department of Marketing, Business Economics and Law

Joffre Swait, Department of Marketing, Business Economics and Law

Yuanfang Lin, Department of Marketing, Business Economics and Law

Andrew Eckert, Department of Economics

Alan Montgomery, Tepper School of Business, Carnegie Mellon University

Dedication

To my parents,

my elder sister Lei Wu,

my twin sister Fen Wu and brother-in-law Feng Chen

ABSTRACT

This dissertation proposes two internal analyses of market structure that can be applied to widely available store-level brand sales and price data. The methodologies, based on either a sales response model (a reduced-form model) or a discrete choice formulation (a structural model), enable researchers to identify the latent asymmetric competitive structure within a pre-defined market.

Chapter 2 estimates a market map of competitive brand relationships that are assumed to jointly underlie cross-price elasticities, own-price elasticities, and brand-specific intercepts in a sales response (i.e., market demand) model. The methodology uses an adaptive Bayesian approach to stabilize the estimation of demand parameters by sharing information across different brands and different model components in a set of demand equations. Drawing upon recent psychometric research, I express the asymmetries present in cross-price elasticities as the difference between what I refer to as brand power parameters, and I identify relationships between a focal brand's power parameter, clout, vulnerability, own-price elasticity, and spatial density. I apply the model separately for two datasets that consist of weekly sales and prices for beer and soft drinks.

Chapter 3 proposes a utility-based structural model and shows how the combination of the aggregate scanner data and forced switching data can help estimate a market map directly from a utility-based formulation. The utility function specification accounts for both vertical and horizontal differentiation

across alternatives, and also incorporates consumer heterogeneity. The specification is modeled to underlie both market outcomes (e.g., unit sales) and forced switching behavior. Conceptually, the proposed model relates the concept of asymmetric competition with fundamental parameters present in the utility function.

The two models developed in this dissertation represent complementary ways of conducting internal analysis of market structure. Chapter 4 discusses the relative advantages and disadvantages of each for application in different contexts. Chapter 1 introduces the topic of market structure analysis and summarizes antecedent research streams that have addressed this topic.

ACKNOWLEDGEMENT

I would like to thank my supervisors, Dr. Paul Messinger and Dr. Terry Elrod, for their constant guidance, encouragement, and support during the course of my study in the PhD program. I would also like to thank Dr. Joffre Swait and Dr. Yuanfang Lin for their invaluable suggestions and encouragement. Besides, I sincerely appreciate Professor Andrew Eckert and Professor Alan Montgomery, who spent much time and effort providing helpful comments and suggestions for the revision of this dissertation. Finally, I would like to thank my parents, my sisters and brother-in-law, for their endless love and continuous support during my work in the PhD years.

Table of Contents

List of Tables

List of Figures

Figure	Title	Page

Chapter 1： General Introduction

1.1 The Contribution of the Dissertation

This dissertation describes two unified internal analyses of market structure for inferring spatial representation of asymmetric brand competition from aggregate data.

1.1.1. Definition of Internal Market Structure Analysis

In the academic literature, demand-side market structure analysis[1] (Day, Shocker & Srivastava, 1979) refers to the class of techniques developed so far to help researchers and marketers understand the nature of competition in a predefined market (Bucklin & Gupta 1999), through explaining the extent to which the marketing offerings (e.g., products, services) under consideration are substitutes (Elrod et al. 2002).[2] The problem of identifying market structure helps practitioners know which of their competitors are most affected by and which will have a greater influence on their strategic behavior (Ruzo et al. 2006). Therefore, market structure analysis is pertinent to a large number of marketing decisions (Day, Shocker & Srivastava 1979) and is particularly "critical to the formulation of a firm or a retailer's own competitive strategy and the success or otherwise of this strategy" (Baker 1985).

Market structure analyses are typically of two types: external and internal (Carroll 1972). External analysis assumes that consumers' choices and behaviors

---

[1] Economists understand market structure mostly from the supply side, as in the degree to which production is dominated by one or a handful of companies. Marketing researchers, on the other hand, emphasize the demand-side explanation of market structure.

[2] Elrod et al. (2002) also emphasize the importance of understanding the complementary relationship among marketing offerings in market structure analysis. However, complementary is not a research focus of this dissertation.

evidencing underlying perceptions of brand substitutability are driven by brand attributes, and the values of the brands on these attributes are known to researchers. Examples of this approach that are of particular interest for my purpose include Pinkse, Slade, and Brett (2002) and Pinkse and Slade (2004), who assume that the brand-related parameters (e.g., price elasticities) of a demand model are functions of a vector of distance measures calculated from certain pre-specified brands attributes. In contrast, with internal analysis of market structure, consumers' choices and brand substitutability behaviors are explained by latent dimensions, where the number of dimensions and the location of the brands on these dimensions are determined solely by the data. Typically, managerial judgment is needed to interpret these dimensions in terms of either observed or unobserved brand attributes. An example of this approach that is quite relevant to current dissertation is the choice map proposed by Elrod (1988a, 1988b). While each approach has its own advantages, internal analysis of market structure is well known for its capability of identifying some important dimensions that researchers or marketers do not anticipate. For a detailed review and comparison of external analysis and internal analysis of market structure, see Elrod (1991) and Elrod et al. (2002).

The current dissertation focuses on internal analysis of market structure. Such analyses yield market maps that spatially locate brands and subjects' preferences in multidimensional latent attribute space. In these maps, brands that are closer in distance to each other are generally assumed to be more similar and, therefore, exhibit higher extent of competition. The interpretation of subjects' preferences for each brand along the brand attributes, on the other hand, is determined by what spatial model is used, the two primary candidates being an ideal point formulation (Coombs 1964) and a vector version (Slates 1960; Tucker 1960). With the ideal point formulation, brands that are closer in distance to an ideal point in any direction within a market map are considered to be more preferable by subjects; while for the vector version, brands positioned farther out in the direction of a vector are more preferred by subjects. The underlying assumption of this latter formulation is "the more the better." Additionally, the ideal-point model is more general than the vector model since the latter is hierarchically nested within the former: as the ideal-point goes to infinity, the two formulations become equivalent.

1.1.2. Use of Aggregate Data

Due to the importance of understanding market structure analysis, the past two decades have witnessed numerous techniques using panel data or survey data to facilitate the understanding of the market structure (e.g., choice map, Elrod 1988a, 1988b; Chintagunta 1994, Elrod & Keane 1995, DeSarbo et al. 1998, Park & DeSarbo 2008).

Despite the rich information these disaggregate data contain and the insights these techniques can provide us, in recent years, more and more attention has been paid to aggregate data-based market structure analysis. Researchers are aware that although marketers may often have the need to understand competitive structure, the disaggregate data are generally expensive to collect, and are typically available only for a relatively limited number of product categories, store, etc (Chen & Yang 2007). Besides, these data could suffer from the problem of being "unrepresentative of population" (Bucklin and Gupta, 1999; Gupta and colleagues, 1996; Bodapati & Gupta, 2004) and may lack time continuity (Nayga 1992). On the other hand, the aggregate data are detailed, timely and accurately recorded; readily available; indicate what people actually do rather than what they might do under certain circumstances; collected continuously, rather than at discrete and perhaps infrequent intervals, and constitute a census of all transactions rather than a statistical sample, which reduces sampling error (Feenstra & Shapiro 2003). Hence, the aggregate data constitute a good and sometimes even better alternative to both consumer panels and survey data. This circumstance justifies interest in developing analytical tools focused on aggregate data in managerial decision-making (a detailed review regarding the published use of aggregate scanner data in various marketing areas can be found in Bucker & Gupta 1999). Current development in aggregate data-based market structure analysis, however, comes far from fully exploiting the potentialities of such data.

1.1.3. Prevailing Methods for Deriving Market Structure from Aggregate Data

When aggregate data are used, the market structure analysis can be generally built upon two models in the existing literature:

1. *Reduced-form models.* In these models, a system of aggregate demand equations is estimated, one for each product. Each equation specifies the demand for a product as a function of its own price, the prices of other products and other variables (e.g, Allenby 1989; Cooper 1988; Montgomery & Rossi 1999).

2. *Utility-based models.* In these models, the aggregate sales observed in the market arise from a population of heterogeneous consumers making discrete choices based on a utility maximization framework (e.g., Berry 1994; Berry, Levinsohn & Pake 1995; Zenor & Srivastava 1993; Kim 1995).

Based on the type of the model used, market competition is typically measured by either:

1. *elasticity patterns*, particularly cross-price elasticity patterns, in the sense that products that compete more with each other tend to have higher positive cross-price elasticity (e.g., Allenby 1989; Cooper 1988); or
2. *covariance matrices of brand preferences*, which have a factor structure (i.e., vector model) with the values of the brands along the factors being interpreted as brand locations in market maps. (Chintagunta 1999; Chintagunta, Dube & Singh 2002, 2003).

These two approaches, with a few exceptions (e.g., Kamakura & Srivastava 1986; Chintagunta, Dube & Singh (2002, 2003)), normally take a multi-stage analysis of market structure, in the sense that the demand function parameters are estimated in the first stage, and then post-analysis (using, for example, a multi-dimensional scaling procedure) of cross-price elasticities or brand preferences' covariance matrix are conducted in the second stage to arrive at a market map. Conceptually appealing as these approaches are, in practice, they can suffer from the "too many parameters" problem: a system of J product alternatives gives rise to J^2 elasticities in the first approach, and $J(J+1)/2$

covariances and variances in the second approach to estimate. When the number of alternatives, *J*, grows, both the elasticity matrix and the covariance matrix become increasingly difficult to identify. This estimation problem becomes severe when marketing mix variables across alternatives tend to co-vary, a common phenomenon, unfortunately, in an industry featuring a large number of products.

A quick review of relevant literature shows that few researchers actually focus their attention directly on analysis of market structure when aggregate data are used. A large amount of research effort so far, instead, has been devoted to estimating valid demand parameters with aggregate data. And I believe part of the reason for this focus is due to the above mentioned problem and the logic that without good estimation of demand function parameters, analysis of market structure using either cross-price elasticities or brand preferences will be biased or even impossible.

Furthermore, the lack of direct research focus on market structure analysis gives rise to certain unaddressed research questions that partly motivate the current dissertation. These questions can be categorized into two types: those related with what information should be used to derive market maps; and those regarding what spatial models should be adopted.

The first type of research questions includes two issues: (a) Why are market maps typically derived only from cross-price elasticities (CPEs)? Intuition tells us that own-price elasticities (OPEs) and even intercepts also carry important market structure information. For example, a brand with higher absolute OPE should face more substitutes and, therefore, be close to more product alternatives in a multidimensional attribute space. And (b) shouldn't market maps to be used to account for not only the covariance matrix of preferences, but also the mean preferences of consumers in the market? Although the importance of explaining the mean preferences of the population in terms of the common attributes has been explicitly addressed when household level panel data is used (Elrod & Keane 1995), studies using aggregate data have ignored it (Chintagunta 1999, 2001; Chintagunta, Dube & Singh 2002, 2003). The current dissertation takes a position that emphasizes the complete account of market structure information in a demand function. And there are at least two benefits for such an emphasis: (a) empirically, the information is more completely and efficiently used to generate market maps; and (b) conceptually, the model can be used to predict all (major) brand-related parameters, as well as demand or shares for new or repositioned brands once their positions on the common factors have been specified.

Based on the first type of questions, the second type of research questions involves at least three different issues: (a) what spatial model formulations should be used when all important brand-related parameters (intercepts and price elasticities) are utilized to derive a market map in a reduced-form demand function? (b) how can we understand and model the asymmetric substitution pattern? And (c) why are the covariance matrices of brand preferences typically assumed to have a vector structure? It is well-known that a vector model of preferences assumes "the more the better," and, therefore is more suitable for describing vertical differentiation.[3] While for some attributes involving individual "tastes" such as sweetness of temperature, which feature horizontal differentiation[4], an ideal-point formulation is more appropriate not only for purposes of capturing the functional form, but also from the viewpoint of interpretation (Kamakura & Srivastava 1986). Kamakura & Srivastava (1986) constitutes the only paper that I am aware of adopting an ideal-point formulation of preference using aggregate data. However, considering the fact that the two types of characteristics, horizontal and vertical characteristics, are embodied actually in most of the products in the market (Anderson, Palma & Thisse 1992, Chapter 8), it appears that a utility-based model that can embody both vector and ideal-point formulations would be more desirable.

1.1.4. This Dissertation's Contribution

As a point of departure from much of past research, this dissertation focuses attention on internal analysis of market structure with aggregate data, and tries to address all the above mentioned problems and research questions. The methods proposed in this dissertation help one to infer a market map directly from demand relationships, formulated in one of two ways: either as a reduced-form demand model or in terms of the underlying utility structure. The methods, therefore, are unified in nature, which is in direct contrast to the traditional multi-stage analysis. The benefits of this unified approach are at least twofold: (a) the approach enables more efficient usage of information (e.g., the avoidance of an

[3] Vertical differentiation occurs in a market where products can be ordered according to their objective quality from the highest to the lowest.
[4] When products are different according to features that can't be ordered in an objective way, a horizontal differentiation emerges in the market. (Piana 2003)

errors-in-variables problem[5]); and (b) estimation of demand function parameters is no longer a pre-requisite for conducting a market structure analysis. Furthermore, the market map generated from each of the proposed methods has two advantages: (a) it accounts for all (or most) important brand-related competition information in a demand function; and (b) it explains the asymmetric nature of competition.

The core of this dissertation involves two essays. The model developed in the first essay is built upon the traditional sales response model and is elasticity-based. Drawing upon research in economics, marketing and psychometrics, the proposed methodology facilitates reliable estimation of the demand function and at the same time derives a market structure map that (a) accounts for the competition information embedded not only in cross-price elasticities, but also in own-price elasticities and intercepts; (b) represents the competitive asymmetry; (c) reasonably integrates common marketing specifications (e.g., vector and dominant point formulations) to represent the various demand-model components (e.g., own-price elasticities, competitive asymmetry feature). Conceptually, I identify underlying relationships among some important marketing and psychometric concepts: brand power, vulnerability, clout, and spatial density. The approach to modeling proposed in this essay is closest to Pinkse, Slade, and Brett (2002) and Pinkse and Slade (2004). One key difference is that their models constitute external analysis, and mine is internal analysis of market structure.

[5] An Errors-in-variables problem occurs when including exogenous variables known to contain error. This problem is known to bring biased and inconsistent ordinary least square estimates (Theil 1971) , and may cause non-concavity of the likelihood function when the logit model is estimated with maximum likelihood method (McFadden 1974).

The second essay builds upon a utility-based structural model. By applying spatial models to preference parameters, I am able to derive market structure maps directly from market demand data. The method proposed in this essay is closely related to the choice map by Elrod (1988a, 1988b), to Kamakura & Srivastava (1986), and also to Chintagunta, Dube & Singh (2002, 2003). Nevertheless, my model differs from these models in the following dimension： (a) choice map uses panel data, while mine uses aggregate data, (b) Kamakura & Srivastava (1986) carry out an external analysis of preferences, while mine is internal analysis; (c) Chintagunta, Dube & Singh (2002, 2003) use only the covariance matrix to infer the market map, but mine takes more full account of competition information in the utility function (wherein market structure is assumed to underlie parameter means as well as covariances); and (d) all three models adopt either the vector version or the ideal-point version of brand preferences, while mine incorporates both spatial formulations in the same utility function, which enables me to capture not only vertical characteristics but also horizontal characteristics of a brand. Similar to the method proposed in first essay, the market map generated is used to represent both types of characteristics.

Compared with other research in the literature, I use a different approach to enable accurate estimation of heterogeneity parameters. Rather than rely on the consumer demographic information, I make use of survey data, which takes the form of forced switching data. I show how a utility formulation can be modeled in a coherent way to underlie both the market outcome (e.g., unit sales) and the forced switching behavior. Again, conceptually, the proposed model relates the asymmetric substitution pattern with key terms present in the utility function (i.e., those products attributes that mainly feature vertical differentiation).

1.2. Literature Review

In this section, I will provide a detailed literature review. As I have discussed in section 1.1.3, few researchers actually take a direct market structure focus when they estimate a demand function with aggregate data. however, I believe it is important to discuss their work in detail and emphasize that this research provides important conceptual and methodological insights to form the basis for the current dissertation.

To be consistent with the discussion of section 1.1.3, the following development will take two lines: one is based on a reduced-form model and another is utility-based.

1.2.1. The Reduced-form model

Marketing literature has witnessed a tradition of estimating a system of demand equations, where quantity demanded of a product can be modeled as decreasing in its own price and increasing in the prices of its rivals. A common specification for this is the constant elasticity form, while alternate specifications include linear and semi-log functional forms. The estimated cross-price elasticities in particular give insights into the competitive market structure through patterns of substitutability among the various brands. Higher positive cross-price elasticity implies stronger competition between two brands. Since cross-price elasticities alone do not directly characterize market structure, researchers in practice post-analyze these estimates using such tools as factor analysis (e.g., Cooper 1988) or multidimensional scaling (e.g., Allenby 1989).

Certain limitations, however, arise from this approach:

1. **Too many parameters.** This problem can be severe when a market contains many products (*i.e.*, n is large) because the number of cross elasticities grows in rough proportion to the square of the number of products in the market. However, in industries characterized by horizontal differentiation, the existence of large number of products is a common feature.

2. **Unreliable estimates or infeasibility of estimation.** When no restrictions are placed to the demand function, own- and cross-price elasticity estimates may be of incorrect signs or of unreasonable magnitudes; e.g., positive own-price elasticities or negative cross-price elasticities (Montgomery and Rossi 1999). When firms systematically adjust the prices of all products in a product line in parallel or "lock-step," unreliability may become extreme and estimation of demand function may even become infeasible due to severe multicollinearity. Often all prices are the same within product lines that vary mainly along a horizontal dimension, such as color, scent, or flavor (Draganska and Jain 2003). This problem arises most often when using sales data from a single store, but it can also arise when using sales data from a single chain.

   It is important to point out here that for analysis of sales quantities, as in a typical constant elasticity model, true cross-price elasticities can be negative even among substitutes due to category expansion effects (Russell et al. 2008). This possibility means that the existence of negative cross-price elasticities does not necessarily indicate unreliability of estimation.

3. **Multi-stage estimation of Market Structure.** Market structure analysis based on post analyzing cross-price elasticities is multi-stage, and it gives rise to several drawbacks. As in all multi-stage estimation approaches, inefficiency arises because information about estimation error is not passed from one stage to the next. In addition, this particular multi-stage approach does nothing to solve the common problem of extreme instability and even infeasibility of elasticity estimates at the first stage. Constraints imposed at stage one can also be inconsistent with assumptions made at stage two about market structure. Further, there is no theoretical consensus for how to convert a matrix of elasticities into a matrix of brand distances. Finally, such market structure analysis has been linked only to cross-price elasticities, and not to own-price elasticities or demand. As a result, multistage estimation approaches say nothing about the change of own-price elasticities or intercepts of existing brands arising from the introduction or deletion of a product from the market. Thus, while such a multi-stage approach to estimation of market structure has value in exploratory work, it is not a substitute for a unified model of market structure.

4. **Symmetric Market Structure.** Current practice in the marketing literature that tries to derive spatial representation of market structure from cross-price elasticities is still preliminary as how to capture the asymmetric competition between brands. This may lead some researchers to simply focus on symmetric competition, which, however, may not match the actual competitive interaction because the directional competitive effects between any pair of brands are not usually the same

The techniques that have been developed so far for estimating aggregate models are able to address some, but not all, of the above-mentioned challenges. In the following sections, I focus, in particular, on whether the past methods are (a) parsimonious, reliable, and feasible; (b) done in multiple stages; and (c) able to incorporate explicitly the asymmetric nature of competition embedded in cross elasticities.

Parsimony, Reliability and Feasibility of Estimation

As previously discussed, the challenges with these models include (a) is the explosion in the number of parameters that need to be estimated as the number of products increases and (b) the poor quality of estimates when data is at a low level of aggregation such as store level (Montgomery and Rossi 1999). These two problems are usually addressed together in past work, since the accomplishment of parsimony can stabilize the variability in the model and improve the reliability of the estimates. One useful approach to solving these problems is to impose structural constraints on the pattern of cross-price elasticities exactly. For example, Allenby (1989) imposes IIA-style (independence of irrelevant alternatives[6]) elasticity restrictions derived from a microeconomic model. This method is not used in my paper for two reasons. First, it requires a preliminary analysis of unconstrained cross-price elasticity estimates to identify hypothetical submarkets. When such estimates are not immediately obtainable, as in one of my applications, the method is not feasible. Second, my goal is to let the data drive the market structure pattern across brands as much as possible, so I avoid imposing any *a priori* structural constraints.

Another stream of literature, widely used when aggregate data is at the chain level, involves imposing restrictions through a stochastic prior framework, in which a Bayesian shrinkage technique is used to "borrow" information across the units of analysis. Blattberg and George (1991), for example, shrink own-price elasticities across stores or brands to single points (i.e., zero for instance). Montgomery (1997) shrinks both own- and cross-price elasticities across stores toward a regression line. Montgomery and Rossi (1999) use a differential shrinkage approach, and impose prior information of price elasticities based on restrictions imposed by additive utility models. Wedel and Zhang (2004), in a study of cross-category price effects, use a two-component representation to decompose the variability in price effects. The shrinkage technique is used to pool data across items and hierarchically account for heterogeneity across stores. All of the above analyses are done with data involving multiple brands and multiple stores or chains.

[6] The IIA property states that the relative probability of any two alternatives being chosen is unaltered by the presence or absence of any other alternative from a choice set.

Since the aggregate data I use in current dissertation comes from a single store, and the only information I have is the units sales and prices for each brand, I do not directly use these approaches. However, the methods proposed in this dissertation do utilize a form of Bayesian shrinkage, more specifically, the adaptive Bayesian shrinkage technique, to share information across brands, so as to stabilize the brand-specific estimates. As background, a shrinkage estimator, in statistics, is an estimator that improves on an estimate by combining it with other related information. Adaptive Bayesian shrinkage improves on a Bayesian estimate by combining it with information about other Bayesian estimates, by supposing that they share a common prior distribution, and the hyper-parameters of the common prior distribution are jointly determined with the family of related Bayesian estimates. Adaptive Bayesian shrinkage is in contrast with the use of separate informative priors for various parameters, which express specific, definite information about a variable. This is also in contrast with the use of weakly informative (uninformative) priors for various parameters, which do not include other information, and which result in minimal shrinkage (information sharing). Generally, the posterior distributions arising from adaptive Bayesian shrinkage models tend to move (or shrink) away from the maximum likelihood estimates towards their common mean. For detailed background regarding adaptive Bayesian shrinkage, see Efron & Morris (1975), Rossi, Allenby & McCulloch (2005), and Montgomery and Rossi (1999).

It is worth noting that it is also possible to achieve parsimony and reliability of estimation by imposing Russell's (1992) Latent Symmetric Elasticity

Structure pattern (LSES) on *market share* price elasticities. The LSES model includes Allenby's (1989) approach as a particular case. In particular, in the LSES model, the market share price elasticities are decomposed into two parts: a symmetric substitution index revealing the strength of competition between brand pairs, and a brand specific coefficient which reveals the overall impact of a brand on its competitors. Parsimony and reliability are then realized either by directly parameterizing the symmetric substitution index as the distances between brands (Gonzalez-Benito et al. 2009), or by estimating the symmetric substitution index using information from brand switching data (Bucklin, Russell & Srinivasan 1998; Russell, Petersen & Divakar 2008). One advantage of the LSES model is that it provides some theoretical guidance for how to derive competitive structure from a matrix of elasticities. It does not require any prior knowledge of the possible market structure pattern among the brands. I did try to apply the LSES model to the data in the first essay. However, the model fails to converge for both datasets, which seems to imply that the assumptions underlying the LSES model are too restrictive for my data. Besides, the LSES model, like other past research, only addresses the market structure information in cross-price elasticities, and, hence, suffers from the problems I have discussed in section 1.1.3.

Multi-stage Analysis of Elasticity Matrix

The numerous elements in the elasticity matrix make it difficult to conceptualize the nature of competition. Therefore, a useful elasticity-based market structure analysis should be capable of both reliably estimating the elasticities *and* inferring the market structure from the complex information present in the elasticity matrix. Existing models in the literature have achieved these two goals mainly in a multi-stage process. Besides, in comparison with the considerable effort employed in perfecting the estimation stage in the past, modeling attempts in the analysis stage have been relatively limited.

Kamakura & Russell (1989) and Cooper (1988) introduce two summary measures of brand competition for elasticity matrices, involving what they called competitive clout and vulnerability. These two concepts help researchers assess cross-brand effects in terms of either how a focal brand exerts influence on the other brands in a market or how a focal brand is influenced by the other brands in a market. Cooper (1988), utilizing these two concepts, proposes a three-mode factor analysis technique for structured exploration of elasticities. His approach, which presumes that elasticities can be feasibly estimated and that such estimates are reliable, involves a two-stage estimation process. Allenby (1989) does not rely on the reliability of the estimates to infer market structure. Actually, it is the identification and testing nature of his approach that helps lead to reliable estimates. However, again, this approach cannot address the feasibility of estimation problem, and involves a three-stage estimation process. Gonzalez-Benito et al. (2009), as discussed before, propose a one-stage estimation method, but this method cannot address the feasibility issue of estimation either and is applicable only when market share information is available.

It is noteworthy that none of the above work explores a market structure map that can account for not only cross-price elasticities, but also own-price elasticities and intercepts. Pinkse, Slade, and Brett (2002) and Pinkse and Slade

(2004) are among the few researchers that explicitly relate those brand-related parameters of a demand function to market structure information featured by some common product attributes. However, their model involves an external analysis of market structure, so the choice of attributes induces some level of subjectivity, and when important attributes are omitted (possibly due to data limitations), price elasticities derived from the model would be misleading.

Asymmetric Pattern in Elasticity Matrix

Competitive asymmetry is defined as the phenomenon that the degree to which brand A may influence brand B does not equal the degree to which brand B influences brand A (Desarbo, Grewal & Wind 2006). In terms of cross elasticities, this asymmetric nature of competition implies that the impact of brand *j*'s price change on brand *i*'s demand does not always equal the impact of brand *i*'s price change on brand *j*'s demand. Such inherently asymmetric competitive patterns in cross-price elasticities are well documented in the marketing literature (e.g., Desarbo, Grewal & Wind 2006, Blattberg & Wisnewski 1989). Despite the existence of much research trying to empirically describe such asymmetric patterns of competition in cross-price elasticities, a review of the extant literature reveals only a few modeling attempts to explicitly analyze the asymmetry present in cross-price elasticity matrices. Cooper (1988) provides an example of how the asymmetric elasticity matrix can be structured by three-way factor analysis. Russell's (1992) LSES model captures the competitive asymmetry by including a brand specific coefficient that reveals the overall impact of a brand on its competitors. In contrast with this literature, the current dissertation uses the asymmetric multidimensional scaling (MDS)[7] method developed in psychometrics to model and depict the asymmetric substitution pattern across brands. The benefits of this method are its simplicity and flexibility in terms of the model form and the visualization of the asymmetries. And the particular asymmetric MDS method I use provides a simple way to test the existence of asymmetric effects in the model.

1.2.2. The Utility-based model

In the market structure literature, one common approach to estimating aggregate data that is in direct contrast with the reduced form method is a structural approach, which is based on optimizing behavior of agents (e.g., utility maximizing by consumers) (Chintagunta, Erdem, Rossi & Wedel 2006). More specifically, when such an approach is adopted, a discrete choice model (e.g., a logit or probit model) is developed at the individual level, a distribution of consumer preferences over products is assumed, and these preferences are then aggregated into a market-level demand system (Berry, Levinsohn & Pakes 1995).

The main issue when analyzing aggregate data with this kind of approach concerns how to explicitly account for the heterogeneity (e.g., in preferences and price sensitivities) across consumers and endogeneity of marketing mix variables (i.e., the potential correlation of a brand's price with other marketing activities not incorporated in the model) in the utility function. Without controlling for these issues, estimated model parameters are known to be biased (Villas-Boas & Winer 1999; Allenby & Rossi 1999) and therefore, any derived parameters such as price

[7] The asymmetric multidimensional scaling is a method which is specifically designed to analyze asymmetric relationships among members and display them graphically by plotting each member in a certain multidimensional space (Saburi & Chino 2008).

elasticities are biased too. For example, when heterogeneity is not accounted for in a logit demand model, the IIA assumption applies, and the substitution pattern between products is driven completely by market shares and not by how similar the products are. In addition, strong economic implications arise that indicate that lower priced products tend to have lower (absolute) own-price elasticities, which implies higher markups in price. This implication, however, depends on the functional form of how price enters into the utility function (see Nevo 2000 for a detailed illustration of these problems).

Because of the importance of accounting for heterogeneity and endogeneity problems in such models, most of the literature focuses its interest on developing various estimation methods that could address either one or both of the problems with aggregate date. So far, the proposed methods can be categorized into three general groups: non-likelihood based estimation methods, likelihood-based methods and Bayesian estimation methods. Although my research focus is on internal analysis of preferences, correctly estimating the utility function is also of interest to my research purpose in this dissertation. Therefore, the remainder of this section will discuss the literature using these three categories.

Non-likelihood Based Estimation Method

The first literature relies on non-likelihood based estimation methods. In this approach, model parameters are estimated by either minimizing the discrepancy between observed market share and predicted market share or equating the two by introducing the unobserved product characteristics into the utility function. While the former estimation strategy will yield estimates of the parameters that determine consumers' heterogeneity, it does not account for endogeneity. In contrast, the latter strategy explicitly accounts for both issues. This latter approach has its roots in Berry (1994) and Berry, Levinsohn & Pakes (1995, henceforth will be referred to as BLP method), which has received wide application in both economics and marketing (e.g., Nevo 2000; Nevo 2001; Chintagunta 1999, 2001; Sudhir 2001).

In particular, the BLP method adds an unobserved brand characteristic in the traditional utility function, and allows the related parameters to be correlated with marketing mix variables such as prices, which leads to endogeneity problem. Since no distribution is assumed on the unobserved brand characteristics, these characteristics are actually estimated as fixed-effects. With aggregate data, directly estimating the fixed effects of unobserved product characteristics can be problematic due to the lack of degrees of freedom (i.e., for J brands and T time periods (or markets), researchers need to estimate J times T number of fixed effects). The BLP method (1995) then circumvents the direct estimation of fixed effects by using a numerical inversion method together with GMM.

Not imposing distribution assumptions for the unobserved product characteristics, constitutes the strength of the BLP method, since this implies that it does not impose restrictions on the forms of pricing behavior. Park & Gupta (2009) point out that the distribution imposed on unobserved product characteristics may have important implications for pricing behavior.

Two shortcomings, however, are related with the BLP method (interested readers can refer to Park & Gupta 2009 for a much more detailed illustration). First, the BLP method assumes no sampling errors in the observed market shares of alternatives (in statistics, the sampling error or estimation error is the error caused by observing a sample instead of the whole population). Randomness in shares is assumed to come only from unmeasured product characteristics. If the no-sampling-error assumption is invalid, the BLP is neither consistent nor asymptotically normal (Berry, Linton & Pakes 2004). This restricts the application of the BLP method to those applications where the underlying sample of consumers is very large. Second, the BLP method is unable to recover heterogeneity parameters precisely when only aggregate data are available (Albuquerque and Bronnenberg 2006, Petrin 2002). To overcome this problem, researchers typically combine aggregate data and consumer-level data (e.g., Goldberg 1995; Petrin 2002; Albuquerque and Bronnenberg 2005).

Two papers that are of particular interest to my purpose are Chintagunta, Dube and Singh (2002, 2003), who use mixed logit utility specifications with normally-distributed random coefficients to estimate heterogeneous demand systems amenable to internal analysis of market structure with aggregate data. These papers use particular specifications for the utility functions for the different brands in a market. The covariance of the various brand-specific intercepts is decomposed into a matrix of latent attributes for each of the brands in the market and a vector of latent consumer tastes for these brands. This covariance structure, assumed to hold for the population of households in the market, is then estimated. One of these papers also includes covariates associated with particular stores, such as neighborhood demographics, as underlying the mean brand-specific intercept and the mean marginal utility of income. These models differ from my proposed model in that the latent product attributes in their models are assumed to underlie only the covariance matrix of brand-specific intercepts (across brands); whereas in my model, the latent product attributes for different brands are also assumed to underlie the mean brand-specific intercepts. In addition, their model essentially adopts a version of a vector model of brand preferences (i.e., a linear latent structure underlying brand-specific intercepts), whereas while my model incorporates both vector and ideal-point spatial formulations underlying the brand-specific intercepts.

Likelihood Based Estimation Method

The second approach is to specify a likelihood function for the aggregate data. (e.g., Bodapati & Gupta 2004; Kim 1995; Zenor & Srivastava 1993). This approach assumes that purchase of every unit of each brand each week is made independently, and in aggregate, the data follows a multinomial process. Compared to the BLP method, this approach explicitly takes the sampling error into account. However, the performance of the aggregate model in recovering the parameters relies on the validity of the multinomial assumption.

Bodapati & Gupta (2004) actually provide a justification for this multinomial assumption (page 354): *"under the assumption that the sample of households that makes a trip to the store is small relative to the population of shoppers, the sampling can be treated as being with replacement. Therefore, the outcome of each trip is independent of the outcomes of other trips (conditional on* *model parameters), and the aggregate data are multinomial counts."* Kim (1995) has further pointed out three factors that may violate this assumption. The first factor is the correlation between the heterogeneity parameters and the unit purchased. For example, if price-sensitive consumers tend to buy more units, a demand model assuming multinomial assumption might overestimate the mean of price sensitivity in the population. Second, the correlation between the heterogeneity parameters and the purchase frequency also matters. For example, if a price sensitivity consumer tends to make purchase more frequently, again, the proposed demand model will overestimate the number of price-sensitive households existing in the population, which leads to biased estimation for price coefficients. The third factor is whether or not the households are exposed to different marketing mix. Furthermore, the multinomial assumption actually combines two sources of heterogeneity: within-household and between-household heterogeneity. However, for mature products, within-household heterogeneity might be limited (Kim 1995).

Further, when likelihood-based estimation strategy is adopted, two types of heterogeneity structure have been explored in literature: parametric and semi-parametric. Kim (1995) is an example who assumes continuous heterogeneity structure and he uses a Maximum Likelihood estimation method. Some other researchers apply a latent class model to the aggregate data (e.g., Besanko, Dubé, and Gupta 2003; Draganska and Jain 2002; Seetharaman 2001; Zenor and Srivastava 1993). For example, Zenor & Srivastava (1993) uncover the heterogeneity in aggregate data by proposing a latent segment logit approach, which implies a semi-parameteric structure of heterogeneity. The Expected Maximum Likelihood Method is used in the paper. Bodapati & Gupta (2004) provide theoretical guidance for identifying latent class heterogeneity parameters from aggregate level data. And they show that under specific assumptions and when the household-level model is correctly specified, most of a latent-class segmentation structure is recoverable with only store-level aggregate data, although sometimes unreasonably large sample sizes are required.

It is worth mentioning that Kamakura & Srivastava (1986) propose an interesting method that addresses the heterogeneity issue with aggregate data in quite a different way. In particular, they present a multinomial ideal point probit model and account for the heterogeneity by assuming a distribution of ideal points. Their paper is similar to current dissertation, although they focus on external analysis of preferences, while I focus on internal analysis of preferences. Besides, my method is more general in the sense that my utility function incorporates and combines features of both vector and ideal-point models.

Finally, there is research that addresses both endogeneity and heterogeneity issues, constituting a likelihood-based correspondence to the BLP method, as, for instance, Park & Gupta (2009). They propose a "simulated maximum likelihood" method to estimate an aggregate Random Coefficient Logit model that considers both endogeneity and heterogeneity. This can be seen as a generalization of the approach from Petrin & Train (2004) and Villas-Boas & Winer (1999).

Bayesian Analysis of Random Coefficient Logit (RCL) Model

It is worth mentioning that in recent years, several studies have performed

Bayesian analysis of Random Coefficient Logit (RCL) model using aggregate data. Some of them use a data augmentation idea in which the parameters of "pseudo" consumers are estimated (i.e., by imposing typically a joint prior on these parameters and estimating both these parameters and the hyperparameters as if it were the disaggregate data that we observe) together with the set of parameters used in inference. Chen & Yang (2007) propose such a model without considering the endogeneity problem. And they augment the choices of R representative consumers using a Metropolis-Hasting (MH) step. Musalem, Bradlow and Raju (2009) propose a Gibbs Sampling alternative to Yang & Chen's MH method, which enables them to draw augmented individual choices directly from their full-conditional posterior distribution. Another important contribution of Musalem et al. (2009) is that they consider two alternative scenarios that generate the observed aggregate data – one in which there are independent cross-sections of consumers in each period (assuming a multinomial distribution of aggregate data) and one in which there is a panel of consumers. And they show the methods used for the estimation of the system of independent consumers can be used to estimate the demand of consumers from the second system. This actually provides further justification for the multinomial assumption of aggregate data.

Jiang, Manchanda & Rossi (2009) proposed a Bayesian analysis of the aggregate RCL model based on distributional assumptions about the unmeasured product characteristics. So their paper considers both the endogeneity and heterogeneity issues. Similar to the BLP model, this approach assumes no sampling error, and the model estimation involves inverting shares through the

BLP contraction mapping (Park & Gupta 2009).

The method proposed in current dissertation does not use the non-likelihood based estimation method, since I do not have enough information to calculate each brand's market share. I instead use a likelihood-based method and Bayesian analysis to facilitate the estimation.

1.3. Overview of the Dissertation

This dissertation is organized as follows. Chapter 2 of the dissertation develops the reduced-form-based market structure analysis. Chapter 3 illustrates the utility-based analysis of market structure. In Chapter 4, I provide a brief conclusion, including a synthesis of the two modeling approaches, a discussion of the relative strengths and limitations of each, and a discussion of future needed research.

For both Chapter 2 and Chapter 3, I apply the proposed models to analyze cross-brand competition in two product categories, one is for beer products, and the other is for the soft-drink category. While only aggregate scanner data is available for the beer data, the soft-drink data come from two sources: store-level scanner data containing weekly sales and price information, and forced switching data describing consumers' willingness to substitute from one product to another when the former brand is assumed to be out of market.

Bibliography

Albuquerque, Paulo & Bart J. Bronnenberg (2009), "Estimating Demand Heterogeneity Using Aggregate Data: An Application to the Frozen Pizza Category," *Marketing Science*, 28 (2), 356-372.

Allenby, Greg M. (1989), "A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data," *Marketing Science*, 8(3), 265–280.

Allenby, G.M. & P.E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics* 89, 57-78.

Baker, M. J. 1985. Marketing Strategy and Management. New York: Macmillan.

Besanko D, Dube JP, Gupta S. (2003), "Competitive Price Discrimination Strategies in a Vertical Channel using Aggregate Data," *Management Science*, 49, 1121-1138.

Berry, Steven T., (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (2), 242-262.

_____, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometric*, 63 (4).

______ , _____ , ______ (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: the New Vehicle Market," *Journal of Political Economy*, 112 (1), 68-104.

Blattberg, Robert C. and George, Edward I. (1991), "Shrinkage Estimation of Price and Promotion Elasticities: Seemingly Unrelated Equations," *Journal of the American Statistical Association*, 86(414), 304–215.

Blattberg, Robert C, and Ken J. Wisniewski (1989), "Price Induced Patterns of Competition," *Marketing Science, 8 (4),* 291–309.

Bodapati, Anand V., Sachin Gupta (2004), "The Recoverability of Segmentation Structure From Store-Level Aggregate Data," *Journal of Marketing Research*, 41(August), 351-364.

Bucklin, R.E., Sachin Gupta (1999), "Commercial use of UPC scanner data," *Marketing Science,* 17(3), 247–273.

Bodapati, Anand V. & Sachin Gupta (2004), "The Recoverability of Segmentation Structure from Store-Level Aggregate Data," *Journal of Marketing Research*, 41(August), 351-364.

Bucklin, Randolph E., Gary J. Russell, & V. Srinivasan (1998), "A Relationship Between Market Share Elasticities and Brand Switching Probabilities," *Journal of Marketing Research*, XXXV (Feburary), 99-113.

Carroll, J.D. (1972), "Individual Differences and Multidimensional Scaling." In R.N. Shepard, A.K.Romney, & S.B.Nerlove (Eds.) *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*. New York: Seminar Press.

Chen, Yuxin & Sha Yang (2007), "Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice," *Journal of Marketing Research,* 44 (November), 613-621.

Chintagunta, Pradeep.(1994), "Heterogeneous Logit Model Implications for Brand Positioning," *Journal of Marketing Research*, 31, 304-311.

______ (1999), "A Flexible Aggregate Logit Demand Model," working paper.

______ (2001), "Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data," *Marketing Science*, 20 (4), 442-456.

______, Jean-Pierre, Dube and Vishal Singh (2002), "Market Structure Across Stores: An Application of A Random Coefficients Logit Model with Store Level Data, " *Econometric Models in Marketing*, 16, 191-221.

______, _____ and _____ (2003), "Balancing Profitability and Customer Welfare in a Supermarket Chain, " *Quantitative Marketing and Economics*, 1, 111-147.

______, Tulin Erdem, Peter E.Rossi & Michel Wedel (2006), "Structural Modeling in Marketing: Review and Assessment," *Marketing Science*, 25 (6), 604-616.

Coombs, C.H. (1964), *A Theory of Data*. New York: Wiley.

Cooper, Lee G. (1988), "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (6), 707–23.

Day, George S., Allan D. Shocker and Rajendra K.Srivastava (1979), "Customer-Oriented Approaches to Identifying Product-Markets," *Journal of Marketing*, 43 (4), 8-19.

DeSarbo, Wayne S. , Youngchan Kim, Michel Wedel, and Duncan K.H.Fong (1998), "A Bayesian Approach to the Spatial Representation of Market Structure from Consumer Choice Data," *European Journal of Operational Research*, 111, 285-305.

_____, Rajdeep Grewal, and Jerry Wind (2006), “Who Competes with Whom? A Demand–Based Perspective for Identifying and Representing Asymmetric Competition,” *Strategic Management Journal* (27), 101–129.

Draganska, Mikhaila & Dipak C. Jain (2002), “A Likelihood Approach to Estimating Market Equilibrium Models,” *Management Science*, 50 (5), 605-616.

Efron, B. and C.Morris (1975), “Data Analysis Using Stein’s Estimator and Its Generalizations,” JASA 70, 311-319.

Elrod, Terry (1988a), “Choice Map: Inferring a Product-Market Map From Panel Data,” *Marketing Science* 7 (1), 21-40.

______ (1988b), “Inferring an Ideal-Point Product-Market Map From Consumer Panel Data,” *Data, Expert Knowledge and Decisions, 240-249.*

______ (1991), “Internal Analysis of Market Structure: Recent Developments and Future Prospects,” *Marketing Letters,* 2(3), 253-266.

_______, Keane, M.P. (1995), “A Factor-Analytic Probit Model for Representing the Market Structure in Panel Data,” *Journal of Marketing Research*, 32 (February), 1-16.

______, Gary J. Russell, Allan D. Shocker, Rick L. Andrews, Lynd Bacon, Barry L. Bayus, J. Douglas Carroll, Richard M. Johnson, Wagner A. Kamakura, Peter Lenk, Josef A.Mazanec, Vithala R.Rao, Venkatesh Shankar (2002), “Inferring Market Structure from Customer Response to Competing and Complementary Products,” *Marketing Letters*, 13(3), 221-232.

Feenstra, Robert C. and Matthew D. Shapiro (2003), Scanner Data and Price Indexes, University of Chicago press.

Goldberg, Pinelopi Koujianou (1995), "Product Differentiation and Oligopoly in International Markets: the Case of the U.S. Automobile Industry," *Econometrica*, 63 (4), 891-951.

Gonzalez–Benito, O., M. P. Martinez–Ruiz, and A. Molla–Descals (2009), "Spatial Mapping of Price Competition Using Logit-type Market share Models and Store-level Scanner data," *Journal of the Operational Research Society*, 60, 52-62.

Gupta, Sachin,Pradeep K.Chintagunta, Anil Kaul, & Dick Wittink (1996), "Do Household Scanner Panels Provide Representative Inferences from Brand Choices? A Comparison with Store Data," *Journal of Marketing Research*, 33 (November), 383-398.

Jiang, Renna, Puneet Manchanda & Peter E. Rossi (2009), "Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data," *Journal of Econometrics*, 149, 136-148.

Kamakura, Wagner A. & Rajendra K. Srivastava (1986), "An Ideal-Point Probabilistic Choice Model for Heterogeneous Preferences," *Marketing Science*, 5 (3), 199-218.

Kamakura, Wagner A., Russell, Gary J. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 23 (November), 379–390.

Kim Byund-Do (1995), "Incorporating Heterogeneity with Store-Level Aggregate Data," *Marketing Letters* 6:2: 159-169.

Montgomery, Alan L. (1997), "Creating Micro–Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science*, 16(4), 315–337.

Montgomery, Alan L. and Rossi, Peter E.(1999), "Estimating Price Elasticities with Theory–based Priors", *Journal of Marketing Research*, 36 (4), 413–423.

Musalem, Andres, Eric T. Bradlow & Jagmohan S. Raju (2009), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," *Journal of Applied Econometrics*, 24, 490-516.

Nayga, R.M. (1992), "Scanner Data in Supermarkets: Untapped Data source for Agricultural Economists," *Review of Marketing and Agricultural Economics*, 60 (2), 205-212.

Nevo, Aviv (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand*,"* *Journal of Economics & Management Strategy*, 9 (4), 513-548.

______ (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69 (2), 307-342.

Park, Joonwook, Wayne S.DeSarbo and John Liechty (2008), "A Hierarchical Bayesian Multidimensional Scaling Methodology for Accommodating Both Structural and Preference Heterogeneity," *Psychometrika*, 73 (3), 451-472.

Park, Sungho & Sachin Gupta (2009), "Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," *Journal of Marketing Research*, XLVI (August), 531-542.

Petrin, Amil (2002), "Quantifying the Benefits of New Products: The Case of the

Minivan," *Journal of Political Economy*, 110 (4), 705-729.

______ & Kenneth Train (2004), "Omitted Product Attributes in Discrete Choice Models," working paper, Booth Graduate School of Business, University of Chicago.

Pinkse, Joris, Margaret E. Slade, and Craig Brett (2002), "Spatial Price Competition: A Semiparametric Approach," *Econometrica* 70 (3), 1111–1153.

_____, and _____ (2004), "Mergers, Brand Competition, and the Price of a Pint," *European Economic Review 48*, 617–643.

Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Science*. 22(3) 304–328.

Russell, Gary J. (1992), "A Model of Latent Symmetry in Cross Price Elasticities," *Marketing Letters* 3(2), 157–169.

______., Ann Petersen, and Suresh Divakar (2008), "Analysis of Brand Price Competition Using Measures of Brand Similarity," *working paper*.

Ruzo, Emilio, Jose M. Barreiro & Fernando Losada (2006), "Competitive Market Analysis from a Demand Approach: An Application of the Rotterdam Demand Model," *International Journal of Market Research*, 48 (2), 193-236.

Saburi, S. & N.Chino (2008), "A Maximum Likelihood Method for an Asymmetric MDS model," *Computational Statistics and Data Analysis*, 52, 4673-4684.

Seetharaman, P.B. (2004), "Estimating Disaggregate Heterogeneity Distributions

Using Aggregate Data: A Likelihood-Based Approach," working paper, Olin School of Business, Washington University at St. Louis.

Slater, P. (1960), "The Analysis of Personal Preference," *British Journal of Statiscal Psychology*, 13, 119-135.

Sudhir, K. (2001), "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science*, 20 (1), 42-60.

Tucker, L.R. (1960), "Intra-Individual and Inter-Individual Multidimensionality." In H.Gulliksen & S.Messick (Eds.), *Psychological Scaling: Theory and Applications*. New York: Wiley.

Villas-Boas, J. Miguel & Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324-1338.

Wedel, Michel & Jie Zhang (2004), "Analyzing Brand Competition Across Subcategories," Journal of Marketing Research, XLI (November), 448-456.

Zenor, M.J. & Srivastava, R.K. (1993), "Inferring Market Structure with Aggregate Data: a Latent Segment Logit Approach," *Journal of Marketing Research*, 30 (August), 369-379.

Chapter 2: Essay 1: A Reduced-Form Model of Asymmetric Competitive Structure Using Aggregate Scanner Data

2.1. Introduction

The problem of identifying competitive structure among brands in a market (i.e., estimating a map of product differentiation) is important for manufacturers, service providers, and retailers managing many aspects of marketing strategy (pricing, brand repositioning, etc.). In addition, government agencies need information about competitive structure in order to apply antitrust laws concerning mergers and monopolization.

Extant methods in marketing for calibrating competitive structure maps rely mainly on disaggregated data about consumers' perceptions, preferences, and choices (e.g., see Elrod et al. 2002) or on panel data (e.g., Hendry model; see Butler and Butler 1970, 1971). These forms of data, however, may be unavailable, costly, and unrepresentative of a market of interest. Scanner data at the store or chain level, by contrast, are readily available from retailers and provide timely information about the structure of competition specific to given localities. In this article, I consider issues that arise when using store-level scanner data to estimate competitive structure for a product category.

With market-level or store-level scanner data, a common approach to calibrating market structure involves, first, estimating the parameters of a system of demand equations that relate quantity demanded for each product to the prices of the product and its rivals, and, then, post-processing the cross-price elasticity estimates to arrive at a structural map using such tools as factor analysis (e.g., Cooper 1988) or multidimensional scaling (e.g., Allenby 1989).

A problem with past applications of this approach is that market structure has been linked only to cross-price elasticities, and not to own-price elasticities or demand intercepts. By contrast, I think that competitive structure should underlie all of these demand parameters. Standard marketing intuition, for example, suggests that desirable brands (i.e., brands with high brand equity) might be associated with large brand intercepts, low own-price elasticities (in absolute value), and large cross-price effects on competitors' quantities.

A second problem with the above approach is that it uses a sequence of estimation steps (i.e., estimation of demand parameters in a first stage, and analysis of market structure in a later stage), which gives rise to three shortcomings: (a) conceptual inconsistency; (b) inefficient estimation; and even (c) infeasibility of estimation. The first shortcoming pertains to the latter-stage analysis making assumptions about competitive structure inconsistent with the first-stage analysis (such as assuming that cross-price effects are symmetric). The second shortcoming pertains to the estimation error not being passed from the first model to the second. This estimation error can be substantial when the number of brands is large which may lead to estimates that are unstable (i.e., high variance) or unreliable (i.e., incorrect signs or unreasonable magnitudes) (Montgomery and Rossi 1999). The third shortcoming pertains to the model in the first stage being inestimable. This can occur when manufacturers change the prices of all their brands in a product category in parallel, resulting in perfect price colinearity. The second data set analyzed in this article illustrates this problem.

The current article addresses these problems by (a) recognizing and modeling explicitly the connections between the underlying competitive structure and various brand-specific demand parameters, (b) estimating the demand parameters and the underlying structural map in one step. My analysis provides contributions at three levels: conceptual, methodological, and managerial.

Conceptually, the proposed model, built on related concepts in marketing, economics, and psychometrics, establishes theoretical linkages between a market structure map and own-price elasticities, cross-price elasticities as well as intercepts in the demand model. Specifically, I show how the asymmetric feature of competition contained in cross-price elasticities is related with some brands' power parameters. And I then develop a proposition that shows the relationship between my brand power parameter and the marketing concepts of *competitive clout* and *vulnerability* (Cooper 1988; Kamakura and Russell 1989). I also develop a second proposition that shows the relationship between own-price elasticity, my brand power parameter, and the psychometric concept of *spatial density*.

Methodologically, I derive a market structure map directly from demand functions by utilizing two techniques – dimension reduction and adaptive Bayesian shrinkage, which allow the estimation of cross-price elasticities even for products that have highly collinear prices. Intuitively speaking, when price histories are highly collinear, the cross-price elasticities are hard to be reliably estimated since the information contained in the data itself (represented as the variation in the independent variables) is too limited. To enable estimation, some extra information is needed. The techniques proposed here follow this logic by borrowing information across different demand function parameters (through hard constraints imposed by dimension-reduction) and across different brands (through soft constraints imposed by Bayesian shrinkage). The underlying idea is that those high-dimensional brand-specific demand function parameters (intercepts, price elasticities) are determined by some common low-dimensional market map. A product's position on this map might tell information about the other products' positions on the same map.

Since asymmetric multidimensional scaling literature is applied here to help project demand parameters on the market map, the cross-price elasticities estimated may be influenced by the functional form that I choose. On the other hand, the extent of the information-sharing across brands' positions is solely determined by the data, therefore, the estimates do not rely on any utility theory assumptions typically used in the literature (e.g., Montgomery & Rossi 1999).

A further methodological contribution is that I engage in model selection from among common specifications in marketing well-suited to the modeling purpose, including the vector model, a dominant point model, or a fully saturated formulation (i.e., a formulation that uses separate parameters for each brand or each pair of brands).

Managerially, analysts can use the proposed model to examine strategy scenarios involving changing competitive positioning in order to predict the impact on demand intercepts, own-price elasticities and cross-elasticities. And managers with access to scanner data at the store or chain level can also readily use the proposed model to track changes over time in asymmetric competitive structure.

Antecedents. My conceptual contribution builds on the work of Pinkse, Slade, and Brett (2002) and Pinkse and Slade (2004), who are among the few researchers that explicitly relate those brand-related parameters of a demand function to market structure information featured by some common product attributes. That work, however, requires collecting additional information describing product-specific characteristics or attributes, and, therefore, is constrained by the type of information that happens to be available. My approach, by contrast, estimates latent product attributes and allows prominent features of the data to emerge, without the influence of data-availability constraints.

My methodological approach of using adaptive Bayesian shrinkage to obtain and stabilize brand estimates utilizes insights gained from prior research in which Bayesian techniques are used to enable the estimation of parameters in the presence of severe multicolinearity, such as with ridge regression (Hoerl and Kennard 1970; Kubokawa and Srivastava 2004; Srivastava and Kubokawa 2005). I also build on the use of Bayesian techniques in marketing to improve the quality of demand parameters estimates (e.g., Blattberg and George 1991, Montgomery 1997, Wedel and Zhang 2004). My application of Bayesian analysis, in particular, emphasizes its adaptive shrinkage nature in the sense that usage of informative priors is avoided by estimating the parameters of these prior distributions and using uninformative priors for the hyper-parameters that characterize these prior distributions. In this way, the degree of shrinkage employed by the model is determined by the data in a fully Bayesian multi-level model. No shrinkage is a feasible limiting case of this model due to the use of uninformative priors for the hyper-parameters. My approach builds on the "adaptive shrinkage" idea originally proposed by Efron & Morris (1975), and given detailed explanation by Rossi, Allenby & McCulloch (2005) and also Montgomery and Rossi (1999). Later in the essay, I will give a more thorough illustration about how the adaptive Bayesian shrinkage approach is implemented in my model and how my implementation differs from this previous literature.

My methodological approach of modeling cross-price elasticities and own-price elasticities adapts ideas from the psychometric asymmetric similarity literature (e.g., Young 1975; Chino 1978, 1990; Holman 1979; Weeks and Bentler 1982; Okada & Imaizumi 2007; Chino 2008). My modeling of price elasticities in this way departs from Gonzalez-Benito et al. (2009), which constitutes the only paper in marketing of which I are aware that also estimates a latent market structure model directly from reduced-form based demand functions. [8] That paper builds upon Russell's (1992) Latent Symmetric Elasticity Structure pattern (LSES) model, but, unlike my approach, the estimated competitive structure map only underlies the cross-price elasticities (there is no information-sharing across brands or across other parameters of the demand model).

[8] I notice there exists another stream of research, which are built upon utility-based demand function and also enable the direct derivation of a market map (Chintagunta, Dube, Singh (2002, 2003), Kamakura & Srivastava 1986). However, I focus my attention in this chapter solely on the reduced-form demand model, and leave any discussion regarding the utility-based demand model to Chapter 3. Besides, comparison of these two types of demand model will be provided in detail in Chapter 4.

The rest of this essay is organized as follows. Section 2.2 introduces my modeling framework and the various equation specifications that implement my framework. Section 2.3 describes my approach to model selection, identification, and estimation. Section 2.4 applies my modeling approach separately to two different datasets both describing weekly grocery sales and prices: one dataset describes beer sales in a test market in the U.S.; and the second dataset describes soft drink sales in a supermarket in St. Louis. Section 2.5 concludes with a summary of limitations of my approach and directions for future research.

2. 2 Modeling Framework

My approach uses several demand model components to situate brands in a multidimensional space. In particular, I assume that latent brand locations underlie (a) the symmetric structure of cross-price elasticities (CPEs), (b) asymmetric dominance relationships embedded within CPEs, (c) own-price elasticities (OPEs), and (d) brand intercepts.

I consider a set of constant-elasticity demand equations[9]

$$\log(E(q_{it})) = \alpha_0 + \alpha_i + \beta_{ii}\log(p_{it}) + \sum_{j \neq i} \beta_{ij}\log(p_{jt}) + f(CV_{it}), \qquad (2\text{-}1)$$

where $i, j = 1,...,n$; $t = 1,...,T$; n is the number of brands; T is the number of time periods covered in the data; $E(q_{it})$ is the expected unit sales for brand i at time t. Expected unit sales is used here since the data used later constains zero quantity demand. β_{ij}, $i \neq j$, are cross-price elasticities; β_{ii} are own-price elasticities; α_i

[9] The constant elasticity assumption can be a restrictive since structure is always the same for price and income. I will consider trying other demand function forms to relax this assumption in the future.

are brand intercepts (mean-centered around α_0); and CV_{it} are covariates, possibly time-dependent. I observe that in many standard econometric models, Equation (2-1) takes a form where q_{it} appears on the left-hand side in place of $E(q_{it})$ and an additive error term is included on the right hand side. I mean-centered log prices across brands and time in my model; and, as a result, α_i describes the log sales differences across brands when all prices are set equal to the mean price in the dataset (so α_i can be interpreted as a measure of the attractiveness of brand *i*) [10].

My basic idea is as follows. I suppose that the parameters β_{ij}, β_{ii}, and α_i are functions of latent coordinates of the associated brands, where the coordinates are described by $M \times 1$ vectors $\boldsymbol{\theta}_i$, for each brand $i = 1,...,n$. In particular, I assume that the key parameters in (2-1) are described by the general relationships

$$\beta_{ij} = g(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j),\ \beta_{ii} = \beta(\boldsymbol{\theta}_i),\ \text{and}\ \alpha_i = \alpha(\boldsymbol{\theta}_i), \tag{2-2}$$

where $i \neq j$. That is, the cross-price elasticity between brands *i* and *j* is described by the relationship between the product locations $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$; and the own-price elasticity and brand-specific intercept for brand i are also influenced by the product location $\boldsymbol{\theta}_i$. In this way, underlying structural information present in the data may inform the estimation of multiple model components. As discussed in the introduction, this is in the spirit of past work in marketing, which either post-processed elasticity estimates (e.g., Allenby 1989) or which involved one-step

[10] It is worth noting that when $f(CV_{it})$ is also brand-dependent, appropriate mean-centering may be necessary to maintain the interpretation of the intercepts.

analysis of cross-price elasticities only (Gonzalez-Benito et al. 2009). This is also in the spirit of work in economics by Pinske and Slade (2004), who assume that relationships similar to g, β, and α are functions of measured product characteristics. My point of departure is that g, β, and α are functions of *latent structural parameters*, $\boldsymbol{\theta}_i$, $i=1,\ldots,n$. Arriving at such latent structure is more challenging, from an estimation perspective, than assuming that the structure is a function of measured product characteristics. In principle, one could think of the relationships $\beta_{ii}=\beta(\boldsymbol{\theta}_i)$ and $\alpha_i=\alpha(\boldsymbol{\theta}_i)$ as reflective of brand equity (Aaker 1991, 1996 and Keller 1993) and $\beta_{ij}=g(\boldsymbol{\theta}_i,\boldsymbol{\theta}_j)$ as reflective of brand substitutability relationships.

In practice, the functions in (2-2) must be determined. This involves selecting suitable specifications for g, β and α and, then, estimating these specifications with the data. The following development outlines alternative specifications suggested by the literature. (As a backdrop to this development, Appendix 1 summarizes the notation of this paper.)

2.2.1 Decomposing Cross-Price Elasticities

In economics, the higher the cross-price elasticity between two brands, assuming they are substitutes in the first place, the more substitutable they are for each other. This constitutes my rationale for modeling β_{ij} as a function of the proximity of the two brand locations $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ when I select a specification for $\beta_{ij}=g(\boldsymbol{\theta}_i,\boldsymbol{\theta}_j)$.

Because β_{ij} is not necessarily equal to β_{ji}, it is desirable for an elasticity-based market structure analysis to be able to explicitly capture asymmetries. Indeed, competitive asymmetries in CPEs are well documented (DeSarbo, Grewal and Wind 2006), and asymmetric patterns are shown to exist between high-share vs. low-share brands (Sethuraman and Srinivasan 2002), high-quality vs. low-quality brands (Blattberg and Wisniewski 1989), national vs. store brands (Kamakura and Russell 1989), and high-priced brands vs. low-priced brands. To model such asymmetries, Cooper (1988) describes how an asymmetric elasticity matrix can be structured by three-way factor analysis. Similarly, Russell's (1992) LSES model and its applications (e.g., Gonzalez-Benito et al. 2009) capture competitive asymmetry by including brand specific coefficients that reveal the overall impact of a brand on its competitors. Innovative as they are, these two approaches appear somewhat cumbersome to use directly as the basis for my specification for $\beta_{ij} = g(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$. Russell's (1992) LSES model, for example, requires CPEs to be market-share based, and, my starting point is instead the standard econometric specification (2-1) for CPEs. And rather than build my model around factor analysis (as Cooper 1988 does), I find it more convenient to work in a multidimensional scaling context following the psychometric asymmetric similarity literature.

I follow a long history of consideration of systematic asymmetries in proximity data[11] (e.g., Young 1975; Chino 1978,1990; Saburi & Chino 2008), all

[11] Proximity data is defined in Shepard (1972) as a rectangular matrix whose cells contain measures of similarity between the row and column objects. Often the row and column objects are the same.

of which departs from traditional distance-based models of similarity (e.g., Shepard 1962a, 1962b; Tversky 1977). Several researchers in psychometrics have proposed what is called the *skew-symmetry* model, which decomposes matrices of (dis)similarity judgments among a set of objects into symmetric and asymmetric components, and represents the latter parsimoniously by as few as one dimension (e.g., see Weeks and Bentler 1982; Saito 1986; Saito and Takeda 1990; Satio 1991, Okada and Imaizumi 1987). Some approaches go a step further in parsimony by representing symmetric and asymmetric components in the same spatial configuration. Examples include the slide-vector model (Zielman and Heiser 1993), multidimensional unfolding models (DeSarbo and Grewal 2007), the hill-climbing model (Borg and Groenen 2005), and multidimensional scaling with a dominance point (Okada and Imaizumi 2007). I will use this last approach below. Yet another approach includes the *additive similarity-bias model* (see Holman 1979, Nosofsky 1991, Carroll 1976, and also Krumhansl 1978).

In particular, following the skew-symmetric approach, I write the matrix of cross-price elasticities β_{ij}, $i \neq j$, together with zero diagonal entries, as $\mathbf{B}$ (i.e., the Greek letter capital beta). This constitutes an asymmetric proximity matrix. I decompose this asymmetric matrix $\mathbf{B}$ into symmetric and skew-symmetric components, $\mathbf{B} = (\mathbf{B} + \mathbf{B}')/2 + (\mathbf{B} - \mathbf{B}')/2 \equiv \mathbf{S} + \mathbf{A}$, which can be rewritten in scalar notation as[12]

[12] There might exist other ways to decompose the cross-price elasticities, for example, the LSES structure proposed by Russell (1992), or SVD decomposition. The skew-symmetric structure is used here since it is less restrictive compared to LSES model, and more tractable than the SVD decomposition.

$$\beta_{ij} = s_{ij} + a_{ij}, \qquad (2\text{-}3)$$

where $s_{ij} = (\beta_{ij} + \beta_{ji})/2$ has the property of symmetry ($s_{ij} = s_{ji}$) and $a_{ij} = (\beta_{ij} - \beta_{ji})/2$ has the *skew-symmetric property*, $a_{ij} = -a_{ji}$ for all $i \neq j$. Note that s_{ij} describes the average level of competition between brands *i* and *j*, and a_{ij} describes the dominance relationship between brands *i* and *j*. That is, when $a_{ij} > 0$, I have $\beta_{ij} > \beta_{ji}$, and I can interpret brand *j* as being more dominant than brand *i*.[13]

I assume that the symmetric similarity measure s_{ij} is a decreasing linear function of an underlying distance metric,

$$s_{ij} = \phi_{CPE} - d_{ij}, \quad i \neq j, \qquad (2\text{-}4a)$$

where ϕ_{CPE} is a constant and d_{ij} is a traditional (symmetric) inter-point distance between brands *i* and *j*.

Next, consistent with the common practice for the skew-symmetry model, I model a_{ij} with a one-dimensional linear form

$$a_{ij} = -x_i + x_j, \qquad (2\text{-}4b)$$

for all *i* and *j*, where x_i and x_j are coordinates of objects *i* and *j* on this one dimension.

I combine (2-4a) and (2-4b) with (2-3) to yield

[13] Here, I use the term "dominant relationship" to describe the dominance data, which in current context can be represented by the skew-symmetric component of cross-price elasticities. Consistent with the multidimensional literature, or in particular, Shepard's (1972) taxonomy of multivariate data used to derive spatial representations of structure, the dominance data describe a square matrix whose rows and columns represent the same set of alternatives, with each cell entries measure the extent to which the row alternative is preferred to, is chosen over, defeats, or otherwise dominates the column object (MCGrath 1985).

$$\beta_{ij} = \phi_{CPE} - d_{ij} - x_i + x_j . \qquad (2\text{-}4)$$

Equation (2-4) constitutes my basic formulation for CPEs.

Prior to considering ways of modeling d_{ij} and x_i, I interpret this formulation. I observe that $x_i > x_j$ implies $\beta_{ij} < \beta_{ji}$, so the price of brand *j* has less of an influence on quantity demanded of brand *i* than the price of brand *i* has on the quantity demanded of brand *j*. This fact allows us to interpret x_i as the relative dominance, or power, of brand *i*. I refer to x_i as the *brand power parameter* of product *i*.

My brand power parameter constitutes a one-dimensional measure that accounts for the asymmetry in cross-price elasticities. If the impact of price changes for brand *i* on the quantities demanded of other brands is consistently higher than the impact of price changes of these other brands on the quantity demanded of brand *i*, then the brand power parameter, x_i, will be large. My definition of brand power is in contrast to two other definitions of brand power. (1) Na et al. (1999) measure "brand power image" as the weighted average of measured brand attributes, benefits, or values. Although my brand power parameter is not directly related to measured brand attributes, I will later show that x_i is modeled to be connected to a brand's location on a market structure map featured by some latent product characteristics. (2) Steenkamp and Dekimpe (1997) operationalize power (of store vs. national brands) along two dimensions: "intrinsic loyalty" (to a brand) is defined to be a brand's ability to keep its current customers; and "conquesting power" is "the proportion of the market's non-loyal customers that one is able to attract in a given time period." It is worth noticing that their dimensions, "intrinsic loyalty" and "conquesting power", are closely related to the familiar marketing concepts of competitive vulnerability and clout (Cooper 1988; Kamakura and Russell 1989). And this coincides with my proposition 1, as discussed below. Specifically, in the following proposition, clout describes the extent to which a focal brand *i* exerts influence on all other brands; and I define $Clout_i = \sum_{j \neq i} \beta_{ji}$. Vulnerability describes the extent to which all other brands collectively exert an influence on a focal brand; and I define $Vul_i = \sum_{j \neq i} \beta_{ij}$.[14] I can now show the following intuitive relationships between x_i (power of brand *i*), β_{ij} (cross-price elasticity between brands *i* and *j*), $Clout_i$, and Vul_i:

Proposition 1. Equation (2-4) implies the following relationships:

(a) $x_i > x_j$ implies $\beta_{ij} < \beta_{ji}$ for all $i \neq j$.

(b) $x_i = \dfrac{Clout_i - Vul_i}{2n}$, if $\sum_i x_i = 0$.[15] Here *n* is the total number of brands.

---

[14] These definitions differ from Kamakura and Russell (1989), who define $Clout_i = \sum_{j \neq i} \beta_{ji}^2$ and $Vul_i = \sum_{j \neq i} \beta_{ij}^2$, and from Cooper (1988), who defines $Clout_i = \sum_{j \neq i} \beta_{ji}^2 + \beta_{ii}^2$, $Vul_i = \sum_{j \neq i} \beta_{ij}^2 + \beta_{ii}^2$. Compared with their definitions, I believe mine are more suitable for quantity-based price elasticities, which are not necessarily to be positive due to the category expansion effects (Cooper 2002). Taking the squared value of cross-price elasticities when estimating clout and vulnerability would mask this negative effect.

[15] I can also relate the brand power parameter, x_i, to an early model called the additive similarity-bias model (Holman 1979; Nosofsky 1991), also known as a hybrid model (Carroll 1976). This model assumes that a proximity measure δ_{ij} between two objects *i* and *j* is some increasing function (often assumed linear, for simplicity) of a symmetric similarity measure, $\bar{s}_{ij}$; a row bias function, r_i ; and a column bias function, c_j : $\delta_{ij} = \bar{s}_{ij} + r_i + c_j$. The

Proof. See Appendix B.

Discussion. Statement (a) allows us to interpret x_i as a measure of *brand power* of product *i*, as I noted earlier. Statement (b) indicates that higher clout and lower vulnerability imply higher brand power, x_i, as might be expected. Thus, my proposed power parameter captures asymmetric pricing effects and links these with the concepts of clout and vulnerability.

2.2.2 Modeling the Symmetric and Asymmetric Components of Cross-Price Elasticities

I now consider modeling specifications for the symmetric and asymmetric components, d_{ij} and x_i. Since the asymmetric component is relatively new to marketing and economics, I devote more attention to this.

Symmetric Component. Note that the smaller the inter-brand distance, d_{ij}, the greater the cross-price elasticity, which constitutes greater substitutability between these brands. In particular, following the traditional multidimensional scaling (MDS) model (Shepard 1962a, 1962b), I assume that d_{ij} is the Euclidean distance (in an M-dimensional metric space) between brands *i* and *j*:

$$d_{ij} = \sqrt{\sum_{m=1}^{M} (\theta_{im} - \theta_{jm})^2}\ , \tag{2-5}$$

where θ_{im} describes the coordinate of brand *i* on dimension *m*.

Skew-Symmetric Component. I identify three ways of modeling the

justification for this form is that the row and column biases reflect distinguishing properties of individual items. For such a model, I can show a third relationship: (c) $x_i = (c_i - r_i)/2$, if $\sum_i x_i = 0$.

brand power term, x_i.

1. Estimate separate parameters, x_i, for each *i*.

2. Express x_i in terms of the brand locations θ_{im} using a form of *dominance (ideal) point model*:

$$x_i = -\omega_1 \sqrt{\sum_{m=1}^{M} (\theta_{im} - y_m)^2} \,. \tag{2-6}$$

3. Express x_i in terms of the brand locations θ_{im} using a *vector model*:

$$x_i = \sum_{m=1}^{M} \upsilon_{1m} \cdot \theta_{im} \,. \tag{2-6'}$$

In specifications 2 and 3, $\omega_1 \geq 0$, and y_m and υ_{1m}, $m = 1,...,M$, are all constants to be estimated.

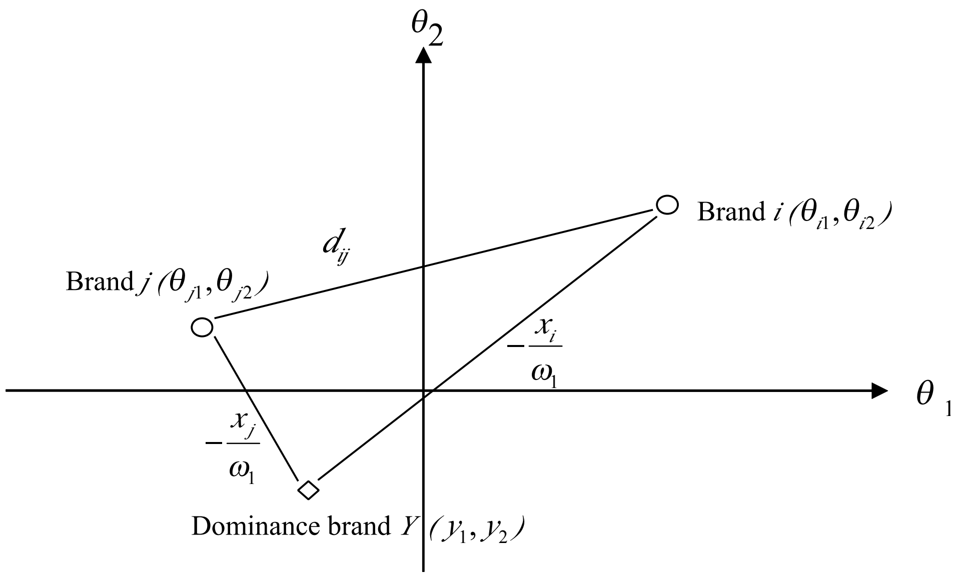
Operationally, when the number of brands in the market is small, it may be feasible to estimate separate parameters, x_i, for each *i*. Since the same basic relationships among the brands are invariant with respect to an additive constant added to x_i; for all *i*, without loss of generality, I set $\sum_i x_i = 0$ as an identification restriction for the first formulation. Note that if the number of brands is large, this approach of estimating separate parameters, x_i, for each *i*, may be undesirable.

The formulation in (2-6) above is a modified adaptation of Okada and Imaizumi's (2007) model of MDS with a dominance point. Here, the hypothetical point of greatest dominance $Y = (y_1, ... y_M)$ is situated in the same space as the symmetric components, and the dominance parameters are structured in relation to each brand's distance to this point, Y. Since ω_1 is restricted to be nonnegative, the closer a brand is to Y, the stronger the brand's power. Furthermore, I can interpret ω_1 as the asymmetry weight because this represents the salience of the asymmetric model component in describing the cross-price elasticity relationships. When ω_1 equals zero in the dominance point model, there is no asymmetry present in the structure of cross-price elasticities.

It is worthwhile comparing (2-6) with the formulation used in Saburi & Chino (2008). The latter adopts the model of Okada & Imaizumi (1987) to represent the asymmetric proximity data, which models the skew-symmetric part of this data matrix using a one-dimensional construct (a treatment similar to (2-4b)), and interprets each object's coordinates on this one-dimensional construct as the radius of the circle attached to this object. An object with larger radius is considered as more proximate to an object with smaller radius and *vice versa*. I do not use this "radius" interpretation of skew-symmetric component for the following reasons: (a) this "radius" interpretation is less parsimonious than the formulation in (2-6). While the latter requires estimating only one more set of locations (of the dominant point), which adds M (the number of dimensions) parameters, the former involves estimating n independent radiuses for n alternatives. As in a typical MDS method, M is usually much smaller than n. (b) A test of the asymmetry is more straightforward and simpler in (2-6) than with the radius interpretation. With (2-6), the significance of ω_1 directly indicates the salience of asymmetry effect, while in Saburi & Chino (2008), various independent tests of asymmetry need to be developed.

I graph the assumed structure for the dominance point model, when $M = 2$, in Figure 2-1. Here, the smaller the distance $-x_i / \omega_1$, the larger x_i, and the more dominant is brand *i*.

Figure 2-1 Modeling Asymmetric CPEs with a Dominance Brand



Lastly, the vector model formulation (2-6') describes an increasing progression of magnitude of the dominance parameters x_i, along the direction of the estimated gradient vector $(\upsilon_{11}, \ldots, \upsilon_{1M})$. Note that when $(\upsilon_{11}, \ldots, \upsilon_{1M}) = 0$ in the vector model, there is no asymmetry present in the structure of the cross-price elasticities. Also note that the vector model can be shown to be a limiting case of the dominance-point model as the distance of the dominance point *Y* from the origin approaches infinity.

2.2.3 Modeling Own-Price Elasticities and Brand Intercepts

I now specify formulations for the other model components $\beta_{ii} = \beta(\boldsymbol{\theta}_i)$ and $\alpha_i = \alpha(\boldsymbol{\theta}_i)$.

Own-Price Elasticities. I note three ways to model own price elasticities:

1. Estimate separate parameters, β_{ii}, for each $i = 1, \ldots, n$.

2. Express β_{ii} in terms of the brand locations θ_{im} using a form of *ideal (dominance) point model*:

$$-\beta_{ii} = \phi_{OPE} + \omega_2 \sqrt{\sum_{m=1}^{M} (\theta_{im} - z_m)^2}, \omega_2 > 0 . \quad \text{(2-7)}$$

3. Express β_{ii} in terms of the brand locations θ_{im} using a form of *vector model*:

$$-\beta_{ii} = \phi_{OPE} + \sum_{m=1}^{M} \upsilon_{2m} \cdot \theta_{im} . \quad \text{(2-7')}$$

In specifications 2 and 3 above, ω_2 is a constant constrained to be positive, and ϕ_{OPE} and υ_{2m}, $m = 1,...M$ are constants without sign constraints. All these constants are to be estimated.

When the number of brands in the market is small, it may be simple enough to estimate separate parameters, β_{ii}, for each i=1,…,n. The shortcoming of this method, however, is that OPEs do not contribute to the estimation of the market map as reflected in (2-5), therefore, cannot be explained by the derived market map either. This undermines the usefulness of the market structure model for making predictions for repositioned, new, or deleted brands because the effects of such changes on own price elasticities and therefore on demand cannot be predicted.

Embedding the OPEs in competitive map solves this problem. Equations (2-7) and (2-7') achieve this by expressing own-price elasticities in terms of the brand locations θ_{im}. In particular, the formulation in (2-7) describes the magnitude (i.e., absolute value[16]) of brand *i*'s own-price elasticity as a linearly increasing function of the squared distance between the brand and the hypothetical brand on the map located at $Z = (z_1, ... z_M)$. The closer the brand to the hypothetical brand $Z = (z_1, ... z_M)$, the smaller the OPE is in absolute value. By contrast, the vector model describes an increasing progression of magnitude of the OPEs, β_{ii}, along the direction of the estimated gradient vector $(\upsilon_{21}, \ldots, \upsilon_{2M})$.

In order to interpret these formulations, I derive the following relationships.

Proposition 2. Assuming utility-maximizing consumers under a linear budget constraint and writing the income elasticity of brand *i* as g_i, I have

(a) $-\beta_{ii} = (\sum_{j \neq i} \beta_{ij}) + g_i = Vul_i + g_i$; and (2-8)

(b) $-\beta_{ii} = (n-1) Density_i - nx_i + g_i$, if $\sum_i x_i = 0$; (2-9)

where $Density_i \equiv (\phi_{CPE} - \sum_{j \neq i} d_{ij} / (n-1))$.

Proof. See Appendix B. [Note (a) arises from utility maximizing behavior under a

[16] Since own-price elasticity should be negative, the negative of OPE equals its absolute value.

linear budget constraint; and (b) arises from (2-4).]

Discussion. I can make several observations from Proposition 2.

First, when the income effect g_i is small, the OPE equals minus the vulnerability of brand *i*, as an approximation.

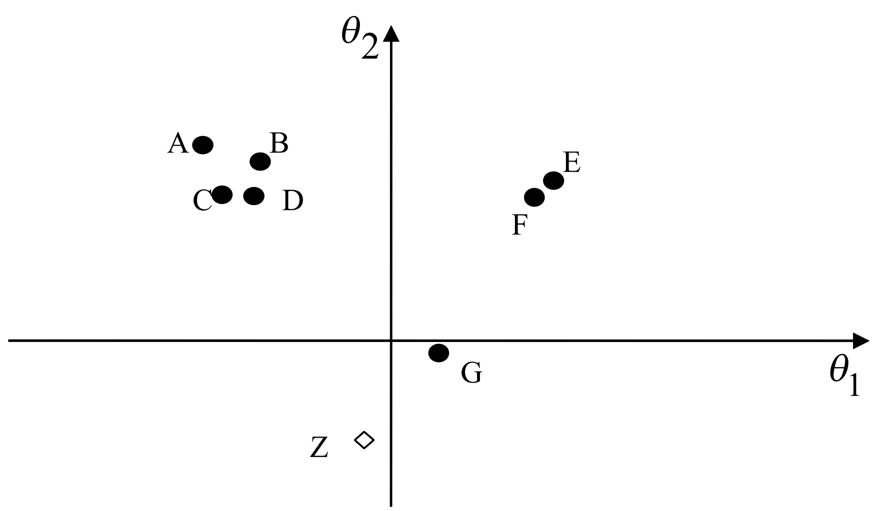
Second, when OPEs are represented according to the ideal (dominance) point formulation of (2-7) and income effects for the product category under study are small (or similar across brands)[17], the hypothetical brand $Z = (z_1, ... z_M)$ can be interpreted as the least vulnerable brand on the map.

Third, according to Proposition 2 (b), when the income effect g_i is small, a brand's own-price elasticity may be viewed as depending on at least two factors: (i) its distinctiveness from all the other brands in the market, as indicated by a spatial density measure, $Density_i \equiv (\phi_{CPE} - \sum_{j \neq i} d_{ij} / (n-1))$; and (ii) its dominance parameter, x_i . Thus, Proposition 2 (b) indicates that the lower the spatial density around a product *i*, *ceteris paribus*, the less price sensitive the product will be (- β_{ii} low in magnitude). Intuitively, a product that is relatively unique will be less price elastic. Furthermore, the higher a product *i*'s power relative to all other products, *ceteris paribus*, the less price sensitive the product will be (i.e., the product will face inelastic demand.). Thus, a powerful product will be less price elastic. I illustrate this idea of spatial density in Figure 2-2, in which seven brands are shown, denoted by *A, B, C..., G* . Holding all the other

[17] When the income effects are zero, the Slutsky equation together with the symmetry of substitution terms for Hicksian demand (Varian 1992, pp. 120 and 123) implies symmetric cross-price effects, i.e., $\partial q_i / \partial p_j = \partial q_j / \partial p_i$. This suggests further constraints that could be included in the model, which I do not impose in order to avoid loss of generality.

factors constant, brand *G* should have the smallest absolute own-price elasticity and the hypothetical brand *Z* (interpreted as having the "least vulnerable" possible brand location) should be located in a sparse space, likely near brand *G* (or further out southwest in the third quadrant). Furthermore, any two brands that are very similar to each other should have similar own-price elasticities because brands similar to each other should also have similar distinctiveness relative to other brands.

Figure 2-2 Modeling OPEs with a Least Vulnerable Brand



Overall, my Propositions 1 and 2 enable us to establish the linkage between cross-price asymmetries, clout, vulnerability, own-price elasticity, and spatial density.

Note that my measure of spatial density (or distinctiveness) is similar to that of Krumhansl's (1978) distance-density model, in which an object's distinctiveness is related to the spatial density in a region surrounding it in the multidimensional configuration, and this spatial density is measured by the sum of a monotonically decreasing function of the distances of all other objects to the focal object. In this context, the denser this region, the more difficult it is for an object to distinguish itself.

It is worth noting further that the spatial density explanation of own-price elasticity has also been alluded to in the economics and marketing literatures. Pinkse and Slade (2004) allow a brand's OPE to depend on the number of neighbors that the brand has in an exogenously determined product-characteristic space. In a related development, Bronnenberg and Vanhonacker (1996) examine the implications of a consumer only responding to price variation of brands in the consumer's choice set (which they call *local price response*). They show that the fewer the number of other brands appearing with a focal brand in a consumer's choice set, the lower the focal brand's vulnerability (and OPE). They also argue that the more frequently a focal brand appears with other brands in consumers' choice sets, the higher the focal brand's clout.

Brand Intercepts. I next express the intercepts α_i in (2-1) in terms of the brand location parameters, θ_{im}, $m = 1,...,M$. In particular, I suggest two ways of modeling α_i:

1. Estimate separate parameters, α_i, for each $i = 1,\ldots,n$.

2. Express α_i in terms of the brand locations θ_{i1} on the first dimension in (2-4):

$$\alpha_i = \omega_0 \cdot \theta_{i1}, \omega_0 \neq 0\,, \tag{2-10}$$

where θ_{i1} is the coordinate of i^{th} brand on the first dimension, which I interpret as "perceived attractiveness" for brand *i*. Linking the intercept α_i to first dimension θ_{i1} aids in map interpretation, in the spirit of Bentler and Weeks (1978), who proposed that for any multidimensional study, external information (i.e, prior analysis, theory) if of vital importance. It can not only facilitate the interpretation of the resulting configuration, but also help achieve parsimony of the model with even improved model performance. The authors then introduced a class of restricted MDS models wherein certain parameters may be fixed as known constants, or imposed with the proportional constraints. In specific, the authors suggested that when some scale is important to the psychological process which produces a set of data, a dimension *k* may be added to the analysis, subject to the restriction that the coordinates of the objects on this dimension *k* is proportional to this scale.

Similar to my earlier treatment of OPEs, expressing intercepts in terms of the brand locations in the market structure map is conceptually appealing because a useful map of competition should have the possibility of accounting for all brand-specific components in a model. Here I recognize that the brand's perceived attractiveness is an important attribute that influences the observed switching patterns reflected in cross-price elasticities: two brands differing in perceived attractiveness cannot have identical locations in a map that purports to explain brand competition.

Covariates. Lastly, one must determine the function *f* according to which various covariates, CV_{it}, enter the model in (2-1). For the two data sets considered later in this article, I model CV_{it} to be time-specific[18] and capture the two basic classes of components of most time series patterns: trend and seasonality. The former represents a "long term" movement that does not repeat within the time range captured by the data. The latter repeats itself in systematic intervals over time. These two components may coexist in my data. For example, for soft drink data, sales of Coke can increase over time due to such factors as population growth, but they still follow consistent seasonal patterns (e.g., the sales increase in summer season, but decrease in winter season). In particular, I use the linear function to capture the time trend, and assume the seasonal pattern follows sine and cosine functions (Cowpertwait et al. (2009) page 101-102; Stolwijk et al. (1999)).[19]

$$f(CV_t) = \gamma_1 t' + \gamma_2 \sin\frac{2\pi t'}{52} + \gamma_3 \cos\frac{2\pi t'}{52} \tag{2-11}$$

where $i = 1,\ldots,n$, $t = 1,\ldots,T$, and t' is the mean-centered t ($t' = t - \bar{t}$). In (2-11), t' is divided by 52, which represents the periodicity of the cycle when dealing with weekly data.

2.2.4 Modeling the Distribution of the Dependent Variable

For packaged-goods data sets at the store or chain level, the quantity

---

[18] When available, CV_{it} can also include other brand- and time- specific variables such as display,features, store expenses, and etc.

[19] It is well-known that sine and cosine functions can be used to build smooth variation into a seasonal model. A sine wave with frequency f (cycles per sampling interval), amplitude A, and phase shift ϕ can be expressed as

$$A\sin(2\pi ft + \phi) = \alpha_s \sin(2\pi ft) + \alpha_c \cos(2\pi ft)$$

where $\alpha_s = A\cos(\phi), \alpha_s = A\sin(\phi)$. (Cowpertwait et al. (2009) page 101-102).

demanded is integer-valued, with values ranging from 0 to several thousand. For such data, two suitable modeling alternatives consist of the Poisson distribution and the Negative Binomial distribution (NBD).

When the Poisson distribution is assumed to describe unit sales q_{it} , the probability mass function is

$$\Pr(q_{it} \mid \lambda_{it}) = (\lambda_{it})^{q_{it}} \exp(-\lambda_{it}) / (q_{it})!, \quad \lambda_{it} > 0 , \qquad \text{(2-12)}$$

where λ_{it} is the rate for this Poisson process. The expected sales in (2-1) is given by $E(q_{it}) = \lambda_{it} = V(q_{\text{it}})$.

The Poisson model is sometimes restrictive, however, because it assumes that the mean and the variance are both equal to λ_{it}. Often important determinants of the mean are unobserved, causing the variance to exceed the fitted mean. In such instances, the negative binomial distribution (NBD) may be used, which generalizes the Poisson distribution by allowing the rates λ_{it} to be distributed according to a gamma distribution across the population of customers. The gamma density is given by

$$\Pr(\lambda_{it} \mid \rho, \eta_{it}) = (\eta_{it})^{\rho} (\lambda_{it})^{\rho-1} \exp(-\eta_{it} \cdot \lambda_{it}) / \Gamma(\rho), \qquad \text{(2-13)}$$

where η_{it} is the scale parameter for the gamma distribution, differing across brands and time, and ρ is the shape parameter. Combining Equations (2-12) and (2-13) and integrating out λ_{it} yields the negative-binomial density

$$\Pr_{NBD}(q_{it} \mid \rho, \eta_{it}) = \binom{q_{it} + \rho - 1}{\rho - 1} \left(\frac{\eta_{it}}{\eta_{it} + 1}\right)^{\rho} \left(\frac{1}{\eta_{it} + 1}\right)^{q_{it}}, \ \rho > 0, \eta_{it} > 0 . \qquad \text{(2-14)}$$

When the NBD model is adopted, the expected sales in (2-1) are given by $E(q_{it})$ =

ρ/η_{it}.

2.2.5 Summary of Model Components

I conclude my discussion of model development by summarizing in Table 2-1 various possible model specifications that are alternatives to freely estimating all the parameters of (2-1). Table 2-1 shows the decisions required for each model component. Component 1 pertains to the integer dependent variable, which is given either a Poisson or negative-binomial distribution (NBD). Component 2 involves the number of dimensions (M) for the symmetric CPE structure (the "map"). Component 3 concerns whether the brand power parameter assumed to underlie CPE asymmetries is dropped altogether, freely estimated, or embedded in the map using a dominance-point or vector formulation. Component 4 addresses whether the OPEs are freely estimated or included in the map using an ideal-point or vector formulation. Component 5 describes the brand intercepts, which can be freely estimated or tied to the horizontal axis of the map. This paper restricts attention to formulations in Table 2-1 – which are all based on (2-1) and (2-11); but I acknowledge that future research may wish to consider other formulations as well.

The choice of model specification has an impact on the number of parameters that need to be estimated. I briefly provide an overview of the number of parameters that are implied by different model specifications. For Component 1, using the NBD formulation adds an additional parameter relative to the Poisson distribution. Components 2 and 3 include alternatives to freely estimating $n(n-1)$ CPEs. Component 2 imposes a symmetric CPE structure involving approximately

nM location parameters (θ_{im}; $i=1,...n$; $m=1,...,M$) – actually after the identification restrictions described in Section 3.2 below, Component 2 involves adding *nM*-(*M-1)M/2* parameters. Component 3 (the skew-symmetric structure of CPEs) adds either (a) n – 1 dominance parameters, x_i, $i=1,\ldots,n$, (imposing $\bar{x}=0$), (b) $M+1$ parameters ($\omega_1, y_m, m=1,...,M$), or (c) M parameters ($\upsilon_{1m}, m=1,...,M$). Component 4 (the structure of OPEs) involves either (a) n OPE parameters (β_{ii}, $i=1,\ldots,n$), (b) $M+1$ parameters ($\omega_2, z_m, m=1,...,M$), or (c) M parameters ($\upsilon_{2m}, m=1,...,M$). Component 5 (the structure for the brand intercepts) entails either (a) n separate intercept parameters (α_i, $i=1,\ldots,n$) or (b) one parameter (ω_0) that links the intercepts with the horizontal axis. Adopting parsimonious forms can, therefore, reduce the parameter count of these five model components from as many as n^2+n to as few as $nM+2M+1-(M-1)M/2$. Thus, for $n=20$ brands and dimensionality $M=2$, there are 420 brand-related demand parameters in a fully saturated model and only 44 brand-related structural parameters in the most parsimonious form. These calculations are based on a simple count of the number of parameters in model components 2 through 5. These parameters constitute the focal parameters of interest.

Table 2-1 Specification Decisions for the Model Components

Model Component		*Alternative Formulations*	*Equation*
1.	**Dependent variable distribution**	(a) Poisson formulation (b) Negative-binomial formulation	(2-12) (2-14)
2.	**Symmetric CPE structure**	(a) Number of dimensions (*M*) for map	(2-4a), (2-5)[1]
3.	**Skew-symmetric CPE structure**	(a) Leave x_i out of model (b) Freely estimate x_i for each brand *i* (c) Dominance-point formulation (d) Vector model formulation	– (2-4) (2-4), (2-6)[2] (2-4), (2-6')
4.s	**OPE structure**	(a) Freely estimate β_{ii} for each brand *i* (b) Dominance -point formulation (c) Vector formulation	(2-1) (2-7) (2-7')
5.	**Intercept structure**	(a) Freely estimated α_i for each brand *i*, (b) Make α_i proportional to horizontal axis	(2-1) (2-10)

[1]The symmetric CPE structure is set in (2-3), (2-4a), and (2-5), assuming $a_{ij} = 0$ Equation (2-5) determines the structure underlying d_{ij}.

[2]The asymmetric CPE structure is set, in particular, by (4b) – the relationship $a_{ij} = -x_i + x_j$, $i \neq j$ — in Equation (2-4). Equation (2-6) and (2-6') specify the structure underlying x_i.

2.3 Model Selection, Identification, and Estimation

Table 2-1 describes many possible model specifications. And I notice the model selection process can be quite computationally demanding when the dataset is large: with two to four options for each component of a demand model as in Table 2-1, we face totally 48 combinations ($2 \times 4 \times 3 \times 2$) of the model specifications. However, there are certain conceptual considerations that will implicitly simplify this process. First, I believe it is always conceptually desirable to pursue the most parsimonious model, the one in which full market structure information (i.e., price elasticities and intercepts) are used to derive the market map. Second, it is conceptually consistent to think own-price elasticity takes the same formulation (a dominant point or a vector structure) as the power parameter in cross-price elasticities. And these greatly reduce the model selection process to a selection between a vector or a dominant point formulation, a choice a researcher sometimes has to make even with a modeling attempt of disaggregate data (Park, Desarbo, Liechty 2008).

However, in order to find good initial values for the most parsimonious model, in terms of the order of estimation of models, I found it convenient to start with a saturated model (with as many parameters as possible freely estimated), and then to successively replace various model components with more parsimonious formulations. By proceeding in this sequence, I could then immediately identify problems with convergence and identification and trace them to their causes.

2.3.1 Model Selection

For model selection, I apply the deviance information criterion (DIC) to *focal* parameters (Spiegelhalter et al. 2002). In particular, let $\overline{D}$ denote the mean deviance for the model allowing all parameters to vary randomly according to their posterior distributions. Let φ denote a vector of model parameters deemed focal (i.e., parameters that are the purpose of the analysis), as distinct from other nuisance parameters. Finally, let $D(\overline{\varphi})$ denote the mean deviance for the model holding focal parameters at their posterior means while allowing all other parameters to vary randomly. Then the DIC is (Spiegelhalter et al. 2002)

$$DIC = 2\overline{D} - D(\overline{\varphi}) . \tag{2-15}$$

Generally, DIC is an adaptation of Akaike's information criterion (AIC) to a Bayesian context.

AIC (Akiake's Information Criteria) is defined as the deviance of a model plus two times the number of parameters estimated. It is designed to identify the "best" model, where best is taken to be the model that has the highest likelihood for additional data generated by the same process. Since AIC was developed in a frequentist and nonhierarchical modeling context, all parameters are estimated as fixed effects. The near equivalent in a Bayesian context is a model in which all parameters are given uninformative priors. For such a model, AIC and DIC agree. However, most Bayesian models are hierarchical. That is, some parameters (such as the means of observed responses) are given distributions that are not uninformative and in fact are often functions of other, higher-level, parameters which also have distributions. In such cases, direct application of AIC is insufficient for two reasons. First, the number of parameters being estimated is unknown *ex ante* because the prior distributions for parameters introduce information sharing (and hence dependence) in their estimates. Thus the effective number of parameters being estimated is less than the number of parameters that appear in the model. Spiegelhalter et al. solve this problem using (2-15). Second, the criterion of best model is no longer clear, because a model's ability to predict depends upon which parameter estimates are held constant when making these predictions. In other words, models often contain nuisance parameters in addition to the parameters central to the purpose of the analysis (*focal parameters*). Spiegelhalter et al. recommend that the second term in (2-15) be calculated with φ including only focal parameters that are constant. If all focal parameters are given uninformative prior distributions, then they may be simply counted and AIC can be used instead of (2-15).

A popular alternative to AIC and DIC in a non-Bayesian context is BIC (Schwarz, 1978), and in a Bayesian context is the mean posterior of the likelihood of the data holding no parameters constant (the Bayes factors approach). Under certain conditions, these criteria will identify the true model from those estimated with probability one as the sample size grows to infinity, and no modeling purpose need be specified. However, DIC avoids assuming that one of the models considered is true, and instead chooses which model is best for a specific modeling purpose. Also, unlike BIC and Bayes factors, DIC is not subject to Lindley's paradox, so it can be used to compare models that make some use of weakly informative priors.

My purpose in modeling competitive market structure is to estimate all parameters characterizing the competitive map together with any brand-specific parameters in (2-1) that the researchers may have chosen to estimate independently of the competitive map (such as brand-specific intercepts). Thus, for my analysis these parameters are the focal ones. Calculating DIC using (2-15) requires first estimating all parameters, and then rerunning the model holding focal parameters at their posterior means (and letting nuisance parameters vary randomly). The DIC statistic produced by WinBUGS (Lunn et al. 2000) in the course of estimation of all parameters does not suffice. The program does not know the modeling purpose. Instead, it guesses that the purpose is to estimate the means of the observed responses, i.e., E(q_{it}) for (2-1). I accordingly had to calculate DIC by hand (with a routine I wrote) using (2-15).

2.3.2 Model Identification

As with all MDS procedures, my model requires identification conditions in order to obtain unique estimates. From (2-5), I see that $\Theta = [\theta_{.1}, ...\theta_{.m}, ..., \theta_{.M}]$ enters into the model mainly through the distance d_{ij} . However, inter-brand distance calculations are invariant to orthogonal, scale-preserving rotations of the brand locations given in $\mathbf{\Theta}$ and to mirror-image "flips" of the map. Therefore, some constraints must be imposed in order for the model to be identified. In particular, for the matrix of parameter locations $\mathbf{\Theta}$, I fix the elements above the diagonal to be zero and the diagonal elements to be positive. When M=2, this involves imposing the constraints $\theta_{12} = 0,\ \theta_{11} > 0,$ and $\theta_{22} > 0$.[20] Generally for any dimensionality *M*, this involves imposing *M(M+1)/2* constraints. In particular, for M>2, $\theta_{jm} = 0, j = 1, ...M-1; m = j+1, ..., M,$ and $\theta_{mm} > 0$. In addition, the indeterminacy of the origin is resolved by requiring that $\sum_{i=1}^{M} \theta_{im} = 0, m = 1, 2.$

[20] Occasionally, instead of imposing a constraint of the form $\theta_{mm} > 0$, it is more convenient to impose a constraint that $\theta_{im} > \theta_{jm}$ for two arbitrarily selected brands i and j . This equally well can avoid lack of model identification due to mirror-image "flips" of the map.

2.3.3 Model Estimation

I adopt a hierarchical Bayes modeling approach which I estimate using a Markov Chain Monte Carlo (MCMC) method. The hierarchical Bayes method is a powerful tool that recognizes that a family of parameters may be theoretically related and that can express such relationships with rich statistical formulations (Gelman and Hill, 2006). This section demonstrates the approach for the particular focal specification described in Table 1 by Components 1b, 2(M=2), 3d, 4c, 5b – analogous development applies for other specifications from Table 1.

The hierarchical Bayes version of the NBD model for my unit sales data is as follows. Using Δ to represent all parameters, I can write the posterior distribution as

$$\Pr(\Delta, \rho, \{\eta_{it}\}, \{\lambda_{it}\} \mid \{q_{it}, p_{\cdot t}, t\}) \propto \prod_t \prod_i \Pr(q_{it} \mid \lambda_{it}) \Pr(\lambda_{it} \mid \eta_{it}, \rho) \Pr(\eta_{it} \mid \Delta, \{p_{\cdot t}, t\}) \pi(\rho) \pi(\Delta) \quad , \qquad (2\text{-}16)$$

with $\Pr(q_{it} \mid \lambda_{it})$ from (2-12); $\Pr(\lambda_{it} \mid \eta_{it}, \rho)$ from (2-13), and $\Pr(\eta_{it} \mid \{p_{\cdot t}, t\})$ constituting the conditional distribution of the scale parameter for the gamma distribution of λ_{it}. $\pi(\rho), \pi(\Delta)$ are prior distributions for the shape parameter, as well as the demand model parameters. I specify an uninformative gamma prior for $\rho \sim dgamma(0.01, 0.01)$. The heart of the model concerns my formulation of the conditional distribution $\Pr(\eta_{it} \mid \{p_{\cdot t}, t\})$ and the prior distribution for Δ.

It is worth mentioning that the multiplication of likelihood functions in (2-16) assumes demands are independent over time. This assumption may be violated when demands are influenced by factors such as inventory effect.

Ignoring the possible existence of autocorrelation would lead to inconsistent estimates of price elasticities, which, under the context of current paper, implies that the estimated market structure map will be biased too. Therefore, although assuming independence over time simplifies the estimation problem and helps me focus attention on market structure related concepts, it certainly constitutes a potential limitation of current research. How to combine the time-series techniques and my proposed model is a topic that merits future attention.

Hard constraints. I begin my characterization of the prior distribution of η_{it} under the focal specification by recognizing a set of conditional relationships arising from the applicable model equations, as follows.

Conditional Relationship	**Applicable Equations**	
$\eta_{it} \mid \{p_{.t}, t\}, \alpha_i, \beta_{ii}, \{\beta_{i.}\}, \alpha_0, \gamma_1, \gamma_2, \gamma_3$;	(2-1) & (2-14)	
$\beta_{ij} \mid \theta_{i1}, \theta_{j1}, \theta_{i2}, \theta_{j2}, x_i, x_j, \phi_{CPE}$, $j \neq i$;	(2-4)	
$x_i \mid \theta_{i1}, \theta_{i2}, \upsilon_{11}, \upsilon_{12}$;	(2-6’)	(2-17)
$\beta_{ii} \mid \theta_{i1}, \theta_{i2}, \upsilon_{21}, \upsilon_{22}, \phi_{OPE}$; and	(2-7’)	
$\alpha_i \mid \theta_{i1}, \omega_0$.	(2-10)	

For example, Equation (2-1) establishes the relationship between the standard demand parameters (together with price data) and log($E(q_{it})$), which, in turn, is related to η_{it}, once I recognize that $E(q_{it}) = \rho/\eta_{it}$, under the NBD model (of Equation (2-14)). Similarly, Equations (2-4), (2-6’), (2-7’) and (2-10) link the CPEs, power parameters, OPEs, and demand intercepts, respectively, to the underlying locations $(\theta_{i1}, \theta_{i2})$ and $(\theta_{j1}, \theta_{j2})$ of brands *i* and *j*. Recognizing these hard constraints, I complete the characterization of the distribution of η_{it} by specifying priors for the constant coefficients and the brand-specific coefficients in the model.

Priors for brand-invariant coefficients. I assume weakly informative priors for all brand-invariant coefficients not subscripted by a particular brand) in the model (e.g., $\alpha_0, \phi_{CPE}, \phi_{OPE}, \upsilon_{11}, \upsilon_{12}, \upsilon_{21}, \upsilon_{22}, \omega_0, \gamma_1, \gamma_2, \gamma_3$). For example, I assume that $\alpha_0 \sim N(0,10^4)$.[21]

Priors for brand-specific coefficients. I use a hierarchical Bayesian structure to describe the brand-specific parameters not predetermined by the assumed hard constraints (in (2-17)). I describe my approach for the estimation of brand locations θ_{im} ($i = 1, 2..., n$, $m = 1,..., M$). The basic idea is to recognize that the brand locations θ_{im}, $i = 1, 2..., n$, in a given dimension *m*, are related; and these locations may be reasonably thought of as having been drawn from a common underlying distribution. This allows for information sharing across brands (because the hyper-parameters of the common underlying distribution are jointly estimated with the estimates of the locations θ_{im}, $i = 1, 2..., n$, themselves). Since the level of information sharing is determined by data (with the common distribution itself is estimated), this approach is referred to as *adaptive Bayesian shrinkage.*[22]

---

[21] Although not necessary for the focal model (Components 1b, 2(M=2),3d,4c,5b in Table 1), some other model formulations require imposing suitable sign constraints on the coefficients.
[22] For the focal specification of this section, the only brand-specific coefficients not predetermined

Thus, I do not estimate the brand locations using informative priors, or using weakly informative priors, as in (2-18),

$$\theta_{im} \sim N(0,10^4),\ i = 1,2...,n. \tag{2-18}$$

I, instead, estimate brand locations using a prior with hyper-parameters that are, themselves, given weakly informative priors on brand locations, as in (2-19):

$$\begin{aligned} &\theta_{im} \sim N(\mu_m, \sigma_m^2) \\ &\mu_m \sim N(0,10^4) \quad .\, i = 1,2...,n \\ &\sigma_m \sim Unif(0,50) \end{aligned} \tag{2-19}$$

Following Gelman and Hill (2006), a broad uniform prior is used for the standard deviation of θ_{im}.

The distributions in (2-19) imply a prior distribution for θ_{im} that is virtually identical to (2-18). (The mean is the same and the variance is one percent larger.) However, (2-18) asserts the brand locations have nothing in common — that they are independently distributed. That is, it assumes that knowing the θ_{im} values for any $n-1$ of the brands provides no information about the value for the remaining brand. Equation (2-19) makes no such assumption. Rather than assert, as in (2-18), that $\mu_m = 0$ and $\sigma_m = 100$, (2-19) estimates these values. Although the values of these two parameters assumed by (2-18) both lie well within the prior distributions given by these hyper-parameters in (2-19), comparison of these assumed values to their posterior densities invariably shows that the assumed

by the hard constraints are the brand locations θ_{im} — all other brand-specific coefficients are conditional on the brand locations θ_{im} (or on constant coefficients) through the constraints in (2-17). Thus, for the focal specification of this section, adaptive Bayesian shrinkage is used only for the brand locations θ_{im}. For other specifications in Table 2-1, however, hard constraints do not necessarily predetermine OPEs, brand intercepts, or the power parameters, x_i. When this is the case, adaptive Bayesian shrinkage is applied for estimating such families of coefficients.

value $\sigma_m = 100$ in (2-18) lies well outside its 95% credible intervals.

The advantage of using (2-19) is that it allows for shrinkage of the brand parameters towards a common mean μ_m estimated by the data whenever $\sigma_m \ll 50$, and since the common mean μ_m is given a weakly informative prior, its estimate is determined by the data. However, if the θ_{im} values are greatly different from each other, then the posterior estimate of σ_m can be much larger than 100, which implies a prior for θ_{im} that is even less informative than (2-18).

Of course, many Bayesian models in marketing (and elsewhere) estimate distributions for model parameters (Rossi and Allenby 2003), but they generally do so when those model parameters are viewed as random effects. In such models, the parameters of the underlying distribution (μ_m and σ_m) are regarded as focal. When the purpose, instead, is to estimate those model parameters themselves (i.e., when the θ_{im} values for the $i = 1, 2..., n$ brands are focal), then a weakly informative prior such as (2-18) has been used, mirroring a fixed-effects specification. In such cases, I suggest that a prior for parameters should be estimated (as in (2-19)). This is the position taken by Gelman and Hill (2006, p. 246), who argue that the latter approach should be taken, regardless of the focus. However, two conditions should be met: (a) weakly informative distributions should be used for the prior distribution's hyper-parameters. This ensures that the estimated prior subsumes both the fixed-effect and random-effect specifications. (b) There must be at least two (latent or observed) variables that share this prior distribution. Otherwise there is no opportunity for sharing information among parameters, and (2-19) is simply an inefficient means for implementing (2-18).

I note that the advantages of an "adaptive shrinkage" approach for the estimation of demand models such as (2-1) have been demonstrated by Montgomery and Rossi (1999). Their implementation differs from my in several respects, however. First, they employ adaptive shrinkage for some parameters but not others, whereas I apply it consistently to all parameters that satisfy the two criteria (a) and (b) above. Second, they make use of informative priors at the highest level of the hierarchy, whereas I use pre-specified weakly informative priors at the highest level throughout. Third, they use strong additive utility as their theoretical model wherein all competitive effects are through the income effect only. This precludes differential competition among brands. The model here allows for some brands to compete more closely than others, and asymmetrically.

2.4 Applications

I now apply the model and estimation methods to two very different data sets: one is informative about brand-specific parameters, and the other has price histories that are perfectly collinear in thirteen of twenty-one dimensions.

2.4.1. Beer Market

I estimate my model using data from Information Resources consisting of aggregate weekly store-level scanner data for beer in a test market in the US from 1989 to 1996. The data set contains 365 weekly observations on unit sales, price, and other variables. I considered the top five brands: Budweiser, Miller, Busch, Old Milwaukee and Milwaukee's Best. Price is the inflation-adjusted average weekly price per ounce.

I used the same brand-level data that formed the basis for Srinivasan, et al.'s (2000) analysis. Those authors first created brand-level variables from SKU-level variables.[23] This aggregation from SKUs (stock keeping units) to brands may be subject to aggregation bias due to possible heterogeneity among promotional variables (Christen et al., 1997; Pesaran and Smith, 1995). Srinivasan et al. (2000) control for this bias by performing pooling tests to determine whether it was reasonable to pool the different varieties for a brand; over 95% of sales could be pooled.[24] They also found that this data set produced elasticities that matched prior estimates (e.g., Tellis, 1988), which also alleviated this concern to some extent. The benefit of aggregating from SKUs to brands is that it avoids problems associated with colinearity and a large state space, pointed out by Kopalle et al. (1999) and Bucklin and Gupta (1999). Indeed, my analysis of the second soft-drink data set will illustrates how severe colinearity can make estimation problematic, and we show how our modeling framework can assist with estimation. We begin with the less challenging beer data set.

Table 2-2 below reports some descriptive statistics for the data:

---

[23] I am indebted to Information Resources and the authors of this article for making these aggregated brand-level data available.

[24] Christen et al. (1997) point out that the aggregation bias is likely to be quite small in data characterized by three conditions: frequent promotions, frequent price cuts, and small own price elasticities. We note that the first two conditions are met for the beer data.

Table 2-2 Summary Measures for Bear Drinks

	Budweiser	Miller	Busch	Old Milwaukee	Milwaukee's Best.
Mean Quantity	1097	589.5	1170	575.8	1404
Mean Price	10.61	7.64	9.78	6.41	7.47
S. D. of Price	0.88	0.81	1.17	0.54	0.70
Minimum of Price	7.33	5.15	7.39	4.65	5.95
Maximum of Price	12.56	9.9	12.31	7.71	9.49
Number of Price Changes	218	216	233	170	205

Model Selection. My model selection process (described earlier) identified the following three models, which I will compare: [25]

Model 1. This comprises the NBD model of (2-14), together with the following demand equation (arrived at by combining (2-1) and (2-11)):

$$\log(1/\eta_{it}) = \alpha_0 + \alpha_i + \beta_{ii}\log(p_{it}) + \sum_{j\neq i}\beta_{ij}\log(p_{it}) + \gamma_1 t' + \gamma_2 \sin\frac{2\pi t'}{52} + \gamma_3 \cos\frac{2\pi t'}{52}, \quad \text{(2-1')}$$

where $i = 1,\ldots,5$; $t = 1,\ldots,365$; $t' = t - \overline{t}$ (mean-centered); and prices are inflation-adjusted and also mean-centered across time and brands. This is a standard demand model in the literature. I estimate this model by imposing weakly informative priors for all demand parameters, which therefore leads to

[25] In this application, Model 1 is the fully saturated model, consisting of model components (1, 2, 3a, 4a, 5a) in Table 2. Model 2 consists of model components (1, 2, 3d, 4a, 5a) in Table 2. Model 3 consists of model components (1, 2, 3d, 4c, 5b).

estimates that are equivalent to the ML estimators. Notice that in the above equation, I drop the shape parameter ρ to add simplicity, since it is treated as a fixed effect and only influences the constant term on the RHS.

Model 2. This comprises Model 1 together with the additional elements:

$$\beta_{ij} = \phi_{CPE} - d_{ij} - x_i + x_j\,,\ i \neq j \tag{2-4}$$

$$d_{ij} = \sqrt{\sum\nolimits_{m=1}^{M} (\theta_{im} - \theta_{jm})^2} \tag{2-5}$$

$$x_i = \sum_{m=1}^{M} \upsilon_{1m} \cdot \theta_{im} \tag{2-6'}$$

In this model, market structure is used to explain (only) the symmetric and asymmetric competitive patterns in CPEs. The adaptive Bayesian shrinkage approach is applied when estimating brands locations, OPEs, and brand intercepts.

Model 3. This comprises Model 2 together with the following additional elements:

$$-\beta_{ii} = \phi_{OPE} + \sum_{m=1}^{M} \upsilon_{2m} \cdot \theta_{im}, \tag{2-7'}$$

$$\alpha_i = \omega_0 \theta_{i1}\,, \omega_0 \neq 0 \tag{2-10}$$

For this model, the estimates of brand locations, θ_{im}, underlie all brand-related parameters in the demand function. The adaptive Bayesian shrinkage approach is applied when estimating the brand locations. Conceptually, I find this to be the most desirable model.

Tables 2.2 and 2.3 show estimates using WinBugs 1.4.3 of these three models. Model 1 yields DIC of 17407. Models 2 and 3 have DIC of 17403 and 17405, respectively, which suggests that two dimensional models work better.

Table 2-3 Parameter Estimates for Beer Data

		Model 1	Model 2 (*M*=2)	Model 3: (*M*=2)
Intercepts	Budweiser	**1.28***	**1.36***	1.33*
	Miller	**-0.88***	**-0.96***	-0.93*
	Busch	**0.82***	**0.81***	0.93*
	Old Milwaukee	**-1.31***	**-1.35***	-1.35*
	Milwaukee's Best	**0.10**	**0.13***	0.02*
	Weight ω_3			**0.07***
Own-Price (OPEs)**	Budweiser	**-4.33***	**-4.25***	-4.28*
	Miller	**-4.80***	**-5.14***	-5.21*
	Busch	**-3.19***	**-3.36***	-3.72*
	Old Milwaukee	**-3.43***	**-3.53***	-3.48*
	Milwaukee's Best	**-3.91***	**-3.89***	-3.67*
	Weight 1 υ_{21}			**1.08***
	Weight 2 υ_{22}			**2.16***
Cross-Price (CPEs)**	β_{ij} [$i = Budweiser, j = Miller$]	**0.63***	0.50*	0.40*
	β_{ij} [$i = Budweiser, j = Busch$]	**0.19**	0.08	0.16*
	β_{ij} [$i = Budweiser, j = Old\ Milwaukee$]	**0.03**	-0.04	-0.09
	β_{ij} [$i = Budweiser, j = Milwaukee's\ Best$]	**-0.21**	-0.02	0.12
	β_{ij} [$i = Miller, j = Budweiser$]	**0.27**	-0.21	-0.48*
	β_{ij} [$i = Miller, j = Busch$]	**-2.05***	-1.07*	-0.92*
	β_{ij} [$i = Miller, j = Old\ Milwaukee$]	**-0.77***	-0.89*	-0.91*
	β_{ij} [$i = Miller, j = Milwaukee's\ Best$]	**-0.27***	-0.90*	-0.86*
	β_{ij} [$i = Busch, j = Budweiser$]	**0.07**	0.60*	0.65*
	β_{ij} [$i = Busch, j = Miller$]	**0.48***	0.17	0.44*
	β_{ij} [$i = Busch, j = Old\ Milwaukee$]	**0.08**	0.13	0.25*
	β_{ij} [$i = Busch, j = Milwaukee's\ Best$]	**-0.07**	0.16	0.49*
	β_{ij} [$i = Old\ Milwaukee, j = Budweiser$]	**-0.06**	0.71*	0.57*
	β_{ij} [$i = Old\ Milwaukee, j = Miller$]	**0.36***	0.58*	0.62*
	β_{ij} [$i = Old\ Milwaukee, j = Busch$]	**0.90***	0.36*	0.42*
	β_{ij} [$i = Old\ Milwaukee, j = Milwaukee's\ Best$]	**0.93***	0.85*	0.53*
	β_{ij} [$i = Milwaukee's\ Best, j = Budweiser$]	**0.04**	0.74*	0.64*
	β_{ij} [$i = Milwaukee's\ Best, j = Miller$]	**0.27***	0.58*	0.53*
	β_{ij} [$i = Milwaukee's\ Best, j = Busch$]	**1.18***	0.39*	0.52*
	β_{ij} [$i = Milwaukee's\ Best, j = Old\ Milwaukee$]	**0.99***	0.85*	0.39*
Brand Power	Budweiser		0.13*	0.08*
	Miller		0.49*	0.52*
	Busch		-0.13*	-0.17*
	Old Milwaukee		-0.24*	-0.24*
	Milwaukee's Best		-0.25*	-0.18*
	Asymmetry Weight 1 [υ_{11}]		**0.56***	**0.41***
	Asymmetry Weight 2 [υ_{12}]		**0.21**	**0.96***
Coordinates	Budweiser		**0.25***	**0.193***
	Miller		**0.31**	**-0.136***
	Busch		**0.09**	**0.135***
	Old Milwaukee		**-0.34***	**-0.196***
	Milwaukee's Best		**-0.32***	**0.003**
Coordinates	Budweiser		0***	0***
	Miller		**0.68***	**0.61***
	Busch		**-0.52***	**-0.24***
	Old Milwaukee		**-0.07**	**-0.18***
	Milwaukee's Best		**-0.09**	**-0.19***
Number of	**(in bold above)**	**30**	**21**	**14**

* The 95% credible interval excludes zero.

** The CPEs and Brand Power parameters in the last two columns and the OPEs and Intercepts in the last column are derived from the related underlying parameters (shown below them in bold in this table).

***Identification constraint.

Table 2-4 Other Parameter Estimates for the Beer Data

		Model 1	Model 2 (M=2)	Model 3 (M=2)
Covariates	γ_1	0.004	0.006	-0.002
	γ_2	0.006	0.003	0.008
	γ_3	-0.09*	-0.09*	-0.095*
Shrinkage Parameters (a)	σ_1		1.42 (1.18)	0.1834 (0.1275)
	σ_2		0.95 (2.04)	0.8191 (1.043)
NBD Model parameter	ρ	22.72*	20.96*	20.47*
DIC		**17407**	**17403**	**17405**

(a) σ_1 and σ_2 are the standard deviations of $\theta_{.1}$ and $\theta_{.2}$, respectively. These parameters are constrained to be nonnegative. We report (in parentheses) the posterior standard deviations of these estimates.

It is worth mentioning that the beer data that I obtained for analysis already aggregated all SKUs for the same manufacturer with correlated price history. This was done by Srinivasan, et al.'s (2000) to permit estimation by standard models. Since some structural information is lost during this aggregation, it is not a surprise to see that DIC for second model is slightly, although not significantly better than Model 3.

Overall, I observe that DIC favors both structural models (Models 2 and 3) over Model 1. [26] A close examination of Models 2 and 3 shows that most of the underlying structural parameters are estimated significantly (zero being outside their 95% credible intervals). Also, all own price elasticities are negative and in ranges similar to previous analyses (Tellis 1988). I further note evidence of seasonality, which is to be expected for beer consumption.

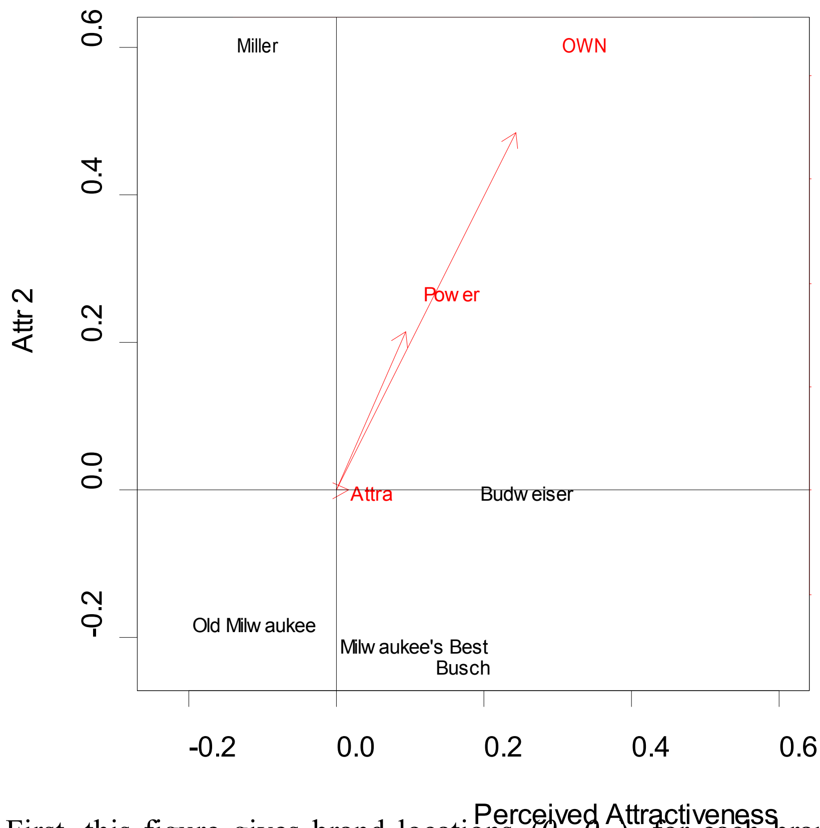
[26] As a rule of thumb provided by Spiegelhalter et al. (2002), models with DIC difference within the 1-2 DIC units of the "best" model are almost as well supported as the best model. Those within 3-7 units of the "best" model are substantially less well supported, and those more than 7 units worse than the "best" model have essentially no support.

I emphasize that even when I impose underlying structure for the CPEs, OPEs, and intercepts, my methodology also recovers implied estimates for the CPEs, OPEs, and intercepts. That I get these implied estimates as a byproduct of the estimation procedure constitutes one of the strengths of MCMC Bayesian methodology. I observe that the estimated values of the intercepts and OPEs are generally robust across Models 1, 2, and 3, even though the later two models impose substantial underlying structure. Furthermore, that the intercepts are as similar in Model 3 as with the other two models, suggests that interpreting axis one of the underlying Euclidean map as describing the intercept does not undermine my estimation. Lastly, many of the CPEs are robust across the three models, and where there are very large changes, the CPE in Model 1 was not significant to begin with. So imposing structure appears to be improving the accuracy of estimation of the cross-price elasticities. Moreover, Models 2 and 3 are very similar in estimates, and imposing structure on the intercepts and OPEs does not appear to change the underlying structure for the CPEs (both the symmetric structure and skew-symmetric brand power estimates). These are reassuring checks. Furthermore, notice that the hyperparameter estimates of brands location variances on each dimension, σ_1 and $\sigma_{\theta.2}$, are small, which indicates that shrinkage plays a role in estimating the brands' locations (Table2.4).

Managerial Implications. Examining the underlying structure, I note that the OPEs are all significantly negative and reasonable, and the CPEs evidence various cross-price effects, some symmetric and others asymmetric. For example, Old Milwaukee and Milwaukee's Best are close symmetric substitutes with each other. On the other hand, Budweiser and Miller are more characterized by one-directional substitutability, where the price of Miller has significant effects on Budweiser's sales, but the price of Budweiser has less of an effect on Miller's sales.

As Model 3 is conceptually my most desirable model (I do not try more than two dimensions, as with only five brands, dimensions more than two can be difficult to reliably estimate), I provide a biplot based on this model. Figure 3 displays the five brand locations and three variables (own-price elasticity, brand power, perceived attractiveness) on two latent attribute dimensions (Figure 2-3). Here, brands are displayed as points while the three variables are displayed as arrows. The horizontal dimension is interpreted as "perceived attractiveness" in accordance with specification component 5b (of Table 1). Several conclusions can be drawn from this map.

Figure 2-3. Plot of Beer Brand Locations in Model 3



First, this figure gives brand locations $(\theta_{i1}, \theta_{i2})$ for each brand *i*. Recall that the symmetric part of the cross-elasticity between any two products is modeled by the Euclidean distance in the map between the brands. Thus Old Milwaukee and Milwaukee's Best are quite similar, while Miller is distinct from all other brands.

Second, the angle between arrows approximates the correlation between the variables (i.e., the smaller the angle, the higher the correlation). I can see from Figure 3 that "brand power" is highly correlated with "own-price elasticities," and both are positively correlated with "perceived attractiveness". This means that brands with more power have less elastic OPEs and are more attractive, which, as expected, reflects Proposition 2. It is also worth noticing two other findings. (a) Since the angle between a vector and an axis indicates the importance of the contribution of the corresponding variable to the axis dimension, I see that the second dimension is more strongly related to brand power and OPEs. (b) Since the length of an arrow is equal to the variance of the corresponding variable, it seems that the "perceived attractiveness" dimension is not that important for these data.

Third, the points (brands) in the biplot can be projected perpendicular on the arrows, and the position of the points along the arrow gives information of the value of the brands on corresponding variables. As I expected, Miller has greater asymmetric dominance than Budweiser, which in turn has greater dominance than the other three brands. Notice the estimates of asymmetry Weights on both dimensions in CPE are significantly different from zero (Table 2-4), which means that competitive asymmetry is an essential feature of this product category. Similarly, Miller has the greatest OPE in absolute value, followed by Budweiser, and then the other three brands. And finally, the brand intercepts (attractiveness), which are proportional to the horizontal axis, Budweiser is represented as the most popular (or mass marketed) brand, along with Busch.

Overall, Figure 2-3 presents a picture of quality tiers. Budweiser appears to be popular, attracting a broad market, but Miller is a beer with greater asymmetric dominance over Budweiser even if it has a smaller market share. There is, thus, a smaller quality market, of which Miller is one brand. Budweiser is at the center. Among the lower dominance segment, consisting of Busch, Milwaukee's Best, and Old Milwaukee, Busch has the highest share. This is intuitive, and perhaps sums up insights that resonate with beer drinkers.

2.4.2 Soft-drink Market

I now turn to analysis of store level aggregate data for carbonated soft drinks (CSD). My selection of context was motivated in part by proposed soft-drink mergers in 1986 when the Federal Trade Commission deterred PepsiCo's proposed acquisition of Seven-Up Co. by threatening antitrust litigation. Unable to similarly dissuade Coca-Cola Co., the FTC went to court to block Coca-Cola Co.'s proposed acquisition of Dr Pepper (*FTC v. Coca-Cola Co.* 641 F. Supp. 1128; D.D.C. 1936). The FTC justified these actions by citing Section 7 of the Clayton Act, claiming in both antitrust issues, that the acquirer and acquiree were selling in the same line of commerce, defined in these cases as "the national carbonated soft drink market." I wish to answer a simple yet important question: To what extent do the brands carried by the four soft-drink manufacturers compete with each other?

Data Description. The data, recorded using the scanner UPC (Universal Product Code) system from a St. Louis supermarket, describes weekly quantity sales and price of the same two liter soft drinks analyzed in the survey and covers 63 weeks commencing 2/22/88. There are in total 48 brands included in our dataset. We focus our analysis on the following 20 brands of four major manufacturers:

PepsiCo: Pepsi, Diet Pepsi, Pepsi Free, Diet Pepsi Free, Mountain Dew, Lemon Lime Slice, Diet Lemon Lime Slice

Coca-Cola Co.: Coke, Diet Coke, Caffeine-Free Coke, Caffeine-Free Diet, Coke, Cherry Coke, Diet Cherry Coke, Sprite, Diet Sprite, MR. PiBB

Dr Pepper Co.: Dr Pepper, Diet Dr Pepper

Seven-Up Co.: 7-Up, Diet 7-Up

The remaining low sales brands (all shares less than 1.5%) were combined into a single "other" brand with price calculated as the average (quantity Weighted) price of the constituent products. A preliminary examination of the data shows that the average Weekly sales ranged from 838 for Pepsi to 5 for Mr.PiBB.

Table 2-5 below reports some descriptive statistics for the data:

Table 2-5 Summary Measures for Soft Drinks Data

	Pepsi	Diet Pepsi	Mount, Dew	Pepsi Free	Diet Pepsi Free	Lemon Lime Slice	Diet Lemon Lime Slice
Mean Quantity	838	425	72	215	347	54	72
Mean Price	1.123	1.124	1.124	1.123	1.124	1.126	1.121
S. D. of Price	0.232	0.232	0.232	0.232	0.232	0.253	0.253
Minimum of Price	0. 71	0. 7	0. 71	0. 7	0. 71	0. 69	0. 69
Maximum of Price	1. 49	1. 49	1. 49	1. 49	1. 49	1. 49	1. 49
Number of Price Changes	9	8	8	8	8	8	7
	Cherry Coke	**Diet Cherry Coke**	**Coke**	**Diet Coke**	**Caff. Free Coke**	**Caff. Free Diet Coke**	**Sprite**
Mean Quantity	47	63	770	453	52	425	158

Mean Price	1.170	1.170	1.170	1.170	1.170	1.170	1.088
S. D. of Price	0.240	0.240	0.240	0.240	0.240	0.240	0.234
Minimum of Price	0. 69	0. 69	0. 7	0. 7	0.69	0.7	0.69
Maximum of Price	1. 49	1. 49	1. 49	1. 49	1.49	1.49	1.49
Number of Price Changes	10	10	10	10	11	10	10

	Diet Sprite	**Mr. PiBB**	**Dr. Pepper**	**Diet Dr. Pepper**	**7-Up**	**Diet 7-Up**	**Other**
Mean Quantity	144	5	61	35	104	61	521
Mean Price	1.088	1.426	1.420	1.420	1.388	1.388	0.986
S. D. of Price	0.233	0.118	0.219	0.219	0.232	0.232	0.124
Minimum of Price	0.69	0.98	0.98	0.98	0.88	0.88	0.44
Maximum of Price	1.49	1.49	1.69	1.69	1.69	1.69	1.249
Number of Price Changes	10	3	8	8	9	9	63

An important feature of my data is the presence of *perfect* multicolinearity among the products within a line of products sold by each of the four major companies. As shown in Table 2-6, the prices of the 21 brands vary in only 8 dimensions! From a managerial perspective, these product lines are being priced as groups—with identical prices over the course of the year. This parallel pricing of products within product lines or groups presents a challenge for estimation. A second challenge for this data set is that I estimate locations of 21 brands (including "other).[27] I show below how the information sharing feature of my modeling framework can enable estimation of the underlying structure among the brands – despite the presence of severe multicolinearity and the large number of brands.

Model Estimation Results. In general, I recommend using the NBD model above. However, when the information in the data is limited, the NBD model can suffer from an over-fitting problem, whereby the resulting map describes random error or noise rather than the underlying relationship. Therefore, for the soft drink data, I adopt only the Poisson distribution (2-12) for the dependent variable. So, instead of (2-1') above, I use the analogous form

$$\log(\lambda_{it}) = \alpha_0 + \alpha_i + \beta_{ii}\log(p_{it}) + \sum_{j\neq i}\beta_{ij}\log(p_{it}) + \gamma_1 t' + \gamma_2 \sin\frac{2\pi t'}{52} + \gamma_3 \cos\frac{2\pi t'}{52}, \tag{2-1''}$$

where $\lambda_{it} \equiv E(q_{it})$; i, j = 1, …, 21; t = 1, …, 63; t' is mean-centered ($t' = t - \overline{t}$); and prices are mean-centered across time and brands.

Given the limitations of this data, the fully saturated model is not estimable, and I instead started by estimating a model that drops the cross-price elasticities ($\beta_{ij} = 0$ for all $i \neq j$) but retains the other terms in (2-1''). That is, this base model has components 1a, M = 0, 3a, 4a and 5a (see Table 2-1). The DIC for this model is 88059, with 42 parameters estimated. Many of the models for M = 1

[27] A managerially interesting feature is that the Coke and Pepsi lines are negatively correlated, apparently due to an implicit understanding between these companies that they offer alternative price promotions. In addition, the Sprite and Slice groups are closely correlated with the Pepsi and Coke groups, respectively, as I might expect.

or 2 and with partially saturated terms (Components 3b, 4a, or 5a) were not estimable, but the dominance-point formulations (3c and 4b) were estimable and worked better than the vector counterparts. The model with components 1a, 2 (with $M = 1$), 3c, 4b, and 5b has a deviance of 86568 and DIC of 86584, with 25 focal parameters estimated,[28] while the $M = 2$ version of this model has a deviance of 76320 and DIC of 76340, with 47 focal parameters estimated, which I report on below. Different from the beer data, the DIC difference here is much bigger, this is due to the Poisson distribution used here, which does not account for the over-dispersion effect.

Table 2-7 shows the estimated price elasticities from the two-dimension model. Own-price elasticities are all less than –1 and, consistent with previous literature on this industry (Dube 2004, 2005), own-price elasticities are very elastic, ranging from –2.5 to –5.5. In addition, the absolute values of cross-price elasticities are an order of magnitude smaller than the own-price elasticities, as expected and as found for the beer data. There are, however, many negative cross-price elasticities, which can arise when category expansion effects overwhelm substitution effects between brands (Russell et al. 2008). Both Diet 7Up and Diet Dr. Pepper have fairly large negative effects on the unit sales of many other brands, which seems to indicate that the price reduction of these two brands might benefit the whole category.

---

[28] The effective number of parameters estimated for both one-dimensional and two-dimensional models will be smaller, as Bayesian shrinkage imposes information-sharing across parameters.

Table 2-6 Soft Drink Price Correlations[29]

Manufacturers	Products	Pepsi	Diet Pepsi	Mount Dew	Pepsi Free	Diet Pepsi Free	Lemon Lime Slice	Diet Lemon Lime Slice	Cherry Coke	Diet Cherry Coke	Coke	Diet Coke	Caff.-Free Coke	Caff.-Free Diet Coke	Sprite	Diet Sprite	Mr. PiBB	Dr. Pepper	Diet Dr. Pepper	7-Up	Diet 7-Up	Other
Pepsi Co.	Pepsi	1	1	1	1	1	0.7	0.7	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2
	Diet Pepsi	1	1	1	1	1	0.7	0.7	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2
	Mountain Dew	1	1	1	1	1	0.7	0.7	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2
	Pepsi Free	1	1	1	1	1	0.7	0.7	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2
	Diet Pepsi Free	1	1	1	1	1	0.7	0.7	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4	-0.2	-0.2	-0.1	0.1	0.1	0.1	0.1	0.2
	Lemon-Lime Slice	0.7	0.7	0.7	0.7	0.7	1	1	-0.3	-0.3	-0.3	-0.3	-0.3	-0.3	0	0	-0.3	-0.2	-0.2	-0.1	-0.1	0.3
	Diet Lemon-Lime Slice	0.7	0.7	0.7	0.7	0.7	1	1	-0.2	-0.3	-0.3	-0.3	-0.3	-0.3	0	0	-0.2	-0.1	-0.1	-0.1	-0.1	0.3
Coca-Cola Co.	Cherry Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.2	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Diet Cherry Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.3	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.3	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Diet Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.3	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Caffeine-Free Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.3	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Caff-Free Diet Coke	-0.4	-0.4	-0.4	-0.4	-0.4	-0.3	-0.3	1	1	1	1	1	1	0.7	0.7	0.1	-0.1	-0.1	-0.2	-0.2	-0.2
	Sprite	-0.2	-0.2	-0.2	-0.2	-0.2	0	0	0.7	0.7	0.7	0.7	0.7	0.7	1	1	0	0	0	0	0	-0.3
	Diet Sprite	-0.2	-0.2	-0.2	-0.2	-0.2	0	0	0.7	0.7	0.7	0.7	0.7	0.7	1	1	0	0	0	0	0	-0.3
	Mr. PiBB	-0.1	-0.1	-0.1	-0.1	-0.1	-0.3	-0.2	0.1	0.1	0.1	0.1	0.1	0.1	0	0	1	0.2	0.2	0.1	0.1	0.2
Dr Pepper Co.	Dr. Pepper	0.1	0.1	0.1	0.1	0.1	-0.2	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0	0	0.2	1	1	0.9	0.9	-0.1
	Diet Dr. Pepper	0.1	0.1	0.1	0.1	0.1	-0.2	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0	0	0.2	1	1	0.9	0.9	-0.1
Seven-Up Co.	7-Up	0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0	0	0.1	0.9	0.9	1	1	0.1
	Diet 7-Up	0.1	0.1	0.1	0.1	0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	0	0	0.1	0.9	0.9	1	1	0.1
	Other	0.2	0.2	0.2	0.2	0.2	0.3	0.3	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.3	-0.3	0.2	-0.1	-0.1	0.1	0.1	1

[29] "Other" refers to the composite of private labels, generics, and lesser known brands. Note the cases of correlation of 1 are identical prices to within one cent (except for one case of a 3-cent price difference in week 55 for Caffeine-free Diet Coke).

Table 2-7 Implied Soft Drink Cross-Price Elasticities

	Pepsi	Diet Pepsi	Mountain Dew	Pepsi Free	Diet Pepsi Free	Lemon-Lime Slice	Diet Lemon-Lime Slice	Cherry Coke	Diet Cherry Coke	Coke	Diet Coke	Caffeine-Free Coke	Caff-Free Diet Coke	Sprite	Diet Sprite	Mr. PiBB	Dr. Pepper	Diet Dr. Pepper	7-Up	Diet 7-Up	Other
Pepsi	-3.91	-0.40	0.50	0.71	-0.57	0.17	-0.27	0.14	0.44	0.95	-0.43	-0.25	0.65	0.69	-0.89	0.02	0.67	-1.16	0.63	0.01	0.63
Diet Pepsi	-0.36	-4.09	0.03	-0.30	0.81	-0.64	0.55	0.50	-0.26	-0.38	0.91	0.58	-0.63	0.38	0.39	0.04	0.59	0.16	0.67	-1.20	0.56
Mountain Dew	0.12	-0.39	-3.77	0.46	-0.52	0.31	0.00	0.46	0.68	0.09	-0.44	0.01	0.05	0.87	-0.72	0.49	0.83	-0.99	0.67	-0.33	0.50
Pepsi Free	0.59	-0.46	0.71	-3.94	-0.62	0.43	-0.20	0.24	0.70	0.56	-0.51	-0.19	0.58	0.78	-0.88	0.22	0.73	-1.16	0.63	0.09	0.55
Diet Pepsi Free	-0.43	0.91	0.00	-0.35	-4.35	-0.67	0.60	0.51	-0.29	-0.45	0.87	0.63	-0.69	0.34	0.56	0.06	0.55	0.33	0.62	-1.24	0.50
Lemon-Lime Slice	0.28	-0.57	0.80	0.67	-0.70	-4.63	-0.16	0.30	0.88	0.26	-0.62	-0.16	0.37	0.71	-0.89	0.44	0.66	-1.16	0.51	0.24	0.36
Diet Lemon-Lime Slice	-0.43	0.35	0.22	-0.24	0.30	-0.43	-4.12	0.81	-0.08	-0.45	0.28	0.96	-0.62	0.49	0.25	0.40	0.68	-0.02	0.66	-1.09	0.45
Cherry Coke	-0.31	0.02	0.40	-0.08	-0.08	-0.25	0.52	-3.72	0.09	-0.33	-0.05	0.52	-0.47	0.63	-0.21	0.53	0.80	-0.48	0.72	-0.91	0.48
Diet Cherry Coke	0.27	-0.47	0.89	0.66	-0.61	0.60	-0.10	0.37	-4.09	0.24	-0.52	-0.08	0.28	0.81	-0.81	0.44	0.76	-1.09	0.61	-0.03	0.46
Coke	0.97	-0.40	0.49	0.71	-0.57	0.16	-0.28	0.13	0.43	-3.93	-0.43	-0.25	0.66	0.68	-0.89	0.01	0.66	-1.16	0.63	0.01	0.63
Diet Coke	-0.35	0.94	0.01	-0.31	0.80	-0.66	0.51	0.47	-0.28	-0.37	-4.13	0.54	-0.63	0.37	0.37	0.01	0.57	0.15	0.66	-1.20	0.56
Caffeine-Free Coke	-0.42	0.37	0.22	-0.23	0.32	-0.44	0.94	0.80	-0.09	-0.44	0.29	-4.08	-0.62	0.49	0.25	0.38	0.68	-0.02	0.67	-1.09	0.46
Caff-Free Diet Coke	0.80	-0.52	0.58	0.85	-0.68	0.40	-0.32	0.12	0.59	0.79	-0.56	-0.30	-4.33	0.67	-0.97	0.09	0.63	-1.25	0.55	0.34	0.51
Sprite	0.04	-0.30	0.60	0.26	-0.45	-0.05	-0.01	0.43	0.33	0.01	-0.35	0.01	-0.12	-3.21	-0.69	0.28	0.93	-0.96	0.78	-0.61	0.61
Diet Sprite	-0.56	0.68	-0.01	-0.43	0.74	-0.67	0.73	0.56	-0.31	-0.58	0.63	0.75	-0.79	0.29	-4.86	0.15	0.49	0.71	0.53	-1.30	0.36
Mr. PiBB	-0.22	-0.23	0.63	0.10	-0.33	0.09	0.31	0.73	0.37	-0.25	-0.30	0.31	-0.30	0.69	-0.42	-4.19	0.76	-0.68	0.61	-0.63	0.38
Dr. Pepper	-0.13	-0.25	0.42	0.06	-0.39	-0.25	0.04	0.45	0.13	-0.15	-0.30	0.06	-0.31	0.78	-0.63	0.20	-3.07	-0.91	0.81	-0.82	0.60
Diet Dr. Pepper	-0.63	0.65	-0.08	-0.50	0.72	-0.74	0.67	0.50	-0.39	-0.65	0.61	0.68	-0.86	0.22	0.91	0.09	0.42	-5.25	0.45	-1.38	0.30
7-Up	-0.11	-0.11	0.31	0.01	-0.26	-0.35	0.08	0.43	0.03	-0.13	-0.16	0.10	-0.34	0.69	-0.54	0.11	0.87	-0.82	-2.80	-0.88	0.74
Diet 7-Up	0.53	-0.71	0.58	0.74	-0.86	0.66	-0.40	0.06	0.66	0.52	-0.75	-0.39	0.72	0.56	-1.10	0.13	0.51	-1.38	0.39	-5.22	0.31
Other	0.06	-0.05	0.31	0.11	-0.22	-0.32	0.03	0.36	0.06	0.04	-0.09	0.06	-0.20	0.68	-0.54	0.05	0.82	-0.81	0.91	-0.79	-2.46

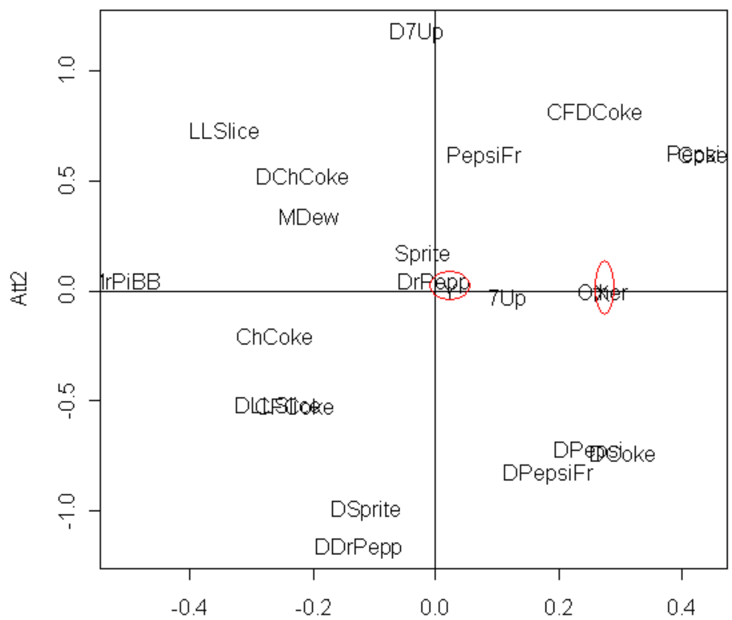
Table 2-8 provides the estimates of three Weights (w_1, w_2, w_3) associated with skew-symmetric component, own-price elasticities, and intercepts respectively—cf. (2-6), (2-7), and (2-10)—as well as their corresponding standard deviations. All of these estimates are statistically significant, and I can conclude that asymmetry is an essential feature for my cross-price elasticities matrix.

Table 2-8 Estimates of Weights for the Soft-drink Data

	w_1 Asymmetry Weight	w_2 Own-price Elasticity Weight	w_3 Intercept Weight
Posterior Mean	0.58	2.27	4.17
Posterior St. Dev.	0.01	0.07	0.18

Figure 2-4 gives the estimated two-dimensional map (please refer to Appendix F for the corresponding brand names to the abbreviations in the Figure), which displays the 21 brands' locations and the two hypothetical brands locations (red circles) on two latent dimensions (describing the Dominance Brand Y and the Least Vulnerable Brand Z, respectively). The horizontal dimension is restricted to be proportional to the brand intercepts. From Figure 2-4, I observe some interesting findings.

Figure 2-4 Two-Dimensional Map for Soft Drinks



First, the horizontal dimension distinguishes the coke vs. non-coke products, with majority of the coke products considered superior to non-coke products. Here I interpret "*perceived attractiveness*" as increasing as I move to the right on the horizontal axis.

Second, the distance between points reflects the perceived similarity between the corresponding two brands and the competition between them. Figure 2-4 indicates that Pepsi and Coke are in head-to-head competition in both dimensions. On the other hand, Diet Pepsi, Diet Coke and Diet Pepsi Free are all diet products and they are similar, although they seem to share a different submarket from Pepsi and Coke. Dr. Pepper products are in direct competition with Sprite products and 7UP, with Diet Dr.Pepper competes with Diet Spirit, while Dr.Pepper competes with both Sprite and 7UP. Furthermore, there seems to be a close relationship between Cherry Coke products and Lemon-Lime Slices products. However, the structure of this competition seems somewhat counterintuitive, as Cherry Coke is in direct competition with Diet Lemon-Lime Slice, while Diet Cherry Coke with Lemon-Lime Slice. A probable explanation might be that there exist features other than Diet vs. Non-Diet that play a role in this competitive pattern.

Third, concerning brand power and own-price elasticity, the power of a brand should be interpreted as related to the distance between a brand and the hypothetical brand labeled as “Y” (the left circle), where a brand’s power is larger, the closer the brand’s location is to “Y”. Similarly, the brand that is closest to the hypothetical brand labeled as “Z” (the right circle) in the map is considered the least sensitive to its own-price changes and also least vulnerable to other brands’ price changes. As I can see from the map, most of the Coke and Pepsi products have relatively lower vulnerability, measured as smaller absolute value of own-price elasticity. This effect might be due to the high market share of these products, and also the relatively sparse spatial density of these products. The sparse spatial density of these products can be attributed to the fact that the Coke and Pepsi products have been differentiated (e.g., diet vs. non-diet, caffeine vs. caffeine-free, etc) so well that finer submarkets are formed among them than other companies’ products. Besides, the low average prices of these products may also play a role. When prices are already low (average $1.1 per 2 liter), consumers’ sensitivity to these products’ price reductions would be very low. Now consider the power of all brands on the market. An interesting finding is that the brands with the highest power are 7UP, and Dr. Pepper, rather than Coke and Pepsi (because 7Up and Dr. Pepper are closest to the hypothetical Dominance Brand Y). Since both 7UP and Dr. Pepper are also quite near the least vulnerable brand Z, I can conclude that these two brands' high power is mainly due to their low vulnerability (rather than their high clout). One explanation for their low vulnerability can be that both 7UP and Dr. Pepper position their brands quite differently from the other brands in the market. In addition, these two brands are closer to non-coke products on the left panel of the Figure than those coke products on the right, but closer in "*perceived attractiveness*" to coke products. Pepsi and Coke, one the other hand, also have relatively low vulnerability, but their power (and their clout) appears less than that of 7Up and Dr. Pepper. On the other hand, both have very high brand intercepts (as shown by their position on the right on the horizontal axis).

Fourth, MrPiBB seems like a true outlier in the soft drink market. It does not exert influence on other brands, nor is it influenced by them. From the consumers' perspective, it has the lowest "*perceived attractiveness*" level but highest average price. This is a classic niche brand that enjoys a quasimonopoly, albeit with a very small segment.

Last, I derive each brand's average clout and vulnerability (Table 2-9). Recall the definition of clout and vulnerability are: $Clout_i = \sum_{j \neq i}^{J} \beta_{ji}$ and $Vul_i = \sum_{j \neq i}^{J} \beta_{ij}$.

The entries in the table are actually $\frac{Clout_i}{J}$ and $\frac{Vul_i}{J}$.

Table 2-9 Clout and Vulnerability

	Pepsi	Diet Pepsi	Mountain Dew	Pepsi Free	Diet Pepsi Free	Lemon-Lime Slice	Diet Lemon-Lime Slice	Cherry Coke	Diet Cherry Coke	Coke	
Vul/21	0.29	0.31	0.28	0.32	0.32	0.33	0.31	0.25	0.32	0.29	
Clout/21	0.17	0.15	0.54	0.33	0.06	0.09	0.35	0.58	0.37	0.15	

	Diet Coke	Caffeine-Free Coke	Caff-Free Diet Coke	Sprite	Diet Sprite	Mr. PiBB	Dr. Pepper	Diet Dr. Pepper	7-Up	Diet 7-Up	Other
Vul/21	0.30	0.31	0.32	0.19	0.32	0.28	0.12	0.30	0.12	0.29	0.14
Clout/21	0.11	0.37	0.05	0.72	-0.12	0.40	0.79	-0.36	0.74	-0.36	0.59

Antitrust Implications. The findings from the model provide us an interesting feature of the competition in late 80's soft drink market: products that belong to 7Up and Dr. Pepper companies compete with each other, but in quite a different market than products of both the Coca-Cola and Pepsi companies. A direct implication of this finding is that the approval of both PepsiCo's proposed acquisition of Seven-Up Co and Coca-Cola Co.'s proposed acquisition of Dr Pepper may *not* necessarily "substantially lessen competition, or to tend to create a monopoly" in the soft drink market. Actually, with both the economies of scale and economies of scope, consumers may even benefit from these acquisitions. This would not be true if the mergers were between Pepsi and Coca Cola, or between 7Up and Dr.Pepper, since in either case, the competition happens within the same strategic submarket.

It is worth mentioning that the outcome in this industry was that a third party, Hicks and Haas, had, by October of 1986, acquired Dr Pepper, A&W Root Beer, and 7 Up, making Hicks and Haas the third largest U.S. soft drink maker. My finding, however, provides contrary evidence that this outcome may not be better than permitting the previous two merger proposals to have become realized.

2.5 Conclusions

This essay presents a unified framework for estimating a market-structure map that (a) represents the perceived substitutability between brands; (b) accounts for cross-price elasticity asymmetries; (c) reasonably integrates common marketing specifications (e.g., vector and dominant point formulations) to represent the various demand-model components; and (d) facilitates the estimation of price elasticities in the presence of severe colinearity in prices and other data limitations. From a microeconomic perspective, I identify underlying relationships between measures of cross-price elasticity, brand power, vulnerability, clout, own-price elasticity, and spatial density. From a methodological perspective, I demonstrate an adaptive Bayesian approach to estimation that shares information across different brands and different terms in a set of demand equations.

This framework can be applied to inform marketing managers who are selecting prices for the brands in their existing product lines and who are setting the positioning of new brands. The framework is also relevant to policy makers applying antitrust policy relating to mergers and to monopolization. I apply this approach to beer and soft-drink data sets and arrive at plausible estimates in light of both economic theory and past marketing literature.

I acknowledge limitations of this study and directions for extension of this research. The model proposed in this paper assumes that a stationary market structure exists over time. And this assumption certainly makes the method more suitable for a mature product category. In particular, it would be desirable do the following: (a) add covariates in the model such as promotions and displays (and the lack of such information may create imprecision or bias in the estimation of such elements as brand intercepts); (b) analyze the time dimension more explicitly, including lagged effects; (c) consider alternative functional forms (other than (2-1) and (2-5)); and (d) develop flexible, utility-maximization-based models, capable of estimating market structure that could also be useful for simulating consumer welfare effects for policy purposes (Bronnenberg et al. 2005). Concerning this last point, such utility-based models would serve as a complement to the price-elasticity, market-demand-based approach of the current paper.

Overall, price elasticities provide important information about competition in a market, and the availability of detailed scanner data in nearly all retail outlets serves as a potential source of this information. I look forward to continued work on elasticity-based market structure analysis that more fully exploits the potentialities of such data.

Bibliography

Aaker, David A. (1991), *Managing Brand Equity*, San Francisco: Free Press.

Aaker, David A. (1996), “Measuring Brand Equity Across Products and Markets,” *California Management Review*, 38 (3), 102–120.

Akaike, Hirotugu (1974), "A new look at the statistical model identification," IEEE Transactions on Automatic Control 19 (6), 716–723.

Allenby, Greg M. (1989), “A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data,” *Marketing Science*, 8(3), 265–280.

Bentler, P.M. and David G. Weeks (1978), “Restricted Multidimensional Scaling Models,” *Journal of Mathematical Psychology* 17, 138–151.

Berry, Steven T., (1994), “Estimating Discrete–Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25 (2), 242–262.

Blattberg, Robert C. and George, Edward I. (1991), “Shrinkage Estimation of Price and Promotion Elasticities: Seemingly Unrelated Equations,” *Journal of the American Statistical Association*, 86(414), 304–215.

Blattberg, Robert C, and Ken J. Wisniewski (1989), “Price Induced Patterns of Competition,” *Marketing Science, 8 (4),* 291–309.

Borg, IngIr and Patrick J.F.Groenen (2005), *Modern Multidimensional Scaling: Theory and Applications*, Second Edition, Springer.

Bronnenberg B.J, and VanHonacker W.R (1996), “Limited Choice Sets, Local Price Response, and Implied Measures of Price Competition,” *Journal of Marketing Research*, 33(2), 163–173.

Bronnenberg, B.J., P.E.Rossi, N.J.Vilcassim (2005), “Structural Modeling and Policy Simulation,” *Journal of Marketing Research*, 42(1), 22–26.

Bucklin, R.E., Gupta, S. (1999), “Commercial use of UPC scanner data,” *Marketing Science,* 17(3), 247–273.

Butler, D. and B. Butler, Jr. (1970, 1971), Hendro–Dynamics: Fundamental Laws of Consumer Dynamics, Cambridge, Mass: The Hendry Corporation.

Carroll, J.D. (1976), “Spatial, Non–spatial and Hybrid Models for Scaling,” *Psychometrika*, 41, 439–463.

Chino, N. (1978), “A Graphical Technique for Representing the Asymmetric Relationships between N objects,” *Behaviormetrika*, 5, 23-40.

_____ (1990), “A Generalized Inner Product Model for the Analysis of Asymmetry,” *Behaviormetrika*, 27, 25-46.

Christen, M., Gupta, S., Porter, J.C., Staelin, R., Wittink, D.R., (1997), “Using market–level data to understand promotion effects in a nonlinear model,” *Journal of Marketing Research* 34, 322–334.

Cooper, Lee G. (1988), "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (6), 707–23.

Cooper, Lee G, Daniel Klapper, and Akihiro Inoue (1996), "Competitive–Component Analysis: A New Approach to Calibrating," *Journal of Marketing Research,* 33(2), 224–238.

Cowpertwait, Paul S.P., & Andrew V. Metcalfe (2009), *Introductory Time Series with R*, First Edition, Springer.

DeSarbo, Wayne S. and Rajdeep Grewal (2007), "An Alternative Efficient Representation of Demand–Based Competitive Asymmetry," *Strategic Management Journal*, 28 (7), 755–66.

DeSarbo, Wayne S., Rajdeep Grewal, and Jerry Wind (2006), "Who Competes with Whom? A Demand–Based Perspective for Identifying and Representing Asymmetric Competition," *Strategic Management Journal* (27), 101–129.

Dube, Jean–Pierre. (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science,* 23, 66–81.

Dube, Jean–Pierre. (2005), "Product Differentiation and Mergers in the Carbonated Soft Drink Industry," *Journal of Economics and Management Strategy*, 14(Winter), 879–904.

Efron, B. and C.Morris (1975), "Data Analysis Using Stein's Estimator and Its Generalizations," JASA 70, 311-319.

Elrod, Terry, Gary J. Russell, Allan D. Shocker, Rick L. Andrews, Lynd Bacon, Barry L. Bayus, and Venkatesh Shankar (2002), "Inferring Market Structure from Customer Response to Competing and Complementary Products," *Marketing Letters*, 13 (3), 221–32.

Gelman, Andrew and Jennifer Hill (2006), Data Analysis Using Regression and Multilevel/ Hierarchical Models, Cambridge.

Gonzalez–Benito, O., M. P. Martinez–Ruiz, and A. Molla–Descals (2009), "Spatial Mapping of Price Competition Using Logit-type Market share Models and Store-level Scanner data," *Journal of the Operational Research Society*, 60, 52-62.

Hoerl, A.E., Kennard, R.W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, 55–67

Holman, E.W. (1979), "Monotonic Models for Asymmetric Proximities," *Journal of Mathematical Psychology*, 20, 1–15.

Kamakura, Wagner A., Russell, Gary J. (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 23 (November), 379–390.

Keller, Kevin Lane (1993), "Conceptualizing, Measuring, and Managing Customer–Based Brand Equity," *Journal of Marketing*, 57, 1–22.

Kopalle, P., Mela, C. F., Marsh, L. (1999), "The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications," *Marketing Science* 18(3), 317–332.

Krumhansl, Carol (1978), "Concerning the Applicability of Geometric Models to Similarity Data: the Interrelationship Between Similarity and Spatial

Density," *Psychological Review*, 85 (5), 445–463.

Kubokawa, Tatsuya and Muni S. Srivastava (2004), "Improved Empirical Bayes Ridge Regression Estimators Under Multicollinearity," *Communications in Statistics,* 33(9), 1943–1973.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000), "WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility," *Statistics and Computing*, 10, 325--337.

Montgomery, Alan L. (1997), "Creating Micro–Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science*, 16(4), 315–337.

Montgomery, Alan L. and Rossi, Peter E.(1999), "Estimating Price Elasticities with Theory–based Priors", *Journal of Marketing Research*, 36 (4), 413–423.

Na, Woon Bong, Roger Marshall, and Kevin Lane Keller (1999), "Measuring brand power: validating a model for optimizing brand equity," *Journal of Product and Brand Management*, 8, 3, pp.170 – 184.

Nosofsky, R.M. (1991), "Stimulus Bias, Asymmetric Similarity, and Classification," *Cognitive Psychology*, 23, 94–140.

Okada, A. and Imaizumi, T. (1987), "Nonmetric multidimensional scaling of asymmetric proximities," *Behaviormetrika*, 21, 81–96.

Okada, Akinori and Imaizumi, Tadashi (2007), "Multidimensional Scaling of Asymmetric Proximities with a Dominance Point," *Advances in Data Analysis*, Springer.

Park, Joonwook, Wayne S. Desarbo & John Liechty (2008), "A Hierarchical Bayesian Multidimensional Scaling Methodology for Accommodating Both Structural and Preference Heterogeneity," *Psychometrika*, 73 (3), 451-472.

Pesaran, M. H., Smith, R. (1995), "Estimating long–run relationships from dynamic heterogeneous panels," *Journal of Econometrics* 68, 79–113.

Pinkse, Joris, Margaret E. Slade, and Craig Brett (2002), "Spatial Price Competition: A Semiparametric Approach," *Econometrica* 70 (3), 1111–1153.

Pinkse, Joris and Margaret E. Slade (2004), "Mergers, Brand Competition, and the Price of a Pint," *European Economic Review 48*, 617–643.

Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Science*. 22(3) 304–328.

Russell, Gary J. (1992), "A Model of Latent Symmetry in Cross Price Elasticities," *Marketing Letters* 3(2), 157–169.

Russell, Gary J., Ann Petersen, and Suresh Divakar (2008), "Analysis of Brand Price Competition Using Measures of Brand Similarity," *working paper*.

Saburi, S. & N.Chino (2008), "A Maximum Likelihood Method for an Asymmetric MDS Model, " *Computational Statistics and Data Analysis*, 52,4673-4684.

Saito, T, (1986), "Multidimensional Scaling to Explore Complex Aspects in Dissimilarity Judgment," *Behaviormetrika*, 20, 35–62.

Saito, Takayuki (1991), "Analysis of Asymmetric Proximity Matrix By a Model of Distance and Additive Terms," *Behaviormetrika* (29), 45–60.

Saito, Takayuki and Shin–ichi Takeda (1990), "Multidimensional Scaling of Asymmetric Proximity: Model and Method," *Behaviormetrika* (1990), 28, 49–80.

Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *Ann. Statist.* 6(2), 461–464.

Sethuraman, Raj and V.Srinivasan (2002), "The Asymmetric Share Effect: An Empirical Generalization on Cross–Price Effects," *Journal of Marketing Research,* 39 (3), 379–386.

Shepard, R.N. (1962), "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function," Part I. *Psychometrika*, 27, 125–140.

_____ (1962), "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function," Part II. *Psychometrika*, 27, 219–246.

_____ (1972), "A Taxonomy of Some Principal Types of Data and of Multidimensinoal Methods for Their Analysis." *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, vol (1), pp. 21-47. New York: Seminar Press.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, and Angelika Van Der Linde (2002), "Bayesian Measures of Model Complexity and Fit," *J. R. Statist. Sco. B,* 64(4), 583–639.

Srinivasan, Shuba, Popkowski Leszczyc, Peter T.L., and Bass, Frank M. (2000), "Market share response and competitive interaction: The impact of temporary, evolving and structural changes in prices," *International Journal of Research in Marketing*, 17, pp. 281–305.

Srivastava, Muni S. and Tatsuya Kubokawa (2005), "Minimax Multivariate Empirical Bayes estimators Under Multicollinearity," *Journal of Multivariate Analysis,* 93, 394–416.

Steenkamp, Jan-Benedict E. M. and Marnik G. Dekimpe (1997), "The Increasing Power of Store Brands: Building Loyalty and Market Share," *Long Range Planning*, 30, 6, pp. 917 - 930.

Stobart, P. (1994), *Brand Power*, Macmillan, Basingstoke.

Stolwijk, A.M, H Straatman, & G A Zielhuis (1999), "Studying Seasonality by Using Sine and Cosine Functions in Regression Analysis," *J Epidemiol Community Health*, 53, 235:238.

Tellis, G.J., (1988), "The price elasticity of selective demand: A meta–analysis of econometric models of sales," *Journal of Marketing Research* 25, 331–341.

Tversky, Amos (1977), "Features of Similarity," *Psychological Review*, 84 (4), 327–352.

Wedel, Michel and Jie Zhang (2004), "Analyzing Brand Competition Across Subcategories," *Journal of Marketing Research*, 41 (4), 448–456.

Weeks, D.C. and P. M. Bentler (1982), "Restricted multidimensional scaling models for asymmetric proximities," *Psychometrika*, 47, 201–208.

Zielman, Berrie and Willem J. Heiser (1993), "Analysis of Asymmetry by a Slide–Vector," *Psychometrika*, 58 (1), 101–114.

Chapter 3: Essay 2: A Utility-based Model of Asymmetric Competitive Structure Analysis Using Aggregate Scanner Data and Forced Switching Data

3.1 Introduction

In Chapter 2, I described an approach to internal analysis of market structure based on the reduced-form demand structure common in economics, I now turn to an approach to internal analysis of market structure which focuses on embedding market structure into consumer's utility functions themselves, in the spirit of the choice modeling literature. The reason I propose these two different approaches is to provide researchers different angles to view similar research question, and also to address different research needs, as each method has its own weaknesses and advantages.

In particular, as with all other reduced-form models, the model proposed in Chapter 2 has the advantages of being flexible, requires fewer (and weaker) consumer-level behaviour assumptions, is not sensitive to measurement error in the left-hand side variables (Berry 1994), and lets the data "speak." However, how to address the change of the resultant market map in response to the addition or deletion of products can be challenging . The utility-based approach, on the other hand, does not suffer from this limitation. And as a structural model,[30] the utility-based approach has the benefit of being parsimonious. The market structure derived in this way is also more stable, since the addition or deletion or reposition of a brand will not fundamentally change the derived market structure.

[30] By structural models, I mean those models that "use a behavioural specification of economic agents (consumers and/or firms) to derive a relationship between endogenous and exogenous variables that may be observed by an empirical researcher" (Mazzeo 2006)

Furthermore, the model does not "run the risk of producing misleading forecasts of the effects of policy/strategic changes that change the stochastic context in which the decisions are made" (Chintagunta, Erdem, Rossi & Wedel 2006). However, the disadvantage of this approach is that its performance actually relies on the validity of discrete choice assumptions, such as that choice be constrained to be "single-item single-unit." For some product categories where this assumption does not hold, the reduced-form-based approach may be better than the utility-based approach. For a detailed comparison of reduced-form and structural models, see Chintagunta, Erdem, Rossi & Wedel (2006) and Mazzeo (2006).

In both economics and marketing, estimating a utility-based demand function using aggregate data is not new (e.g., Berry 1994; BLP 1995; Kim 1995; Zenor & Srivastava 1993; Kamakura & Srivastava 1986; Chen & Yang 2007; Musalem, Bradlow and Raju 2007). The main focus of this literature is on how to account for the heterogeneity and/or the endogeneity problem so as to derive valid estimates of demand parameters. Since sometimes, aggregate data alone may not enable precise estimates of heterogeneity parameters, previous researchers have used consumer-level data (e.g., Goldberg 1995, Albuquerque and Bronnenberg 2005) or the distribution of consumer-level data (e.g., Petrin 2002) to facilitate the analysis of aggregate data. Although the current essay is built upon this literature, my research interest focuses on internal analysis of market structure directly from demand function with aggregate data. This focus puts this research in direct contrast with the external analysis of preference in Kamakura & Srivastava (1986), and also to the choice map with panel data (Elrod 1988a, 1988b). Similar to Chapter 2, the current essay emphasizes the importance of utilizing all market structure information in the utility function (i.e., both mean preferences in the population and the covariance matrix of preferences) to derive the market map, therefore is also different from Chintagunta et al. (2002, 2003). But more explicitly than in Chapter 2, the proposed utility function has the flexibility of incorporating both vertical latent characteristics (taking the form of a vector formulation) and horizontal latent characteristics (taking the form of an ideal-point formulation). This differs from all previously mentioned work that use either the vector model (Chintagunta et al. 2002, 2003) or the ideal point model (Kamakura & Srivastava 1986) of preferences is assumed. The market map generated is able to represent both types of attributes. Moreover, rather than relying on the consumer-level information to augment the aggregate data (as done in some past work), and relating this information directly with the heterogeneity parameters in the utility function, I propose ways to combine aggregate data with survey data (i.e., forced switching data, defined below) to help estimate not only consumer heterogeneity but also the latent market map in the utility function.

Brand switching data is an important source that has been used to form proximity measures in marketing research (Lehmann 1972; Desarbo & Manrai 1992; Chintagunta 1998). Forced switching data, as one form of brand switching data, measures the substitutability relationships between pairs of brands by providing the probabilities of switching to products in the situation where an individual's most preferred product is assumed (*forced)* to be not available. In the literature, forced switching data had been used to judge a predefined market structure by conducting a series of classical hypothesis tests (Urban, Johnson & Hauser 1984). The basic logic is that when a specific set of submarkets exist, the forced switching probability should be different from what is predicted only by market share information. That is, we should observe

$$p_{ij} > \frac{MS_j}{1 - MS_i} \text{ if j is in the same submarket as i}$$

$$p_{ij} < \frac{MS_j}{1 - MS_i} \text{ if j is NOT in the same submarket as i} \, .$$

The implication here is that brands that have more commonality will have higher than average forced switching probability. However, when the purpose of the research is to *identify* the market structure, this method (which requires hypothesizing a market structure) is not effective anymore, and usually MDS techniques can be adopted instead. Nevertheless, ambiguity still exists about how to transform an asymmetric matrix of forced switching probabilities into a symmetric distance measure to be used with MDS methods. For example, some researchers might convert the forced switching matrix into a distance measure by simply averaging the switching probability matrix with its transpose or forming a symmetric matrix using the switching frequency. Notice that this approach treats any asymmetry as noise. A behavioral justification that can also account for asymmetry in this kind of data is therefore of theoretical interest.

In this essay, I will propose a behavioral justification that helps establish a linkage between aggregate data and forced switching data. The intuitive idea behind this method is that consumers' brand preferences give us important market structure information regarding brands' proximities in some latent attribute space, and the same proximity relationship across brands influences the consumers' forced switching behavior. Furthermore, since the utility function I propose embodies both the vector model and the ideal-point model to incorporate vertical and horizontal characteristics, I am able to show that it is the vertical characteristic that governs the observed asymmetric pattern in forced switching data.

The rest of this essay is organized as follows. Section 3.2 introduces the forced switching data. Section 3.3 illustrates my modeling framework. Section 3.4 describes my approach to model selection, identification, and estimation. Then in section 3.5 my modeling approach is applied to the soft drink aggregate data and the forced switching data. Section 3.6 concludes with a summary of limitations of my approach and directions for future research.

3.2 The Forced Switching Data

The data I use for this essay belong to the same industry as studied in Chapter 2: the Carbonated Soft Drinks, and come from two sources: aggregate scanner data and forced switching data. The former has been given detailed illustration in Chapter 2 (section 2.5.2). Therefore in this section, I will focus only on the second data type, in which consumers were surveyed to determine their willingness to substitute from one product to another. The data were collected in 1988, the same year when the aggregate data was collected. In particular, after giving a brief introduction and showing the respondent a list of 47 two-liter soft drinks, the consumers were asked:[31]

Question 1. Do you purchase any of the products in this list at least once a month? Which ones?

[The *i*th answer is coded as product *i*.]

Question 2. Now assume that one day, while shopping for (product i), you find that there are no more (product i)s on the shelf. Which product in this list would you be most likely to purchase instead of (product i)?

The researchers then repeated Question 2 for each product mentioned in response to Question 1. An answer to Question 1 and the associated answer to Question 2 represent a forced switch from one soft drink to another.

The data were collected in an intercept survey conducted outside the same St. Louis supermarket (again, the same supermarket where the aggregated data were collected) during a one week period. Of the 1,180 shoppers approached, 455 (38.5%) responded; an additional 267 (22.6%) were willing to respond but did not purchase any of the listed products. On average, each of the 455 respondents indicated that he or she purchased 2.44 different two-liter soft drink brands a month.

Table 3-1 shows the survey responses for 20 of the key products involved in the FTC v.Coca-Cola Co. litigation. A row corresponds to a soft drink that a respondent selects in answer to Question 1 ("purchase at least once a month").

[31] The product list actually contained 48 products. We noticed, however, that many respondents meant Coke Classic, but said Coke when they saw the product 'Coke' on the list. We, thus, treat Coke and Coke Classic as one single product, call that product 'Coke,' and perform the analysis using 47 products. The survey was conducted from September 30 through October 6, 1988.

Let N_k denote the number of times the product in the kth row was chosen as an answer to Question 1 (see the left-hand column). A column corresponds to a soft drink that a respondent selects as an answer to Question 2 ("most likely to purchase instead"). I let N_{kj} denote the number of times the product in the jth column was chosen instead when the product in the kth row was not on the shelf. The entry in the Sprite row and 7-Up column, for example, indicates that of the $N_k = 41$ respondents that purchase Sprite "at least once a month," $N_{kj} = 22$ would be "most likely to purchase" 7-Up "instead" if Sprite were not on the shelf. It is worth noting that although the forced-switching data is not readily available, it does provide potential information for strategic stockouts.

Table 3-1 Forced Switching Data

	Question 1 Responses (N_i)	Question 2 Responses (N_{ij})																				
	Row Total	Pepsi	Diet Pepsi	Mount. Dew	Pepsi Free	Diet Pepsi Free	Lemon Lime Slice	Diet Lemon Lime Slice	Cherry Coke	Diet Cherry Coke	Coke	Diet Coke	Caff.-Free Coke	Caff.-Free Diet Coke	Sprite	Diet Sprite	Mr. PiBB	Dr. Pepper	Diet Dr. Pepper	7-Up	Diet 7-Up	Other
Pepsi	148	**0**	1	1	3	0	1	0	5	0	84	3	1	0	0	0	1	3	0	5	1	39
Diet Pepsi	110	6	**0**	1	1	1	1	0	0	1	4	57	0	2	2	0	0	3	3	0	4	24
Mountain Dew	29	1	1	**0**	0	0	4	0	0	0	3	0	0	0	1	0	0	2	1	1	0	15
Pepsi Free	24	3	2	0	**0**	0	1	0	0	0	2	2	7	0	0	0	0	0	0	0	0	7
Diet Pepsi Free	21	1	1	0	1	**0**	0	0	0	0	0	6†	1	9	0	0	0	0	0	0	1	1
Lemon-Lime Slice	7	0	0	0	0	0	**0**	0	0	0	0	0	0	0	1	1	0	0	0	2	1	2
Diet Lemon-Lime Slice	3	0	1	0	0	0	0	**0**	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Cherry Coke	25	7	0	0	0	0	0	0	**0**	0	5	0	0	0	1	0	0	0	0	1	0	11
Diet Cherry Coke	15	2	1	0	0	0	0	0	0	**0**	0	4	0	0	0	0	0	1	1	0	0	6
Coke	168	103	2	0	0	0	0	0	4	0	**0**	5	1	1	3	1	1	6	0	4	0	37
Diet Coke	140	4	74	1	1	2	0	0	1	4	5	**0**	0	3	3	2	0	0	5	1	3	31
Caffeine-Free Coke	16	4	0	0	6	1	0	0	0	0	2	0	**0**	0	0	0	0	0	0	0	0	3
Caff-Free Diet Coke	45	3	5	0	1	14	0	0	0	0	2	3	0	**0**	0	2	0	0	1	1	0	13
Sprite	41	0	0	2	0	0	3	0	0	0	4	1	2	0	**0**	2	0	1	0	22	1	3
Diet Sprite	15	1	0	0	0	0	1	1	0	0	1	0	0	0	0	**0**	0	0	0	0	7	4
Mr. PiBB	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	**0**	1	0	0	0	1
Dr. Pepper	59	6	2	2	0	1	0	0	2	1	8	3	0	0	0	0	11	**0**	0	1	0	22
Diet Dr. Pepper	12	0	2	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	**0**	1	0	6
7-Up	51	1	0	1	0	0	6	0	0	0	2	1	0	0	25	0	0	0	0	**0**	1	14
Diet 7-Up	30	1	1	0	0	0	0	3	0	0	2	1	0	1	1	8	0	0	0	0	**0**	12
Other	147	14	5	2	3	1	2	0	3	0	23	5	0	2	3	2	2	2	0	5	2	71
Column Total	1109	157	98	10	16	21	19	4	16	6	147	93	12	18	40	20	15	19	11	44	21	322

3.3 Modeling Framework

In this section, a utility function that incorporates both vertical characteristic (using vector model) and horizontal characteristics (using ideal-point model) is developed, and assumed to underlie both aggregate demand data and forced switching data. The incorporation of consumer heterogeneity in the utility function permits flexible substitution patterns across alternatives. And the vertical characteristic, interpreted here as quality, explains the observed asymmetric switching patterns in forced switching data.

3.3.1 Modeling Scanner Data

Utility structure. Concerning notation, let $t = 1,...,T$ weeks; $i = 1,...,I$ individuals; $k, j = 1,...,J$ brands; $m = 1,...,M$ latent dimensions for a market structure space; θ_{jm} the *m*th coordinate for brand *j*; w_{im} is the *m*th ideal point coordinate for individual *i*; α_j describes the quality level of product *j*; v describes the consumer *i*'s marginal willingness-to-pay for quality, assumed to be the same across consumers; β describes price sensitivity, again assumed to be the same across consumers; $\ln p_{jt}$ is the log price for product *j* at week *t*. And further let α denote the *J* x *1* vector of quality measures, θ_m the *J* x *1* vector of brands' coordinates on *m*th latent dimension, and $\Theta = [\theta_1 \quad ... \quad \theta_m \quad .... \quad \theta_M]$. Finally, let $\underline{q} = (q_{.1},...,q_{.t},...,q_{.T})$, with $q_{.t} = [q_{1t} \quad ... \quad q_{Jt}]$; and $\underline{p} = (p_{.1},...,p_{.t},...,p_{.T})$, with $p_{.t} = [p_{1t} \quad ... \quad p_{Jt}]$.

I define the latent utility derived by consumer *i* from consuming product *j* at week t as given by the following function:

$$u_{ijt} = v\alpha_j - \sum_{m=1}^{M} (\theta_{jm} - w_{im})^2 - \beta \ln p_{jt} + \varepsilon_{ijt}, \tag{3-1}$$

Some important conceptual issues need to be emphasized for my utility function in Equation (3-1). First, the utility function combines both vertical (i.e., α) and horizontal characteristics [32] (i.e., $\theta_m,\ \ m = 1,...,M$), reflecting the fact that these two types of characteristics are embodied actually in most of the products in the market (Anderson, De Palma & Thisse 1992, Chapter 8). The vertical characteristic is interpreted as quality, although it captures more subtle concepts such as a brand image, attractiveness, familiarity, etc. No matter what interpretation is given to this attribute, the essence of it is that the consumers have a common ordering of the brands in terms of this feature, and when all brands are priced at the same level, all consumers agree that more of this feature is always better, even though they may vary in their willingness to pay for this feature. On the other hand, horizontal characteristics find their origin in consumers' taste dispersion, as reflected by the ideal-point framework in the utility function (Shocker and Srinivasan 1979).

Second, although I use the term "ideal point" for $w_{i.}$ in Equation (3-1), it is better interpreted as the feasible ideal point, or, more accurately, an attraction point. An ideal brand would have infinite quality at zero price, which is not very relevant for consumer choice. By contrast, an attraction brand would be the most preferred brand that lies in the market map spanning all brands. Notice a

[32] Throughout this dissertation, the words characteristics, features, and attributes all share the same meaning.

consumer's attraction brand is not unique, and could differ across purchase occasions. The non-uniqueness property of the attraction brand explains why in reality, we observe different brands being chosen by different consumers in different occasions. For simplicity, from now on, I would still use "ideal point" terminology, although what I really mean is the attraction point.

Third, although this non-traditional utility function blends itself well in the product differentiation literature in both economics and marketing (e.g., Anderson, De Palma & Thisse 1992; Sutton 1986; Vandenbosch & Weinberg 1995), there are two differences that need to be emphasized here: (a) in the current essay, I model both types of characteristics as latent. In particular, the horizontal differentiation of the products is assumed to be explained by M latent characteristics ($\theta_m, \;\; m = 1,...,M$), and the vertical differentiation is captured by one latent variable (α). Therefore, the approach adopted here focuses on internal analysis of preferences (Elrod 1991). It is worth mentioning that adding observed attributes is also possible for this utility function. The observed attributes can be included directly in the right-hand side of the Equation (3-1), or used to explain the latent attributes by adding another hierarchical level in the model structure. The current paper, however, only focuses on latent attributes for simplicity. (b) The horizontal characteristics and the vertical characteristic are not necessarily unrelated to each other. Specifically, in this essay, I will explore modeling the vertical feature as spanned by the same characteristics space of the horizontal differentiation. That is, we will consider the case

$$\alpha_j = \sum_{m=1}^{M} \gamma_m \theta_{jm} , \qquad (3\text{-}2)$$

One potential advantage of this specification is its parsimony, in the sense that it reduces the number of parameters that require estimates, especially when the number of differentiated products is large in the market. Another more important advantage is that it helps researchers to use a single spatial map to explain all brand-related differences, so now the effect of the repositioned, new or deleted brands on quality level can also be predicted. I acknowledge that there certainly exist some contexts that quality is simply a unique feature, and the specification in (3-2) will not work. And under these circumstances, researchers can always estimate α as separate parameters from $\theta_m, \ m=1,...,M$, otherwise any attempt that tries to impose the specification in (3-2) may lead to insignifant parameter estimates for $\gamma_m, \ m=1,...,M$, and poor model performance. However for a product cateogry featured mainly by horizontal differentiation, we can put forward conceptual justification for this specification. For example, in the soft drink industry, quality may be more likely to be related with such concepts as consumers' perceived attachment or familiarity to a certain brand, which may connect to such horizontal features as cola vs. non-cola, name of manufacturers, or even spatial density (see section 2.3.3 for a detailed illustration of spatial density concept) which is captured in a spatial map of horizontal characteristics. Fourth, consumer heterogeneity in the current essay is said to be described by a consumer's ideal point (i.e, $w_{im}, i=1,...,N; m=1,...,M$) with respect to the horizontal characteristics $\theta_m, \ m=1,...,M$. Inclusion of this heterogeneity is necessary for avoiding the unreasonable substitution pattern across brands implied by the IIA assumption of a standard logit model (Berry 1994; BLP 1995).

In fact, it can be shown that the heterogeneity of preferences implies that the choice probability for a given alternative j will depend not only on its own position on the structural map ($\theta_{jm},\ m=1,...,M$), but also on the other alternatives' position (see Appendix D for more detailed illustration, which is logically quite similar to Kamakura & Srivastava 1986). Finally, both willingness-to-pay (i.e., ν) and price sensitivity parameter (i.e., β) are assumed here to be the same across consumers for simplicity. I also consider adding the heterogeneity in these two parameters a future extension of current work. It is worth noting that when quality α is treated as a latent variable and modeled as in Equation (3-2), and the willingness-to-pay parameter ν is considered a fixed effect, we will have an identification problem of estimating both ν and $\gamma_m,\ m=1,...,M$. So from now on, I will drop ν from the equation (3-1).

Fifth, although my proposed utility function assumes ideal-point model for horizontal characteristics, the formulation actually accommodates linear or monotonically increasing preferences, which occur when the ideal points go to infinity (Carrol 1972, Kamakura & Srinivasan 1986). And when this happens, the utility function in Equation (1) will solely depend on vector formulation. Furthermore, according to Kamakura & Srinivasan (1986): *"a distribution of ideal points located at infinity or substantially beyond the relevant range of the attribute space would approximate a random distribution of linear preferences."*

Aggregate Choice Probability. Assuming ε_{ijt} in the utility function is distributed independently according to a Type I Extreme Value distribution with mean zero and unit scale parameter, and combining Equation (3-1) and (3-2), the probability ρ_{ijt}, that the *i*th consumer in period *t* chooses brand *j*, is then given by

$$\rho_{ijt} = \frac{\exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - w_{im})^2 - \beta \ln p_{jt})}{\sum_{l=1}^{J} \exp(\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - w_{im})^2 - \beta \ln p_{lt})} \quad (3\text{-}3)$$

Since the data we observe is aggregate level, it is not possible for us to estimate w_{im} for each household. Hence I impose distributions for these parameters across households, with the parameters for these distributions estimated instead, when applicable. Specifically, I rewrite Equation (3-3) as

$$\rho_{jt} \mid i = \frac{\exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - w_m)^2 - \beta \ln p_{jt})}{\sum_{l=1}^{J} \exp(\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - w_m)^2 - \beta \ln p_{lt})}, \quad (3\text{-}3')$$

and denote $f(w \mid \overline{w}, \Omega)$, where w is *M* by *1* vector, a multivariate normal distribution with mean vector $\overline{w}$ and variance-covariance matrix Ω. Then by integrating the probability in Equation (3-3') over the proposed multivariate normal distribution, we derive the aggregate probability function for choosing each brand *j* at week *t*,

$$\rho_{jt}(p_{.t}; \gamma, \beta, \Theta, \overline{w}, \Omega) = \int_w \frac{\exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - w_m)^2 - \beta \ln p_{jt})}{\sum_{l=1}^{J} \exp(\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - w_m)^2 - \beta \ln p_{lt})} f(w \mid \overline{w}, \Omega) dw$$

(3-4)

where $\gamma = [\gamma_1, ..., \gamma_M]$. I acknowledge the potential limitation of assuming a multivariate normal distribution for heterogeneity parameters. A bi-modal distribution or multi-model distribution may better describe this dataset. However, I leave this conjecture for future research.

Multinomial-distributed Aggregate Sales Data. In this section, consistent with the common practice of marketing literature (Kim 1995; Zenor & Srivastava 1993; Bodapati & Gupta 2004), I assume that the units of brand i purchased at week t are independently made by q_{it} distinct number of households, with each household purchases at most one unit of a product at each week. Therefore the observed aggregate unit sales q_{it} come from a multinomial distribution in which probabilities are given by Equation (3-4). (For a detailed discussion regarding the multinomial distribution assumption for aggregate data, please refer to section 1.2.2) That is, we have

$$q_{.t} \sim multinomial(\rho_{jt}(\underline{p}, \gamma, \beta, \Theta, \overline{w}; \Omega), Q_t), \quad Q_t = \sum_{j=1}^{J} q_{jt}$$

with the density function given by

$$\Pr(q_{.t},\ p_{.t} \mid \gamma, \beta, \Theta, \overline{w}; \Omega) = \frac{Q_t!}{q_{1t}! q_{2t}! \ldots q_{Jt}!} \prod_{j=1}^{J} \rho_{jt}(p_{.t}; \gamma, \beta, \Theta, \overline{w}; \Omega)^{q_{jt}} \quad . \qquad (3\text{-}5)$$

So the likelihood function for aggregate data would be

$$\mathrm{L}_{\mathrm{agg}} = \prod_{t=1}^{T} \Pr(q_{.t},\ p_{.t} \mid \gamma, \beta, \Theta, \overline{w}; \Omega) \qquad (3\text{-}6)$$

3.3.2 Modeling the Forced Switching Data

Utility Structure and Forced Switching Probability. The marketing literature has witnessed a long history of relating observed brand-switching data with brands similarity, which is measured either directly by the distances between pairs of brands (Lehmann 1972; Desarbo & Manrai 1992; Chintagunta 1998), or indirectly by brands' substitutability pattern such as price elasticities (Bucklin, Russell & Srinivasan 1998). The basic assumption underlying this literature is that consumers are more likely to switch between brands which are more similar to each other than those which are less similar (Urban, Johnson & Hauser 1984; Novak & Stangor 1987). Based on this assumption, and given the utility structure in Equation (3-1), I assume that when consumers' first priority brand *k* is out of the market, they would choose a brand *j* that has the closest distance to *k* in the *M*-dimensional structural map *when every other factor (i.e., quality level, prices) is set to be equal*. Following this logic, a consumer at week *t* will choose a brand that maximizes his/her utility function of the form (combining Equation (3-1) and Equation (3-2))

$$u_{j,-k} \mid t = \sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - \theta_{km})^2 - \beta \ln p_{jt} + \varepsilon_{ijt}, j \neq k. \quad (3\text{-}7)$$

Therefore, the probability of brand *j* being chosen when the *k* as the first option is out of market will be

$$\begin{aligned} \pi_{j,-k} \mid t = prob(&\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - \theta_{km})^2 - \beta \ln p_{jt} + \varepsilon_{ijt} \geq \\ &\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - \theta_{km})^2 - \beta \ln p_{lt} + \varepsilon_{ilt}, \forall l \neq j,k) \end{aligned}.$$

Since we have IID Type I extreme value distribution with mean zero and unit scale parameter for ε_{ijt}, we have:

$$\pi_{j,-k} \mid t = \frac{\exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - \theta_{km})^2 - \beta \ln p_{jt})}{\sum_{l \neq k}^{J} \{\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - \theta_{km})^2 - \beta \ln p_{lt}\}} \quad . \qquad \text{(3-8a)}$$

And when $j = k$, we have the following equation due to the nature of the forced switching data:

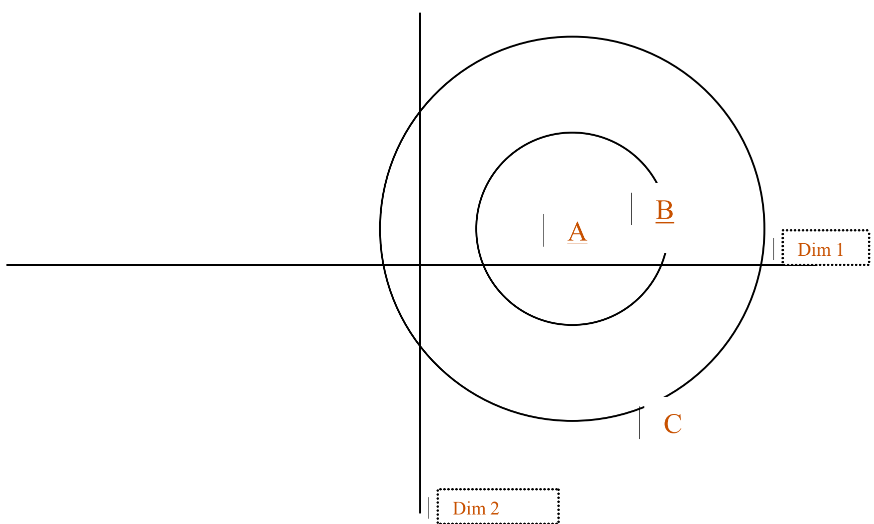
$$\pi_{j,-j} = 0 \ . \qquad \text{(3-8b)}$$

Further, notice that the assumed relationship between brands' similarity and switching pattern is consistent with the substitutability pattern implied by the utility function. Specifically, in section 3.3.1 and Appendix D, I show that, when consumer heterogeneity is included in the utility function, two brands that "resemble" each other tend to have higher covariance in utilities, therefore higher substitutability.

Two issues need to be emphasized for the utility function in Equation (3-7). First, this utility function does not consider consumers' variety-seeking behavior (Chintagunta 1998), and it implies that given a mature product category crowded with competitors, the chosen brand ought to be close to the feasible ideal brand on that occasion. The first limitation serves as a future extension of current research. While for the second point, Figure 3-1 illustrates the idea in a three-brand scenario, where brand A is chosen initially. And when it is removed from the market, it serves as an approximation to the feasible ideal point for the consumer. Now, any brand (in this example, brand B) that is closest to it in this metric space would have higher probability to be chosen, of course conditional on all other factors being the same (e.g., quality, prices). This is actually the position taken by

Lehmann (1972), who proposes that the last-period choice is an indicator of an individual's ideal point. Intuitively speaking, there might exist some brand that, when consumers stick to their true feasible ideal point (i.e., $w_{i.}$), actually generates higher utility than brand *B*. In reality, however, certain concerns may prevent consumers from engaging totally independent new search based on utility function (3-1): the consumption inertia due to, for example, switching cost; uncertainty related with the true ideal point; the salience of certain attributes associated with the first chosen brand *A*; and the correlation among alternatives, etc.

Figure 3-1 Ideal Point Model for Forced Switching Data



Second, the utility function in (3-7) says consumers, when making a switch, consider not only the similarity of the new brand with the first priority brand *k*, but also the quality and price level. So when two brands that have the same distance to *k,* only the brand that has the highest quality-price ratio will be chosen. And this also implies that for any two brands *j* and *k*, the forced switching probabilities are not symmetric, that is $\pi_{j,-k} \mid t \neq \pi_{k,-j} \mid t$. In particular, by appropriate rearranging the components in Equation (3-8), we derive the following formula

$$\pi_{j,-k} \mid t = \frac{W_{jt}\exp(-d_{jk})}{\sum_{l \neq k}^{J} W_{lt}\exp(-d_{lk})} \;,\; j \neq k \qquad \text{(3-8a')}$$

with $W_{jt} = \exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \beta \ln p_{jt})$, $d_{jk} = \sum_{m=1}^{M} (\theta_{jm} - \theta_{km})^2$. Here $W_{jt} > 0$ is the weight for brand *j,* and is related with brands' quality-price ratio level. This formulation coincides with Shepard (1957)'s suggestion regarding adding weights to distance measure to control the asymmetric pattern in observed stimulus and response matrix. Furthermore, Equation (3-8a) also assumes that a brand's price only linearly affects the utility, therefore, consumers, when making comparison between two brands, do not make price-based adjustments for each attribute dimension θ_{jm}.

It is worth mentioning that the above equation (3-8'') also implies the observed asymmetry in forced switching probability is resulted from two sources: (1) weight differences (i.e., $W_{jt} \neq W_{kt}$): products that have higher weight tend to be more dominant (powerful) in the market; (2) product's weighted proximity to all other alternatives in the market: products that are more close to all the other brands, especially those brands that have higher power (i.e., denser space in metric space) have lower corresponding forced switching probabilities

$(\sum_{l \neq k}^{J} W_{lt} \exp(-d_{lk}) \neq \sum_{l \neq j}^{J} W_{lt} \exp(-d_{lj}))$.

Aggregate Forced Switching Probability. Notice that the forced switching probability in Equation (3-8) is time-specific ($\ln p_{jt}$).[33] Since the forced switching data we observe is usually collected at one time, it is not possible for us to know exact p_{jt} information. One approach to solve the time-dependent problem is to calculate the weighted average of prices over time for each alternative, with the weight depending on quantity purchased for this alternative at each time. [34] The number of weeks used for this calculation can be set to equal the total number of weeks for the aggregate scanner data. Intuitively, for a product category featured mainly by vertical differentiation, including $\ln p_{jt}$ is important, since it can explain why most of the times we observe consumers only switch between products that have similar quality level. However, when it concerns a horizontally differentiated product category, as in soft drink industry, a more simple way would be simply dropping this price term from Equation (3-8), since most of the products may be priced very similar, especially when they are within the same product line. Besides, different brands may take turns frequently engaging in promotion activities (e.g., Pepsi and Coke in my data set), so in long run, the difference in their average prices may be small enough to be ignored by consumers. Specifically, when applying the soft drink

[33] When heterogeneity is added to the willingness-to-pay parameter, the forced switching probability is also individual-specific.
[34] Using weighted average of prices may not always be a good idea, as it may put too much weight for prices at discount period. For certain product category (e.g., luxury brand), consumers may take the prices that appear most frequently during a specific period of time as the prices considered in their forced switching behavior.

data, instead of Equation (3-8a), I estimate

$$\pi_{j,-k}(\gamma,\Theta)=\frac{\exp(\sum_{m=1}^{M}\gamma_m\theta_{jm}-\sum_{m=1}^{M}(\theta_{jm}-\theta_{km})^2)}{\sum_{l\neq k}^{J}\{\sum_{m=1}^{M}\gamma_m\theta_{lm}-\sum_{m=1}^{M}(\theta_{lm}-\theta_{km})^2\}} \quad . \tag{3-8a''}$$

Multinomial Distributed Forced-Switching Frequencies. In this section, the notation is consistent with section 3.3. I assume that when brand k was chosen as a first priority, the observed frequency N_{kj} of switching from k to brand j follows a multinomial distribution in which probability is given by Equation (3-8'''). Let $N_{k.}=[N_{k1},...,N_{kJ}]$, and $\underline{N}=[N_{1.},...,N_{J.}]$, then we have,

$$N_{k.}\sim multinomial(\pi_{j,-k}(\gamma,\Theta),N_k),N_k=\sum_{j=1}^{J}N_{kj},N_{kk}=0,$$

with the density function given by

$$\Pr(N_{k.}\mid\gamma,\Theta)=\frac{N_k!}{N_{k1}!...N_{k(k-1)}!N_{k(k+1)}!...N_{kJ}!}\prod_{j\neq k}^{J}\pi_{j,-k}(\gamma,\Theta)^{N_{kj}} \quad . \tag{3-9}$$

So the likelihood function for forced switching data would be

$$\mathrm{L}_{\mathrm{swi}}=\prod_{k=1}^{J}\Pr(N_{k.}\mid\gamma,\Theta). \tag{3-10}$$

The multinomial assumption treats every observed switching comes from an independent consumer. [35] This assumption might still be valid, however, if we consider different first choices for a consumer come from the needs for different household members or for different usage occasions. And these different needs result in multiple ideal points (Lee et al. 2002) for second choice.

[35] The forced switching data I use in this essay actually permited consumers to pick more than one brand as their first priority.

Finally, combining the aggregate scanner data and forced switching data, the total likelihood function would be

$$L(\underline{N}, \underline{p}, \underline{q}; \gamma, \beta, \Theta, \overline{w}, \Omega) = L_{\text{swi}} L_{agg} = \prod_{k=1}^{J} \Pr(N_{k.} \mid \gamma, \Theta) \prod_{t=1}^{T} \Pr(q_{.t}, \; p_{.t} \mid \gamma, \beta, \Theta, \overline{w}, \Omega) \; .$$

(3-11)

3.4 Model Identification and Estimation

3.4.1 Model Identification

My model focuses on the internal analysis of preference, and notice the latent parameters in the model include the heterogeneous ideal-points $w' = [w_1, ..., w_M]$, and the brand location matrix $\Theta = [\theta_1 \quad ... \quad \theta_m \quad .... \quad \theta_M]$. As I discussed before, I assume the M by 1 ideal-point vector w follows a multivariate normal distribution with mean $\overline{w}$ and variance-covariance matrix Ω (i.e, $f(w \mid \overline{w}, \Omega)$). The brand location parameters $\Theta = [\theta_1 \quad ... \quad \theta_m \quad .... \quad \theta_M]$ enter the equation (3-1), (3-2), (3-7), mainly through the ideal-point formulation. They are subject to the same flipping, rotation as well as location problems of a typical inter-brand distance calculation. Therefore, following Elrod (1988b), to solve the rotational problem of the map, Ω is constrained to be diagonal with the form

$$\Omega = \begin{bmatrix} \sigma_1^2 & & & \\ 0 & \sigma_2^2 & & \\ 0 & 0 & .... & \\ 0 & 0 & 0 & \sigma_{\mathrm{M}}^2 \end{bmatrix}.$$

This constraint is without loss of generality. The covariance matrix Ω measures the degree of dispersion of the ideal-points over the attribute space. The smaller the σ_{m}^2 the more homogeneous consumers' preferences are along the mth attribute dimension. Further, similar to section 2.4.2, for the origin indeterminacy, I impose a restriction on the coordinates for each dimension such that $\sum_{j=1}^{J}\theta_{jm} = 0, m = 1,..., M.$ To solve the flipping problem, I constrain M stimulus coordinates to lie in the positive orthant of the derived space. For instance, with M=2, I constrain the elements of Θ on the diagonal on the positive real line (i.e, $\theta_{11} > 0,\ \ \theta_{22} > 0,$).

Finally, to summarize the notation of the heterogeneity distribution, I can write

$$f(w \mid \overline{w}, \Omega) = f(w \mid \overline{w}, \sigma_1^{\ 2}, ..., \sigma_M^{\ 2}) \tag{3-12}$$

3.4.2 Model Estimation

Similar to Chapter 2, a hierarchical Bayesian modeling approach is adopted and MCMC method is used to facilitate the estimation. The model selection is mainly to select the number of dimensions *M*. Again, since adaptive Bayesian shrinkage (see section 2.4.3) is used in estimating the brands locations, the number of parameters being estimated is hard to calculate exactly. So the DIC for focal parameters (see section 2.4.1) is applied for the purpose of model selection.

Given Equation (3-11) and (3-12), the joint posterior distribution for our model is as follows

$$p(\gamma,\beta,\Theta,\overline{w},\sigma_1^2,...,\sigma_M^2,\tau_1^2,...,\tau_M^2 \mid \underline{q},\underline{N}) \propto$$
$$\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(p_{jt};\gamma,\Theta,\beta,\overline{w},\sigma_1^2,...,\sigma_M^2)^{q_{jt}}\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(N_{kj};\gamma,\Theta)^{N_{kj}}$$
$$(\prod_{m=1}^{M}\phi(\gamma_m))\phi(\Theta \mid \tau_m^2, m=1,..,M)\phi(\beta)(\prod_{m=1}^{M}\phi(\overline{w}_m))(\prod_{m=1}^{M}\phi(\sigma_m^2))$$
$$\prod_{m=1}^{M}\phi(\tau_m^2)\ ,$$

(3-14)

where $\phi(.)$ is the prior distribution for the corresponding parameters. In particular, the second line of Equation (3-14) gives the joint likelihood function of the data; the third line the prior distribution for the focal parameters of the model; the fourth line the prior for the hyper-parameters. I adopt the adaptive Bayesian shrinkage idea discussed in section 2.4.3 when estimating Θ, that is, $\phi(\theta_{jm}) \sim N(0,\tau_m^2)$. Then I assign weakly informative priors on all the hyper-parameters and the other parameters. That is, for $\gamma_m, \overline{w}_m, \beta$, a univariate normal distribution $N(0,10^4)$ is imposed; while for all standard deviance parameter, σ_m^2, τ_m^2, an inverse gamma distribution is imposed, say, $\sigma_m^{-2} \sim G(0.01,100)$, with the second term in the gamma distribution a scale parameter. Finally, since no analytical solution exists for the integrations in Equation (3-4), numerical integration is used.

The estimation of the proposed model is carried out by setting up a series Markov chain and iteratively sampling from the conditional distributions of model parameters. The draws for the parameters are complicated due to the approximation in Equation (3-4), so the Metropolis-Hasting algorithm is used throughout the whole simulation process (for details, see Appendix E).

3.5 Application

The model is applied to the data from the carbonated soft drink industry, which consists of the aggregate weekly store-level scanner data and the forced switching survey data. The MCMC algorithm is implemented in R, and for one-dimensional model, the total deviance is 78158.24 and DIC is 78181. The number of focal parameters are 24. While for two-dimensional model, deviance is 69385.04 and DIC is 69430.05. The number of focal parameters are 47. To facilitate better comparison across models in this Chapter and Chapter 2, I further decompose total deviance of the model into two parts: the deviance for soft drink data and the deviance for forced switching data. In particular, for one-dimensional model, the deviance for soft drink data is 72775.84 and for forced switching data is 5382.38. For two-dimensional model, the total deviance is with soft drink data 65986.05 and forced-switching data 3398.988. This suggests the two-dimensional model works much better than one-dimensional model. So I report the findings from the two-dimensional model in the following paragraphs.

Table 3-2 shows the estimated utility function parameters from the two-dimensional model.

Table 3-2 Estimates of Utility Function Parameters from 2D Model

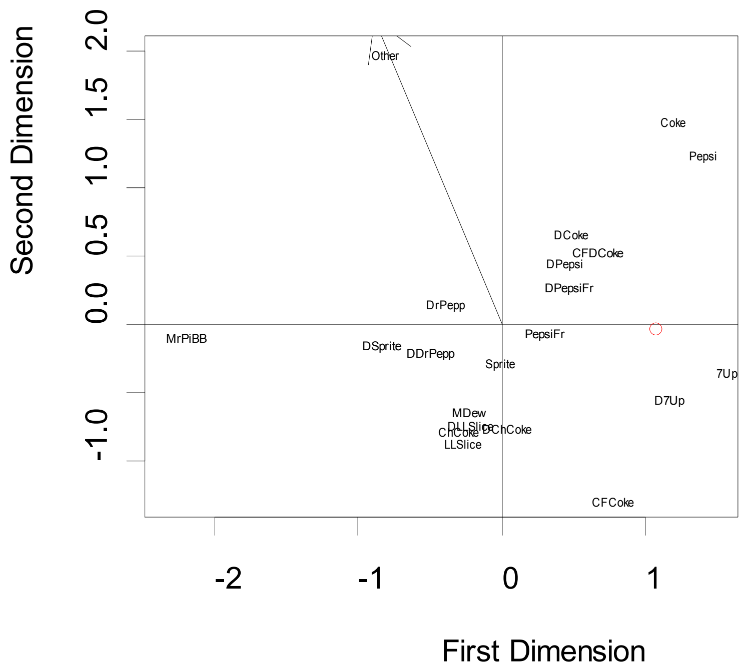
		Posterior Mean (Posterior Sd.)	
Price sensitivity	β	5.7 (0.021)∗	
Quality coefficients	γ_1	-0.89 (0.057) ∗	
	γ_2	2.23 (0.053) ∗	
Heterogeneity Parameters for Ideal Points	$\overline{w_1}$	1.07 (0.029) ∗	
	σ_1^2	0.51 (0.011) ∗	
	$\overline{w_2}$	-0.04 (0.027)	
	σ_2^2	0.36 (0.012) ∗	
Hyperparameters (Shrinkage effects)	τ_1^2	0.82 (0.29) ∗	
	τ_2^2	0.57 (0.20) ∗	
Brands Locations		First Dimension	Second Dimension
	Pepsi	1.31 (0.04) ∗	1.24 (0.03) ∗
	Diet Pepsi	0.31 (0.03) ∗	0.44 (0.02) ∗
	Mountain Dew	-0.35 (0.06) ∗	-0.63 (0.03) ∗
	Pepsi Free	0.16 (0.04) ∗	-0.06 (0.02) ∗
	Diet Pepsi Free	0.3 (0.03) ∗	0.26 (0.02) ∗
	Lemon-Lime Slice	-0.4 (0.05) ∗	-0.87 (0.03) ∗
	Diet Lemon-Lime Slice	-0.38 (0.06) ∗	-0.73 (0.03) ∗
	Cherry Coke	-0.44 (0.07) ∗	-0.78 (0.03) ∗
	Diet Cherry Coke	-0.13 (0.05) ∗	-0.75 (0.03) ∗
	Coke	1.11 (0.04) ∗	1.49 (0.03) ∗
	Diet Coke	0.36 (0.03) ∗	0.66 (0.02) ∗
	Caffeine-Free Coke	0.63 (0.05) ∗	-1.29 (0.03) ∗
	Caffeine-Free Diet Coke	0.49 (0.04) ∗	0.53 (0.02) ∗
	Sprite	-0.12 (0.06) ∗	-0.29 (0.03) ∗
	Diet Sprite	-0.97 (0.05) ∗	-0.15 (0.02) ∗
	Mr. PiBB	-2.34 (0.04) ∗	-0.09 (0.06)
	Dr. Pepper	-0.53 (0.05) ∗	0.15 (0.02) ∗
	Diet Dr. Pepper	-0.67 (0.08) ∗	-0.21 (0.02) ∗
	7-Up	1.49 (0.03) ∗	-0.35 (0.02) ∗
	Diet 7-Up	1.06 (0.04) ∗	-0.55 (0.03) ∗
	Other	-0.91 (0.02) ∗	1.98 (0.03) ∗
DIC		69430.05	

∗ The 95% credible interval excludes zero.

All parameters in the table, except the population mean of ideal-point on the second dimension, are significant in the sense that their corresponding 95% credible interval excludes zero. In particular, consumers exhibit high price sensitivity in the data, consistent with previous literature on this industry (e.g., Dube 2004, 2005). σ_m^2 is a measure of consumers' heterogeneity of ideal-point on *m*th dimension of the latent characteristic space. The results (0.51 for σ_1^2 and 0.36 for σ_2^2) indicates some heterogeneity in the distribution of consumers' ideal-points on both dimensions. It is interesting to point out that both quality coefficients are significant different from zero, implying for soft drink product category, spanning quality on the horizontal characteristic space is a valid assumption. And it also implies that asymmetry is an essential feature for my cross-price elasticities matrix. Finally, I find strong shrinkage effects across brands.

Brand location estimates in Table 3-2 do not provide us a good interpretation of the market structure. A better alternative is to generate a two-dimensional map based on these estimates as in Figure 3-2, which displays the 21 brands' locations on two latent dimensions and represents vertical characteristic "*quality*" using an arrow. From Figure 3-2 (please refer to Appendix F for the corresponding brand names to the abbreviations in the Figure), I observe some interesting findings, some of which are comparable with those corresponding ones in Chapter 3 (section 2.4.2).

Figure 3-2 Two-Dimensional Map for Soft Drinks



First, since the angle between an arrow and an axis indicates the importance of the contribution of the corresponding variable to the axis dimension (i.e., the smaller the angle is, the higher the importance), we can see that "quality" is highly correlated with the vertical dimension. Besides, since the length of the arrow is equal to the variance of the corresponding variable, it seems that the "quality" is an important attribute for these data. Notice, the "quality" attribute in the utility function governs the observed asymmetric switching pattern in the forced switching data.

Second, the horizontal dimension distinguishes the coke vs. non-coke products. Since the brands in the figure can be projected perpendicular on the arrow, and the position of the brands along the arrow gives information of the value of the brand on corresponding variables, we can see that majority of the coke products are considered superior (higher quality) to non-coke products. In particular, Coke and Pepsi are of the highest quality (e.g., familiarity, etc) level compared with the other nineteen soft drink products.

Third, the distance between points reflects the perceived similarity between the corresponding two brands and the competition between them. Figure 3-2 indicates that Pepsi and Coke are in head-to-head competition in both dimensions. On the other hand, Diet Pepsi, Diet Coke and Diet Pepsi Free are all diet products and they are similar, although they seem to share a different submarket from Pepsi and Coke. Dr. Pepper products are in direct competition with Sprite products, with Diet Dr.Pepper competes more with Diet Spirit. Furthermore, there seems to be a close relationship between Cherry Coke products and Lemon-Lime Slices products. Notice these findings are consistent with what I get in Figure 2-4.

Fourth, the (bold) dot in Figure 3-2 represents the average consumers' ideal point in the population. An interesting finding here is that although both Coke and Pepsi are considered as superior in terms of quality level, they are not actually that close to consumers' ideal-points in terms of the horizontal characteristics. Consumers seem to prefer those Coke and Pepsi products that are either diet or Caffeine-free, or both. Besides, consumers do like 7-Up products, although these products seem to be considered as low quality. This finding seems to imply that *"quality"* in current context may be more likely related with such concepts as low familiarity, the low brand equity, or market share, etc .

Fifth, consistent with the Figure 2-4, MrPiBB is a true outlier in the soft drink market. It is far away from the rest of the products. From the average consumers' perspective, it is not that desirable, so it is definitely not a mainstream product. However, it is considered with a moderate quality level, which might reflect the fact that this brand enjoys some brand loyalty in a quasimonopoly, although small segment.

Finally, I employ the following definition of aggregate demand price elasticities, and report them in Table 3-3.

$$\begin{aligned}\eta_{jk} &= \frac{\partial \ln \rho_j(P_j)}{\partial \ln P_k} = \frac{-\beta}{\rho_j(P_j)} \cdot \int_w \psi_j (1-\psi_j) f(w \mid \overline{w}, \Omega) dw, j = k \\ &= \frac{\beta}{\rho_j(P_j)} \cdot \int_w \psi_j \psi_k f(w \mid \overline{w}, \Omega) dw, j \neq k\end{aligned} \tag{3-15}$$

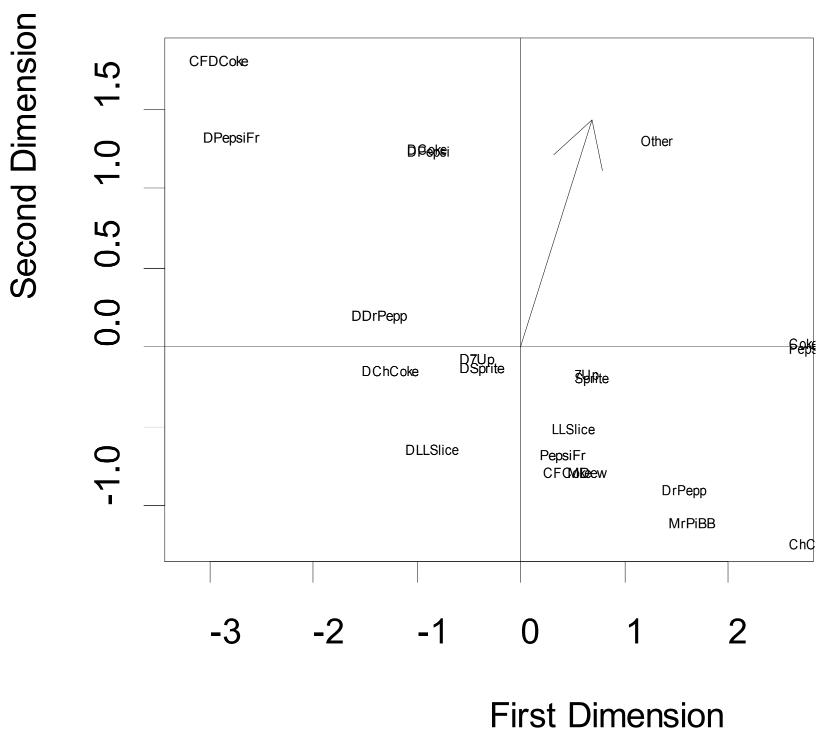
where $\psi_j = \frac{\exp(\sum_{m=1}^{M} \gamma_m \theta_{jm} - \sum_{m=1}^{M} (\theta_{jm} - w_m)^2 - \beta \ln P_j)}{\sum_{l=1}^{J} \exp(\sum_{m=1}^{M} \gamma_m \theta_{lm} - \sum_{m=1}^{M} (\theta_{lm} - w_m)^2 - \beta \ln P_l)}$, and P_j is the average price over time. The parameters used in the above equation are calculated posterior means. Since forced switching data is used to help derive these posterior means, they do not directly, but indirectly, influence the aggregate price elasticities. Again, from Table 3-3, we can see that own-price elasticities are very elastic, ranging from -2.5 to -5.6. Besides, the cross-price elasticities are all positive, a natural result from a utility-based framework, with comparable magnitudes with Dube (2005).

Table 3-3 Implied Soft Drink Cross-Price Elasticities

	Pepsi	Diet Pepsi	Mountain Dew	Pepsi Free	Diet Pepsi Free	Lemon-Lime Slice	Diet Lemon-Lime Slice	Cherry Coke	Diet Cherry Coke	Coke	Diet Coke	Caffeine-Free Coke	Caff-Free Diet Coke	Sprite	Diet Sprite	Mr. PiBB	Dr. Pepper	Diet Dr. Pepper	7-Up	Diet 7-Up	Other
Pepsi	-3.88	0.47	0.02	0.16	0.36	0.01	0.01	0.01	0.01	1.39	0.53	0.01	0.53	0.08	0.02	0.00	0.02	0.01	0.08	0.04	0.13
Diet Pepsi	0.76	-4.98	0.10	0.39	0.60	0.06	0.08	0.05	0.08	0.62	0.67	0.02	0.63	0.29	0.13	0.00	0.07	0.04	0.05	0.05	0.30
Mountain Dew	0.17	0.64	-5.41	0.58	0.63	0.19	0.25	0.17	0.21	0.13	0.48	0.06	0.46	0.58	0.38	0.01	0.13	0.10	0.04	0.05	0.14
Pepsi Free	0.48	0.72	0.17	-5.19	0.66	0.11	0.15	0.10	0.13	0.37	0.61	0.04	0.59	0.42	0.19	0.00	0.10	0.06	0.06	0.06	0.21
Diet Pepsi Free	0.68	0.72	0.12	0.42	-5.07	0.07	0.10	0.07	0.09	0.54	0.65	0.03	0.62	0.32	0.14	0.00	0.08	0.04	0.06	0.05	0.26
Lemon-Lime Slice	0.13	0.60	0.32	0.59	0.62	-5.48	0.28	0.19	0.23	0.10	0.43	0.06	0.43	0.61	0.42	0.02	0.14	0.11	0.04	0.06	0.11
Diet Lemon-Lime Slice	0.15	0.62	0.30	0.58	0.63	0.20	-5.43	0.18	0.22	0.11	0.46	0.06	0.45	0.59	0.40	0.02	0.14	0.10	0.04	0.06	0.13
Cherry Coke	0.13	0.61	0.31	0.58	0.61	0.21	0.28	-5.50	0.22	0.10	0.44	0.06	0.43	0.60	0.44	0.02	0.14	0.11	0.03	0.05	0.12
Diet Cherry Coke	0.20	0.65	0.28	0.60	0.65	0.19	0.25	0.17	-5.48	0.14	0.48	0.07	0.48	0.58	0.31	0.01	0.12	0.09	0.05	0.07	0.11
Coke	1.72	0.48	0.02	0.15	0.35	0.01	0.01	0.01	0.01	-4.24	0.55	0.01	0.52	0.07	0.02	0.00	0.02	0.01	0.05	0.03	0.20
Diet Coke	0.88	0.69	0.08	0.34	0.57	0.04	0.06	0.04	0.06	0.74	-5.02	0.02	0.63	0.24	0.10	0.00	0.07	0.03	0.05	0.04	0.34
Caffeine-Free Coke	0.34	0.64	0.25	0.62	0.67	0.18	0.23	0.15	0.22	0.19	0.46	-5.60	0.52	0.55	0.17	0.00	0.08	0.06	0.12	0.12	0.03
Caff-Free Diet Coke	0.93	0.69	0.08	0.35	0.57	0.05	0.07	0.04	0.06	0.74	0.66	0.02	-5.06	0.24	0.10	0.00	0.06	0.03	0.06	0.05	0.26
Sprite	0.29	0.70	0.23	0.55	0.66	0.15	0.20	0.13	0.17	0.23	0.56	0.05	0.54	-5.19	0.27	0.01	0.12	0.08	0.04	0.06	0.20
Diet Sprite	0.12	0.58	0.28	0.47	0.54	0.19	0.25	0.18	0.17	0.12	0.45	0.03	0.40	0.52	-5.02	0.05	0.17	0.13	0.01	0.03	0.33
Mr. PiBB	0.01	0.15	0.41	0.21	0.16	0.31	0.39	0.31	0.17	0.01	0.10	0.01	0.08	0.43	2.01	-5.45	0.24	0.28	0.00	0.00	0.16
Dr. Pepper	0.26	0.70	0.20	0.48	0.62	0.12	0.17	0.12	0.13	0.24	0.59	0.03	0.52	0.45	0.34	0.01	-5.56	0.08	0.02	0.03	0.43
Diet Dr. Pepper	0.17	0.64	0.26	0.51	0.60	0.17	0.23	0.16	0.17	0.15	0.51	0.04	0.46	0.52	0.48	0.03	0.15	-5.59	0.02	0.04	0.29
7-Up	1.40	0.57	0.06	0.33	0.53	0.04	0.05	0.03	0.06	0.74	0.51	0.05	0.61	0.19	0.03	0.00	0.03	0.01	-5.43	0.15	0.03
Diet 7-Up	0.91	0.65	0.12	0.45	0.62	0.08	0.10	0.07	0.11	0.51	0.54	0.06	0.62	0.32	0.08	0.00	0.05	0.03	0.19	-5.56	0.04
Other	0.40	0.57	0.04	0.21	0.41	0.02	0.03	0.02	0.02	0.48	0.62	0.00	0.46	0.15	0.14	0.00	0.09	0.03	0.00	0.01	-3.74

In order to see what is the information from forced-switching data versus the sales data, I also estimate a two-dimensional market map (Figure 3-3) using solely the former dataset based on the Equation (3-8a'').

Figure 3-3 Two-Dimensional Map solely from Forced Switching Drinks



3.6 Conclusion

In the current essay, I propose a utility-based model of competitive structure that accounts for and spatially represents both vertical (quality) and horizontal characteristics, where I show how the vertical characteristics govern observed asymmetric substitution patterns of the data. I estimate my model jointly with two datasets, aggregate store-level purchase data and forced switching survey (stated preference) data. To facilitate estimation with such two kinds of data, I develop a conceptual linkage between these two kinds of data. The intuition idea behind our modeling approach is that consumers' actual purchase behavior for brands gives us important market structure information regarding brands' proximities in some latent attribute space, and the same proximity relationships across brands influence the consumers' forced switching behavior. By imposing a particular distribution on latent ideal points, I am able to estimate heterogeneity in the proposed model. The model results are shown to be quite consistent with those from Chapter 2.

There are a few caveats of the model that I want to address. The first concern is the identification issue that would exist in Equation (3-1) when Equation (3-2) applies. Specifically, the first term $\nu\alpha_j$ (or $\sum_{m=1}^{M}\gamma_m\theta_{jm}$) and the linear component of the second term after being decomposed in Equation (3-1), $\sum_{m=1}^{M}2\theta_{jm}w_m$, may not be separately identified. To solve this problem, extra information is necessary. One possible source of such information is from the forced switching data. In particular, I may impose the constraint such as $\bar{w} = E(\theta_{first})$, where θ_{first} denotes the first chosen brand's location in the map.

Besides this identification issue, the second limitation is that the model only estimates the heterogeneity in ideal points, leaving willingness-to-pay and price sensitivity parameter homogeneous. This might lead to bias in estimated demand parameters and price elasticities. Since the datasets in current research evidences the perfect multicollinearity problem, estimating heterogeneity of price sensitivity is difficult. How to efficiently use the information in forced switching data to help identify the heterogeneity in these two parameters becomes a potential future research topic. Third, the model ignores the possible variety-seeking behavior among consumers. Fourth, the validity of the model relies on how multinomial assumption holds for our two data sets and on the extent to which the “single-item” assumption holds.

Despite these limitations, the proposed model appears to open up new opportunities for researchers interested in using aggregate data to explore asymmetric competitive patterns in markets.

Bibliography

Albuquerque, Paulo & Bart J. Bronnenberg (2009), "Estimating Demand Heterogeneity Using Aggregate Data: An Application to the Frozen Pizza Category," *Marketing Science*, 28 (2), 356-372.

Allenby, G.M. & P.E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics* 89, 57-78.

Anderson, S.P., A.De Palma & J.F. Thisse 1992, Discrete Choice Theory of Product Differentiation, the MIT Press.

Berry, Steven T., (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (2), 242-262.

______ , James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometric*, 63 (4).

______ , _____ , ______ (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: the New Vehicle Market," *Journal of Political Economy*, 112 (1), 68-104.

Besanko D, Dube JP, Gupta S. (2003), "Competitive Price Discrimination Strategies in a Vertical Channel using Aggregate Data," *Management Science*, 49, 1121-1138.

Bodapati, Anand V. & Sachin Gupta (2004), "The Recoverability of Segmentation Structure from Store-Level Aggregate Data," *Journal of Marketing Research*, 41(August), 351-364.

Bucklin, Randolph E., Gary J. Russell, & V. Srinivasan (1998), "A Relationship Between Market Share Elasticities and Brand Switching Probabilities," *Journal of Marketing Research*, XXXV (Feburary), 99-113.

Chen, Yuxin & Sha Yang (2007), "Estimating Disaggregate Models Using Aggregate Data Through Augmentation of Individual Choice," *Journal of Marketing Research,* 44 (November), 613-621.

Chib, Siddhartha & Edward Greenberg (1995), "Understanding the Metropolis-Hasting Algorithm," *The American Statistician*, 49 (4), 327-335.

Chintagunta, Pradeep (1998), "Inertia and Variety Seeking in a Model of Brand-Purchase Timing," *Marketing Science*, 17 (3), 253-270.

______(1999), "A Flexible Aggregate Logit Demand Model," working paper.

______ (2001), "Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data," *Marketing Science*, 20 (4), 442-456.

______, Jean-Pierre, Dube and Vishal Singh (2002), "Market Structure Across Stores: An Application of A Random Coefficients Logit Model with Store Level Data, " *Econometric Models in Marketing*, 16, 191-221.

______, _____ and _____ (2003), "Balancing Profitability and Customer Welfare in a Supermarket Chain, " *Quantitative Marketing and Economics*, 1, 111-147.

______, Tulin Erdem, Peter E.Rossi & Michel Wedel (2006), "Structural Modeling in Marketing: Review and Assessment," *Marketing Science*, 25 (6), 604-616.

Desarbo, Wayne S. & Ajay K. Manrai (1992), "A New Multidimensional Scaling Methodology for The Analysis of Asymmetric Proximity Data in Marketing Research," *Marketing Science*, 11(1), 1-20.

______ & Rajdeep Grewal (2007), “An Alternative Efficient Representation of Demand-Based Competitive Asymmetry,” *Strategic Management Journal*, 28, 755-766.

Dube, Jean-Pierre, Bart Bronnenberg, Ron Goettler, P.B. Seetharaman, K. Sudhir, Raphael Thomadsen & Ying Zhao (2002), “Structural Applications of the Discrete Choice Model,” *Marketing Letters* 13 (3), 207-220.

Draganska, Mikhaila & Dipak C. Jain (2002), “A Likelihood Approach to Estimating Market Equilibrium Models,” *Management Science*, 50 (5), 605-616.

Elrod, Terry (1988a), “Choice Map: Inferring a Product-Market Map From Panel Data,” *Marketing Science* 7 (1), 21-40.

______ (1988b), “Inferring an Ideal-Point Product-Market Map From Consumer Panel Data,” *Data, Expert Knowledge and Decisions, 240-249*.

______ (1991), “Internal Analysis of Market Structure: Recent Developments and Future Prospects,” *Marketing Letters,* 2(3), 253-266.

______, and Michael P. Keane (1995), “A Factor-Analytic Probit Model for Representing the Market Sructure in Panel Data,” *Journal of Marketing Research* 32, 1-16.

Gupta, S., P.K., Chintagunta, A.Kaul & D.R. Wittink (1996), “Do Household Scanner Data Provide Representative Inferences from Brand Choices: A Comparison with Store Data,” *Journal of Marketing Research* 33, 383-398.

Goldberg, Pinelopi Koujianou (1995), “Product Differentiation and Oligopoly in

International Markets: the Case of the U.S. Automobile Industry," *Econometrica*, 63 (4), 891-951.

Jiang, Renna, Puneet Manchanda & Peter E. Rossi (2009), "Bayesian Analysis of Random Coefficient Logit Models Using Aggregate Data," *Journal of Econometrics*, 149, 136-148.

Kamakura, Wagner A. & Rajendra K. Srivastava (1986), "An Ideal-Point Probabilistic Choice Model for Heterogeneous Preferences," *Marketing Science*, 5 (3), 199-218.

Kim Byung-Do (1995), "Incorporating Heterogeneity with Store-Level Aggregate Data," *Marketing Letters* 6:2: 159-169.

Lancaster, Kelvin J. (1966), "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74 (2), 132-157.

Lehmann, Donald (1972), "Judged Similarity and Brand-Switching Data," *Journal of Marketing Research*, IX(August), 331-334.

Lee, Jack K.H., K. Sudhir & Joel H.Steckel (2002), "A Multiple Ideal Point Model: Capturing Multiple Preference Effects from Within an Ideal Point Framework," *Journal of Marketing Research*, XXXIX (February), 73-86.

Mazzeo, Michael (2006), "Invited Commentary: Marketing Structural Models: 'Keep It Real'," *Marketing Science* 25 (6), 617-619.

Musalem, Andres, Eric T. Bradlow & Jagmohan S. Raju (2009), "Bayesian Estimation of Random-Coefficients Choice Models Using Aggregate Data," *Journal of Applied Econometrics*, 24, 490-516.

Nevo, Aviv (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand*," Journal of Economics & Management Strategy*, 9 (4), 513-548.

______ (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69 (2), 307-342.

Novak, Thomas P. & Charles Stangor (1987), "Testing Competitive Market Structures: An Application of Weighted Least Squares Methodology to Brand Switching Data," *Marketing Science*, 6 (1), 82-97.

Park, Sungho & Sachin Gupta (2009), "Simulated Maximum Likelihood Estimator for the Random Coefficient Logit Model Using Aggregate Data," *Journal of Marketing Research*, XLVI (August), 531-542.

Petrin, Amil (), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110 (4), 705-729.

______ & Kenneth Train (2004), "Omitted Product Attributes in Discrete Choice Models," working paper, Booth Graduate School of Business, University of Chicago.

Seetharaman, P.B. (2004), "Estimating Disaggregate Heterogeneity Distributions Using Aggregate Data: A Likelihood-Based Approach," working paper, Olin School of Business, Washington University at St. Louis.

Shepard, R.N. (1957), "Stimulus and Response Generalization: A Stochastic Model relating Generalization to Distance in Psychological Space," *Psychometrika* 22 (4).

Shocker, A.D. and V. Srinivasan (1979), "Multiattribute Approaches for Product Concept Evaluation and Generation: A Critical Review," *Journal of*

Marketing Research, 16, 159-180.

Sudhir, K. (2001), "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science*, 20 (1), 42-60.

Sutton, John (1986), "Vertical Product Differentiation: Some Basic Themes," *The American Economic Review*, 76 (2), 393-398.

Urban, G. L., P. L. Johnson & J. R. Hauser (1984), "Testing Competitive Market Structures," *Marketing Science*, 3 (2), 83-112.

Vandenbosch, Mark B. & Charles B. Weinberg (1995), "Product and Price Competition in a Two-Dimensional Vertical Differentiation Model," *Marketing Science*, 14 (2), 224-249.

Villas-Boas, J. Miguel & Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324-1338.

Zenor, M. & R.K.Srivastava (1993), "Inferring Market Structure with Aggregate Data: A Latent Segment Logit Approach," *Journal of Marketing Research* 30, 369-379.

Chapter 4 General Conclusions

4.1 Conclusion of the dissertation

4.1.1 Summary

Understanding the extent to which different brands compete with each other in a pre-defined product market is important to marketing managers, researchers and policy makers. The market structure analysis facilitates this understanding via constructing the market map which locates brands in a multi-attribute space. The introduction of the scanning system in mid-1970s provided easy access to the store-level aggregate data which opened tremendous possibilities for developing aggregate data-based tools of market structure analysis. The purpose of this dissertation is to exploit the potential of store-level aggregate data in identifying the pattern of inter-brand rivalry.

Two methods have been proposed in current dissertation for this purpose. Built upon either a reduced-form model or a utility-based formulation, these two methods differ mainly on what information is used to generate market maps: the cross-price elasticity matrix or estimated utility paramters. Despite this difference, they share some common features: (a) both methods are unified (i.e., one-stage), in the sense that the estimation of demand function and the derivation of a market map are achieved in one step; (b) both methods provide internal analysis of market structure, so the number of the dimensions in the market map and the location of the brands on these dimensions are determined by the aggregate data (i.e., aggregate sales data in Chapter 2; aggregate sales data and aggregate choice data in Chapter 3); (c) both methods explicitly address the asymmetric feature of competition, and represent it in the same market space that is typically used to account for symmetric proximity data.

In particular, the first method, built upon a sales response model, estimates a market map that is assumed to jointly underlie cross-price elasticities, own-price elasticities, and brand-specific intercepts. Drawing upon recent psychometric research, I express the asymmetries present in cross-price elasticities as the difference between what I refer to as brand power parameters. Two propositions are proposed to identify relationships between a focal brand's power parameter, clout, vulnerability, own-price elasticity, and spatial density. I apply the model separately for two datasets that consist of weekly sales and prices for beer and soft drinks. One main finding is that the comparison between the proposed model and the saturated model favors the former. Besides, the market maps generated from both data sets provide interpretable results.

The second method, built upon a utility-based structure, derives a market map that can account for both horizontal characteristics (i.e., attribures that are horizontally differentiated) and vertical characteristics (attributes evidencing vertical differentiation) in the utility function. I show how the combination of aggregate scanner data and forced switching data can help estimate a market map directly from the utility function while accounting for consumer heterogeneity. Conceptually, the proposed model shows how the asymmetric feature of competition, represented by the asymmetric switching pattern in forced switching data is related to the vertically differentiated attributes in the utility function. Applying the model to weekly sales data and forced switching data for soft drinks, I recover the demand system.

4.1.2 Comparison of Two Models

Generally, there are at least six criteria that one can use to assess the relative advantages and disadvantages of different models: (a) a more parsimonious model is better; (b) a model with good agreement with data (fit) is better; (c) a model built upon the theory is better; (d) a model requiring fewer assumptions is better; (e) a model requiring less data is better; and (f) a model that can answer a research question directly is better than indirectly doing so (Vapnik 1998). I apply these criteria sequentially to the two models of this dissertation as follows.

First, both methods achieve similar levels of parsimony: for the soft drink data, for example, the two-dimensional model in Chapter 2 relies on 47 focal parameters, the same number of parameters as the corresponding model in Chapter 3.

Second, a direct comparison between the deviance of the soft drink data in Chapter 2 and Chapter 3 implies that the utility-based model is better (e.g., the deviances of two-dimensional model specifications for soft drink data are 76,320 in Chapter 2 and 65,986.05 in Chapter 3). However, this difference is probably due to the fact that the Chapter 3 analysis explains sales conditional on observed category volume (by using the Multinomial distribution), whereas the Chapter 2 analysis explains category volume as well (by assuming a Poisson distribution). In fact, despite this difference, these two methods estimate brand positioning maps with quite consistent market structure results. For example, both Figure 2-4 and Figure 3-2 indicate the following: (i) Pepsi and Coke are the head-to-head competitors; (ii) Mr. PiBB is an market outlier; (iii) products within the Dr. Pepper product line or the Sprite Product line compete with each other, although Figure 2-4 seems to further differentiate this submarket into (Diet Sprite vs. Diet Dr. Pepper) and (Sprite vs. Dr. Pepper); (iv) Lemonlime Slice products, Mountain Dew, and Cherry Coke products belong to the same market, although, again, Figure 2-4 seems to provide further differentiation; (v) major Diet Coke products (Diet Pepsi, Diet Coke, Caffeine-free Diet Pepsi) are close competitors.

Third, in terms of the criteria of (c) (theory driven) and (d) (fewer assumptions), each approach has its own strengths and weaknesses.

In particular, the method proposed in Chapter 2, as a typical reduced-form model, prioritizes fit to the empirical data and let the data "speak." Furthermore, it relies on few behavioral assumptions and allows for flexible functional forms. By assuming that the price-elasticities and intercepts are driven by some latent brand attributes, the researchers are able to, without the need of re-estimating the demand function every time, predict the potential optimum competitive position (e.g., defined by the power, the clout and the vulnerability) of a new entrant, and its impact on the existing brands' own-price elasticities (e.g., through the concept of spatial density). When the research focus is on analyzing the marketing structure, this method therefore can avoid the "static measure" problem[36] of the cross-price elasticities to some extent. This idea would also be of importance to antitrust authorities. In its *Notice to Agreements of Minor Importance*, the

[36] Cross-price elasticities are said to be "*static*" measures, since over time new entrants or departures from a market may affect the cross-elasticity between any two alternatives.

European Court Commission (1986) stated that deciding whether products are interchangeable "must be judged from the vantage point of the user, normally taking the characteristics, price and intended use of the goods together." (Massey 2000).

On the other hand, the utility-based approach proposed in Chapter 3 is theory-based and more capable of explaining cross-price elasticities as the result of consumers' optimizing behavior. By directly relating the market map with consumers' preference parameters, the calibrated market structure is more intuitive and stable (om the face of changes in brand offerings) than the method in Chapter 2. The main problem of this method, however, is that the model relieson more behavioral assumptions than the first model. It imposes relatively strong assumptions on consumers' heterogeneity parameters. And it is more suitably applied to cases where demand for a set of goods is mutually exclusive. In some product categories, where consumers may be variety-seeking, imposing the single-unit, single-brand purchase assumption of the discrete choice model will generate biased consumer responses to marketing mix variables (Hendel 1999; Kim et al. 2002, 2006; Dube 2004,2005; Bhat & Sen 2006). Besides, the model also relies on the heterogeneity distribution assumption. Therefore, the performance of this utility-based approach does rely on the validity of the underlying assumptions.

Fourth, the method in Chapter 2 has a lower data requirement. The data (i.e., the soft drink sales data) used in this dissertation suffers from a severe multicollinearity problem. The model in Chapter 2 can estimate the demand function and the market map with only aggregate sale data. However, this is not the case in Chapter 3. It is well-known that estimating the heterogeneity parameters from a discrete choice demand function can be challenging with only aggregate sales data (Albuquerque and Bronnenberg 2006, Petrin 2002), especially when the data contains limited information (Kim 1995). To overcome this problem, researchers typically combine the aggregate data and consumer-level data (e.g., Goldberg 1995; Petrin 2002; Albuquerque and Bronnenberg 2005). For my case, the perfect multicollinearity in the data aggravates this difficulty of estimation. In order to recover the true distribution of heterogeneity, adding a second data source becomes a necessity in the current context.

Finally, I believe the choice between these two methods should be made based upon the researcher's primary research interest (underlying the need for a market map). When the interest is to understand a brand's market performance, such as its power, its vulnerability, or to estimate the price elasticities, say, for antitrust purpose, the method in Chapter 2 would be preferred. When the research interest is on understanding consumers' intrinsic preferences, and other parameters of the utility function, or inferences at the individual or segment level, the method in Chapter 3 might be better.

In summary, it is hard to tell which method proposed in the dissertation is the better one, as each wins for certain criteria, but loses on others. This dissertation is the first analysis, as far as I know, which provides a side-by-side comparison of a reduced-form-based market structure analysis with a utility-based market structure analysis.

4.1.3 Contributions Revisited

The dissertation provides conceptual, methodological and managerial contributions to the aggregate data-based market structure analysis literature. Conceptually, the dissertation emphasizes the importance of utilizing as much market structure information as possible in deriving market maps, an issue that has not been given enough attention in past literature. A complete account of competitive information in a market structure map maximizes efficient information usage and increases the usability of the derived market maps.

Furthermore, the current dissertation accommodates both vector formulation and ideal-point formulation to derive market maps, and shows how each formulation (in the first method) or the combination of both (in the second approach) help us formalize theoretical linkages among many familiar competitive structure concepts: (a) competitive asymmetry, brand power, clout, vulnerability; (b) own-price elasticity, vulnerability, spatial density; (c) vertical differentiation, horizontal differentiation, asymmetric brand switching, etc. And the dissertation draws upon the knowledge from not only marketing literature, but also economic and psychometric literature.

Methodologically, the Markov Chain Monte Carlo (MCMC) methods are adopted throughout the dissertation to help simulate complex, nonstandard multivariate distribution. The adaptive Bayesian shrinkage approach is used in a way that enables consistent information sharing across brands in the demand function by using weakly informative priors for hyperparameters throughout.

This shrinkage idea, combined with the information sharing obtained by utilizing all market structure information embedded in major brand-specific parameters, help stabilizing the estimates and achieves parsimony in the estimation process.

Managerially, the proposed methods enable managers with access to scanner data at the store level to track changes over time in asymmetric competitive structure; the methods are also useful for predicting the demand for existing and new products offered in existing markets. Specifically, the market map derived in the utility-based approach indicates the ideal points of consumers in terms of horizontal characteristics and the ideal vector of quality. Brands that are closer to the ideal points and project farther out on the quality vector are more preferable for consumers and, therefore, tend to yield highest market share. For the reduced-form-based model, the market map also provides the position of a hypothetical brand with the highest dominance and another hypothetical brand with the least vulnerability in the market. Any brand that is closer to both hypothetical brands should be more optimal to marketers. Further, analysts can use this method to examine strategy scenarios involving changing competitive positioning in order to predict the impact on demand intercepts, own-price elasticities, and cross-elasticities.

4.2 Limitations of Proposed Methods

Besides the limitations that have already been discussed at the close of each chapter, other limitations that are important and sometimes common to both methods include the following:

Interpreting the axes of the map can be difficult. The proposed methods are internal analysis of market structure, therefore the axes of market maps derived need to be labeled by some external means. To label the axes, researchers can examine the location of the brands in the map and check for correlation with the brands' observed attributes. Or a more systematic labeling procedure may be conducted: for example, model the brands' location as a linear function of the observed attributes. However, neither method guarantees the existence of a satisfactory interpretation of the axes.

The assumption of stationary preferences is restrictive. Both methods proposed in current dissertation assume stationarity of market maps over time, which implies that consumers' preferences are also stable. This assumption may be violated by factors such as consumers' variety-seeking behavior. To reduce the chance of seriously violating this assumption, the period of observations should be restricted to relatively short periods of time, and frequently-purchased products may be suitable product categories for study.

The imposed distribution governing consumers' ideal-points may be misspecified. In Chapter 3, I assume consumers' ideal-point on each dimension of the market map is normal distributed. This assumption may be violated when multiple ideal-points exist in the population. Under this situation, a mixture of normal priors for consumers' preferences may be more appropriate.

4.3 Directions for Future Research

The dissertation raises the following potential future research directions:

Generalizing the proposed methods to chain-level data. The proposed methods in current dissertation are developed to deal with store-level data.

However, much aggregate data available to marketers are actually chain-level. From a manager's perspective, a market structure analysis that identifies store-specific inter-brand rivalry enables micromarketing strategies and therefore is usually more desirable. Previous research provide various insights in this area (e.g, Chintagunta et al. 2002, 2003; Montgomery 1997; Montgomery & Rossi 1999). Combining my proposed methods with the ideas of this past work to facilitate chain-level market structure analysis would constitute an interesting future research area.

Enriching the utility-based model with more flexible distribution assumptions for consumer heterogeneity. When aggregate data is used, consumer heterogeneity has been handled in different ways in previous literature. Some previous research estimate heterogeneity using latent class models, assuming the existence of a finite number of segments with their own parameters (e.g., Besanko, Dubé, and Gupta 2003; Draganska and Jain 2002; Seetharaman 2001; Zenor and Srivastava 1993). Others specify a particular functional form for the parameters of the heterogeneity distribution and use consumer-level data to help estimate these parameters (e.g., Goldberg 1995; Petrin 2002; Albuquerque and Bronnenberg 2005). When conducting internal analysis of market structure, Chintagunta et al. (2002, 2003), similar to my model, impose a multivariate normal distribution on consumers' brand perceptions. One interesting extension would be to assume that brand perceptions are drawn from a mixture of normal distributions, with parameters of these distributions following some function of consumer demographic information.

Utilizing the information in forced-switching data to help estimate the heterogeneity in the willingness-to-pay parameters. One limitation of my proposed model in Chapter 3 is that the willingness-to-pay parameter is treated as constant across consumers. Estimating heterogeneity in this parameter can be challenging since it enters the utility functions for both aggregate data and the forced-switching data. How to efficiently use the information in forced-switching data to help estimate the heterogeneity in this parameter then becomes a future extension of current research.

Accounting for the endogeneity problem in the utility function. Some of the previous research that tries to utilize aggregate data to estimate a utility-based demand function deals with not only heterogeneity but also endogeneity (i.e., a situation where the marketing-mix variables could be correlated with the error terms in the latent utilities (Villas-Boas & Winer 1999). Failure to account for this endogeneity can bias the parameter estimates of the marketing-mix variables (BLP 1995). The current dissertation does not deal with this problem due to the lack of data, however, studying how my proposed methods can be combined with approaches proposed in literature to deal with both heterogeneity and endogeneity problems would be desirable.

Overall, this dissertation adds to the market structure analysis literature by proposing two aggregate data-based techniques to help managers, researchers and policy makers understand asymmetric competition among brands. I look forward to continued progress on unified analysis of asymmetric market structure.

Bibliography

Albuquerque, Paulo & Bart J. Bronnenberg (2009), "Estimating Demand Heterogeneity Using Aggregate Data: An Application to the Frozen Pizza Category," *Marketing Science*, 28 (2), 356-372.

Besanko D, Dube JP, Gupta S. (2003), "Competitive Price Discrimination Strategies in a Vertical Channel using Aggregate Data," *Management Science*, 49, 1121-1138.

Berry, Steven T., James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometric*, 63 (4).

Bhat, Chandra R., Sen, Sudeshna (2006), "Household Vehicle Type Holdings and Usage: An Application of the Multiple Discrete-Continuous Extreme Value Model," *Transportation Research*, 40, 35-53

Chintagunta, Pradeep., Jean-Pierre, Dube and Vishal Singh (2002), "Market Structure Across Stores: An Application of A Random Coefficients Logit Model with Store Level Data, " *Econometric Models in Marketing*, 16, 191-221.

_____, _____, _____ (2003), "Balancing Profitability and Customer Welfare in a Supermarket Chain, " *Quantitative Marketing and Economics*, 1, 111-147.

Draganska, Mikhaila & Dipak C. Jain (2002), "A Likelihood Approach to Estimating Market Equilibrium Models," *Management Science*, 50 (5), 605-616.

Dube, Jean-Pierre. (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science,* 23, 66-81.

_____ (2005), "Product Differentiation and Mergers in the Carbonated Soft Drink

Industry," *Journal of Economics & Management Strategy*, 14(Winter), 879-904.

Goldberg, Pinelopi Koujianou (1995), "Product Differentiation and Oligopoly in International Markets: the Case of the U.S. Automobile Industry," *Econometrica*, 63 (4), 891-951.

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

Kim Byund-Do (1995), "Incorporating Heterogeneity with Store-Level Aggregate Data," *Marketing Letters* 6:2: 159-169.

Kim, Jaehwan, Allenby, Greg M., Rossi, Peter E. (2002), "Modeling Consumer Demand for Variety," *Marketing Science*, 21 (3), 229-250.

_____, _____, _____ (2006), Product Attributes and Models of Multiple Discreteness," *Journal of Econometrics*, 5 (December), 1-23.

Massey, Patrick (2000), "Market Definition and Market Power in Competitive Analysis: Some Practical Issues," *The Economic and Social Review*, 31(4), 309-328

Montgomery, Alan L. (1997), "Creating Micro–Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science*, 16(4), 315–337.

Montgomery, Alan L. and Rossi, Peter E.(1999), "Estimating Price Elasticities with Theory–based Priors", *Journal of Marketing Research*, 36 (4), 413–423.

Petrin, Amil (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110 (4), 705-729.

Seetharaman, P.B. (2001), "Estimating Disaggregate Heterogeneity Distributions Using Aggregate Data: A Likelihood-Based Approach," working paper, Olin School of Business, Washington University at St. Louis.

Vapnik, Vladimir N. *Statistical Learning Theory*. New York Wiley

Villas-Boas, J. Miguel & Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324-1338.

Zenor, M. & R.K.Srivastava (1993), "Inferring Market Structure with Aggregate Data: A Latent Segment Logit Approach," *Journal of Marketing Research* 30, 369-379.

Appendix A. Variable Definitions for Chapter 2

Focal Variables

p_{it} price of brand i at time t (log prices are mean centered)

q_{it} unit sales of brand i at time t

Demand Equation Parameters

α_0 overall intercept of demand function

α_i brand-specific intercept (mean-centered)

β_{ii} own-price elasticity of brand i

β_{ij} cross-price elasticity between brands i and j

CV_{it} covariate in i^{th} brand equation at time t

f functional form of covariate term(s)

Latent Structural Parameters

θ_{im} location of brand i on mth dimension

$\boldsymbol{\theta}_i$ location (vector) of brand i

g specification linking β_{ij} and the relationship between $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$

β specification linking β_{ii} and $\boldsymbol{\theta}_i$

α specification linking α_i and $\boldsymbol{\theta}_i$

B (capital beta) asymmetric matrix of cross-price elasticities, with diagonal elements zero

s_{ij} $\equiv (\beta_{ij} + \beta_{ji})/2$

S symmetric matrix with typical element s_{ij}

a_{ij} $\equiv (\beta_{ij} - \beta_{ji})/2$

A skew-symmetric matrix with typical element a_{ij}

Indices

n number of brands in market

i index of brand, $i = 1, \ldots, n$

j index of brand, $i = 1, \ldots, n$

T number of Weeks in data set

t index of Week, $t = 1, \ldots, T$

t' is mean-centered version of index t

M number of dimensions in structural map

m index of map dimension, $m = 1, \ldots, M$

Covariates (time dependent)

γ_1 the coefficient for sine time trend

γ_2 the coefficient for cosine time trend

γ_3 the coefficient for linear time trend

Underlying Structural Model Parameters

ϕ_{CPE} constant in specification for β_{ij}

ϕ_{OPE} constant in specification for β_{ii}

d_{ij} inter-point distance between $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$

x_i brand power parameter for brand i

ω_0 Weight in brand-specific intercept

ω_1 Weight of dominance-point model for x_i

ω_2 Weight of ideal-point model for β_{ii}

$Y = [y_m]$ most powerful location, $m = 1, \ldots, M$

$Z = [z_m]$ least vulnerable location, $m = 1, \ldots, M$

υ_{1m} m^{th} coefficient of vector model for x_i

υ_{2m} m^{th} coefficient of vector model for β_{ii}

Distribution Parameters

λ_{ij} the rate parameter of poison distribution

η_{ij} expected value for gamma distribution

ρ The shape parameter of gamma distribution

Hyper Parameters

μ_m hyper-parameter (prior mean) for brands' coordinates on *mth* dimension

σ_m hyper-parameter (prior standard deviation) for brands' coordinates on *mth* dimension

Brand-Related Measure

$$Clout_i \equiv \sum_{j \neq i} \beta_{ji} \,, Vul_i \equiv \sum_{j \neq i} \beta_{ij}$$

$$Density_i \equiv (\phi_{CPE} - \sum_{j \neq i} d_{ij} / (n-1))$$

Appendix B. Proofs of Propositions in Chapter 2

Proof of Proposition 1. (a) This fact was established in the discussion following (2-4), which I restate here. Suppose $x_i > x_j$. Equation (2-4) states that $\beta_{ij} = \phi_1 - d_{ij} - x_i + x_j$. Then I have

$$\beta_{ij} = \phi_1 - d_{ij} - x_i + x_j < \phi_1 - d_{ij} - x_j + x_i = \phi_1 - d_{ji} - x_j + x_i = \beta_{ji}.$$

The first and last equalities are restatements of Equation (2-4); the inequality arises from $x_i > x_j$; and the next to last equality arises from the symmetry of $d_{ij} = d_{ji}$. So $\beta_{ij} < \beta_{ji},\ \ i \neq j$.

(b) From the definition of clout and vulnerability, and after substitution (2-4),

$$Clout_i = \sum_{j \neq i}^{n} \beta_{ji} = [\sum_{j \neq i}^{n} (\phi_1 - d_{ji} - x_j + x_i)]$$

$$= (n-1)\phi_1 - \sum_{j \neq i}^{n} d_{ji} - \sum_{j \neq i}^{n} x_j + (n-1)x_i$$

$$Vul_i = \sum_{j \neq i}^{n} \beta_{ij} = \sum_{j \neq i}^{n} (\phi_1 - d_{ij} - x_i + x_j)$$

$$= (n-1)\phi_1 - \sum_{j \neq i}^{n} d_{ij} - (n-1)x_i + \sum_{j \neq i}^{n} x_j$$

Subtracting the second equation from the first yields

$$Clout_i - Vul_i = \{(n-1)\phi_1 - \sum_{j \neq i}^{n} d_{ji} - \sum_{j \neq i}^{n} x_j + (n-1)x_i\} - \{(n-1)\phi_1 - \sum_{j \neq i}^{n} d_{ij} - (n-1)x_i + \sum_{j \neq i}^{n} x_j\}$$

$$= -2\sum_{j \neq i}^{n} x_j + 2(n-1)x_i = -2\sum_{j=1}^{n} x_j + 2nx_i.$$

When $\sum_i x_i = 0$, $Clout_i - Vul_i = 2nx_i$. So $x_i = \dfrac{Clout_i - Vul_i}{2n}$.

(c) [in Footnote 13] The additive similarity-bias model can be transformed into skew-symmetric model by properly decomposing its bias component.

$$\bar{s}_{ij} + r_i + c_j = \bar{s}_{ij} + \frac{(r_i + c_i)}{2} + \frac{(r_j + c_j)}{2} - \frac{(c_i - r_i)}{2} + \frac{(c_j - r_j)}{2} = s_{ij} - x_i + x_j,$$

where $s_{ij} = \bar{s}_{ij} + \dfrac{(r_i + c_i)}{2} + \dfrac{(r_j + c_j)}{2}$, and $x_i = \dfrac{(c_i - r_i)}{2}$. Thus, I can interpret each brand's dominance effect (i.e., x_i) as the half of the difference between its column and row bias (i.e., $x_i = \frac{c_i - r_i}{2}$). Done.

Proof of Proposition 2. (a) Assuming utility-maximizing consumers under a linear budget constraint, I evoke the microeconomic property of demand homogeneity of degree zero in prices and income (because, for example, doubling all prices and incomes would have no real effect). This property arises directly from standard utility maximization under a linear budget constraint, and it can be shown to apply at the individual consumer level and at the market level. This property implies that, for any good, the sum of the own price elasticity and all of cross price elasticities equals minus the income elasticity. That is,

$$\beta_{ii} + \sum_{j \neq i} \beta_{ij} + g_i = 0 .$$

Note that $\sum_{j \neq i} \beta_{ij}$ is my definition of vulnerability of brand *i*. Thus, I have (2-8),

$$-\beta_{ii} = (\sum_{j \neq i} \beta_{ij}) + g_i = Vul_i + g_i . \qquad (2\text{-}8)$$

Incidentally, for some categories it is approximately true that $g_i = 0$,

(b) From (2-4) and (2-8), I have

$$\begin{aligned} -\beta_{ii} &= \sum_{j \neq i} \beta_{ij} + g_i = \sum_{j \neq i} (\phi_{CPE} - d_{ij} - x_i + x_j) + g_i = \sum_{j \neq i} \phi_{CPE} - \sum_{j \neq i} d_{ij} - \sum_{j \neq i} x_i + \sum_{j \neq i} x_j + g_i \\ &= (n-1)\phi_{CPE} - \sum_{j \neq i} d_{ij} - \sum_{j \neq i} x_i - x_i + \sum_{j \neq i} x_j + x_i + g_i \\ &= (n-1)\phi_{CPE} - \sum_{j \neq i} d_{ij} - nx_i + \sum_{j=1}^{n} x_j + g_i \\ &= (n-1) Density_i - nx_i + \beta_{iI} + \sum_{j} x_j , \end{aligned}$$

where $Density_i \equiv (\phi_{CPE} - \sum_{j \neq i} d_{ij} / (n-1))$. And, without loss of generality, if $\sum_{j=1}^{n} x_j = 0$, we have (2-9), which I repeat below:

$$-\beta_{ii} = (n-1) Density_i - nx_i + g_i . \qquad (2\text{-}9)$$

Appendix C. Variable Definitions for Chapter 3

Focal Variables

p_{it} price of brand i at time t

q_{it} unit sales of brand i at time t

Utility Function Parameters

u utility level

v willingness-to-pay for consumer i

w_{im} consumer i's ideal point on mth dimension

β consumers' price sensitivity

ε utility function residual

γ_m the weight of quality measure on the mth horizontal characteristic.

ρ the choice probability of aggregate data

η the price elasticities

Latent Structural Parameters

Θ location matrix

θ_{im} location of brand i on mth dimension

$\boldsymbol{\theta}_i$ location (vector) of brand i

α_j brand j's quality

Indices

J number of brands in market

j,k index of brand, $j,k = 1,...,J$

i index of consumers, $i = 1,...,I$

T number of weeks in data set

t index of week, $t = 1,\ldots,T$

M number of dimensions in structural map

m index of map dimension, $m = 1,\ldots,M$

Q_t the total number of unit sales at week t

N_k the number of times that brand k is chosen as the first option in forced switching data

N_{kj} the number of times that brand j is chosen instead when k was not on the shelf in forced switching data

Distribution Parameters

$\overline{w}$ the mean vector for ideal-point parameters

σ_m^2 the dispersion parameter for ideal-point parameters on mth dimension

Ω the distribution parameter for w_{im}

π the distribution function for forced switching probability

f the distribution function for

heterogeneity parameters

ϕ the prior distribution function

Pr the density function for multinomial distribution

L the likelihood function

Hyper Parameters

$\tau_{.}$ hyper-parameter (prior standard deviation) for various parameters

Brand-Related Measure

Weight for brand j at time t

$W_{jt} = \exp(\nu\alpha_j - \beta \ln p_{jt})$

Appendix D. Implied Product Interdependency in Chapter 3

The discussion is similar to Kamakura & Srivastava (1986). Specifically, to show the effect of preference heterogeneity on product substitution pattern, first remember that the utility function has the following form:

$$u_{ijt} = v_i \alpha_j - \sum_{m=1}^{M} (\theta_{jm} - w_{im})^2 - \beta \ln p_{jt} + \varepsilon_{ijt},$$

where ε_{ijt} is distributed independently according to a type I Extreme Value distribution with mean zero and unit scale parameter. And I assume some distributions for consumer heterogeneity parameters. In particular, since α_j is a latent variable, without loss of generality, we can assume $v_i \sim N(v,1)$. Further I assume a multivariate normal distribution of ideal points, that is, we can write $w_{im} = w_m + e_{im}$, where

w_m = mean coordinate of ideal points on latent dimension m;

e_{im} = random disturbance for consumer i, relating to mth dimension. And these disturbances follow a multivariate normal distribution with mean zero and a diagonal covariance matrix (again, this form of covariance matrix is assumed without loss of generality (Elrod, 1988b)).

$$\Omega = \begin{bmatrix} \sigma_1^2 & & & \\ 0 & \sigma_2^2 & & \\ 0 & 0 & .... & \\ 0 & 0 & 0 & \sigma_M^2 \end{bmatrix}$$

Finally, I assume $\text{cov}(v_i, e_{im}) = 0, \forall m = 1,..., M$. Now the utility function becomes

$$u_{ijt} = v_i \alpha_j - \sum_{m=1}^{M} (\theta_{jm} - w_m - e_{im})^2 - \beta \ln p_{jt} + \varepsilon_{ijt}$$

$$= -\beta \ln p_{jt} + v_i \alpha_j - \sum_{m=1}^{M} (\theta_{jm}^2 + w_m^2 + e_{im}^2 - 2 w_m \theta_{jm} - 2 e_{im} \theta_{jm} + 2 w_m e_{im}) + \varepsilon_{ijt}$$

$$= -\beta \ln p_{jt} - \sum_{m=1}^{M} \theta_{jm}^2 - \sum_{m=1}^{M} w_m^2 + 2\sum_{m=1}^{M} w_m \theta_{jm} - 2\sum_{m=1}^{M} w_m e_{im} - \sum_{m=1}^{M} e_{im}^2$$

$$+ v_i \alpha_j + 2\sum_{m=1}^{M} e_{im} \theta_{jm} + \varepsilon_{ijt}$$

So the randomness of the utility function enters into the utility function mainly through the last three components (i.e., $v_i\alpha_j + 2\sum_{m=1}^{M} e_{im}\theta_{jm} + \varepsilon_{ijt}$) of the right-hand side equation. Notice, since the term $-\sum_{m=1}^{M} w_m{}^2 - 2\sum_{m=1}^{M} w_m e_{im} - \sum_{m=1}^{M} e_{im}{}^2$ enters all alternatives' utility function, it is easy to show that the proposed model is equivalent to a multinomial logit model with utility function for alternative j equals to

$$u_{ijt}' = -\beta \ln p_{jt} - \sum_{m=1}^{M} \theta_{jm}{}^2 + 2\sum_{m=1}^{M} w_m\theta_{jm} + v_i\alpha_j + 2\sum_{m=1}^{M} e_{im}\theta_{jm} + \varepsilon_{ijt}$$

with

$$\operatorname{cov}(u_{ijt}', u_{ikt}') = \operatorname{cov}\{(v_i\alpha_j + 2\sum_{m=1}^{M} e_{im}\theta_{jm} + \varepsilon_{ijt}), (v_i\alpha_k + 2\sum_{m=1}^{M} e_{im}\theta_{km} + \varepsilon_{ikt})\}$$

$$= \alpha_j\alpha_k + 4\sum_{m=1}^{M} \sigma_m^2\theta_{jm}\theta_{km}$$

Therefore, the choice probability for a given alternative j will depend not only on its own position on the structural map ($\theta_{jm},\ m = 1,...,M$) and its quality level (α_j , if it is unrelated with $\theta_{jm},\ m = 1,...,M$), but also on the other alternatives' position and quality level.

Appendix E. Markov Chain Monte Carlo Algorithm for the Proposed Model in Chapter 3

The joint posterior of all parameters in the model is given by

$$p(\gamma,\beta,\Theta,\overline{w},\sigma_1^2,...,\sigma_M^2,\tau_1^2,...,\tau_M^2 \,|\, \underline{q},\underline{N}) \propto$$
$$\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(p_{jt};\gamma,\Theta,\beta,\overline{w},\sigma_1^2,...,\sigma_M^2)^{q_{jt}}\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(N_{kj};\gamma,\Theta)^{N_{kj}}$$
$$(\prod_{m=1}^{M}\phi(\gamma_m))\phi(\Theta\,|\,\tau_m^2, m=1,..,M)\phi(\beta)(\prod_{m=1}^{M}\phi(\overline{w}_m))(\prod_{m=1}^{M}\phi(\sigma_m^2))$$
$$\prod_{m=1}^{M}\phi(\tau_m^2)$$

The estimation of the model parameters proceeds by recursively sampling from the following distributions:

First consider demand parameters (β , $\gamma_{M\times 1}$, Θ).

1. Generating β

The posterior of β conditional on all other parameters is proportional to

$$\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\beta,rest)^{q_{jt}}\}\phi(\beta)$$

where *rest* means other parameters in the likelihood. I use Metropolis-Hasting algorithm to generate draws of β with a random walk chain. Let $\beta^{(p)}$ denote the previous draw, and then the next draw $\beta^{(n)}$ is given by

$$\beta^{(n)} = \beta^{(p)} + \varsigma_\beta ,$$

with the accepting probability α given by

$$\alpha = min\{\frac{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\beta^{(n)},rest)^{q_{jt}}\phi(\beta^{(n)})}{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\beta^{(p)},rest)^{q_{jt}}\phi(\beta^{(p)})},1\}.$$

And ς_β is a draw from the $N(0,s_\beta)$, and s_β is the scaling constant, and set to achieve the acceptance rates close to the target value of 0.25 (same for all the other scaling parameters below).

2. Generating $\gamma_{M\times 1}$

The posterior of γ_m conditional on all other parameters is proportional to

$$\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\gamma_m,rest)^{N_{kj}}\}\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\gamma_m,rest)^{q_{jt}}\}\phi(\gamma_m)$$

I generate the candidate draw of $\gamma_{M\times 1}$ from the distribution

$$\gamma_m^{(n)} = \gamma_m^{(p)} + \varsigma_\gamma$$

And the accepting probability α is given by

$$\alpha = min\{\frac{\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\gamma_m^{(n)},rest)^{N_{kj}}\}\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\gamma_m^{(n)},rest)^{q_{jt}}\}\phi(\gamma_m^{(n)})}{\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\gamma_m^{(p)},rest)^{N_{kj}}\}\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\gamma_m^{(p)},rest)^{q_{jt}}\}\phi(\gamma_m^{(p)})},1\}.$$

3. Generating Θ

As discussed in section 3.5.1, the brand coordinates parameters can be separated into two parts: one with no constraints, the other with constraints on the positive orthant. Let $\theta_{jm(c)}$ denotes those constrained brand coordinates; $\theta_{jm(nc)}$ those unconstrained brand coordinates.

For $\theta_{jm(nc)}$, the posterior distribution conditional on all other parameters is proportional to

$$\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\theta_m, \ rest)^{N_{kj}}\}\cdot\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\theta_m, \ rest)^{q_{jt}}\}\phi(\theta_m|\tau_m^2)$$

where the prior distribution is normal $\phi(\theta_{jm(nc)}|\tau_m^2)\sim N(0,\tau_m^2)$. I find it is more efficient to generate the candidate draw of $\theta_{jm(nc)}^{(n)}$ from the distribution

$$\phi(\theta_{jm(nc)}^{(n)}|\tau_m^{(p)2})\sim N(0,\tau_m^{(p)2})$$

And the accepting probability α is given by

$$\alpha = min\{\frac{\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\theta_{jm(nc)}{}^{(n)},\ rest)^{N_{kj}}\}\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\theta_{jm(nc)}{}^{(n)},\ rest)^{q_{jt}}\}}{\{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\theta_{jm(nc)}{}^{(p)},\ rest)^{N_{kj}}\}\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\theta_{jm(nc)}{}^{(p)},\ rest)^{q_{jt}}\}},1\}.$$

This is the third family of Metropolis-Hasting algorithms, as described by Chibs and Greenberg (1995, p.330).

For $\theta_{jm(c)}$, similar to the treatment in Park, Desarbo & Liechty (2008), I adopt a random-walk Metroplis-Hasting algorithm with a weakly informative gamma prior $\phi(\theta_{jm(c)}) \sim G(0.01, 0.01)$ and a gamma proposal to generate the candidate draw. The proposal density takes the form

$$\theta_{jm(c)}{}^{(n)} \sim G(k\theta_{jm(c)}{}^{(p)2}, \frac{1}{k\theta_{jm(c)}{}^{(p)}})$$

Again the scaling parameter k is set to achieve an adequate acceptance rate. And the accepting probability α is given by

$$\alpha = min\{\frac{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\theta_{jm(c)}^{(n)}, rest)^{N_{kj}}\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\theta_{jm(c)}^{(n)}, rest)^{q_{jt}}}{\prod_{k=1}^{J}\prod_{j\neq k}^{J}\pi_{j,-k}(\theta_{jm(c)}^{(p)}, rest)^{N_{kj}}\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\theta_{jm(c)}^{(p)}, rest)^{q_{jt}}} \times \frac{\phi(\theta_{jm(c)}{}^{(n)})p(\theta_{jm(c)}{}^{(p)}|\theta_{jm(c)}{}^{(n)})}{\phi(\theta_{jm(c)}{}^{(p)})p(\theta_{jm(c)}{}^{(n)}|\theta_{jm(c)}{}^{(p)})},1\}.$$

Second, consider the heterogeneity parameter ($\overline{w}, \Omega$).

4. Generating $\overline{w}$

The posterior of $\overline{w}$ conditional on all other parameters is proportional to

$$\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\overline{w}, rest)^{q_{jt}}\}\phi(\overline{w})$$

I use Metropolis-Hasting algorithm to generate draws of $\overline{w}$ with a random walk chain. Let $\overline{w}^{(p)}$ denote the previous draw, and then the next draw $\overline{w}^{(n)}$ is given by

$$\overline{w}^{(n)} = \overline{w}^{(p)} + \varsigma_{\overline{w}},$$

with the accepting probability α given by

$$\alpha = min\{\frac{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\overline{w}^{(n)}, rest)^{q_{jt}}\phi(\overline{w}^{(n)})}{\{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\overline{w}^{(p)}, rest)^{q_{jt}}\}\phi(\overline{w}^{(p)})}, 1\}$$

where $\varsigma_{\overline{w}} \sim N(0, s_{\overline{w}})$, and $s_{\overline{w}}$ is the scaling constant.

5. Generating $\Omega = \begin{bmatrix} \sigma_1^2 & & & \\ 0 & \sigma_2^2 & & \\ 0 & 0 & .... & \\ 0 & 0 & 0 & \sigma_{\mathrm{M}}^2 \end{bmatrix}$

I use the similar strategy with $\theta_{jm(c)}$ here for σ_{M}^2. In particular, And a random walk gamma proposal is adopted to generate draws:

$$\sigma_m^{-2(n)} \sim G(k_2\sigma_m^{-2(p)\,2}, \frac{1}{k_2\sigma_m^{-2(p)}})$$

And the accepting probability α is given by

$$\alpha = min\{\frac{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\sigma_m^{2(n)}, rest)^{q_{jt}}}{\prod_{t=1}^{T}\prod_{j=1}^{J}\rho_{jt}(\sigma_m^{2(p)}, rest)^{q_{jt}}}\frac{\phi(\sigma_m^{-2(n)})p(\sigma_m^{-2(p)}|\sigma_m^{-2(n)})}{\phi(\sigma_m^{-2(p)})p(\sigma_m^{-2(n)}|\sigma_m^{-2(p)})}, 1\}$$

Finally, the hyperparameters, τ_m^2, are generated using Gibbs Sampler, where the candidate draws are derived from the following full conditional posterior distributions:

$$\tau_m^{2(n)} | \theta_m^{(p)} \sim IG(\frac{J-T}{2} + 0.01, \frac{1}{\frac{\sum_i (\theta_{im(nc)}^{(p)})^2}{2} + 100})$$

where T is the number of constrained parameters.

Appendix F Abbreviates for Brand Names in Figure 2-4 and 3-2

Pepsi	"Pepsi"
Diet Pepsi	"DPepsi"
Mountain Dew	"MDew"
Pepsi Free	"PepsiFr"
Diet Pepsi Free	"DPepsiFr"
Lemon-Lime Slice	"LLSlice"
Diet Lemon-Lime Slice	"DLLSlice"
Cherry Coke	"ChCoke"
Diet Cherry Coke	"DChCoke"
Coke	"Coke"
Diet Coke	"DCoke"
Caffeine-Free Coke	"CFCoke"
Caff-Free Diet Coke	"CFDCoke"
Sprite	"Sprite"
Diet Sprite	"DSprite"
Mr. PiBB	"MrPiBB"
Dr. Pepper	"DrPepp"
Diet Dr. Pepper	"DDrPepp"
7-Up	"7UP"
Diet 7-Up	"D7UP"
Other	"Other"

Appendix G Simulation Study for Chapter 3

In order to verify the validity of MH algorithm in Appendix E, I perform the following simulation experiment. Using the estimates in Table 3-2, where $\beta = 5.7, r = c(-0.89, 2.23), w \sim MVN(\begin{bmatrix} 1.07 \\ -0.04 \end{bmatrix}, \begin{bmatrix} 0.51 & 0 \\ 0 & 0.36 \end{bmatrix})$, and the θ values equaling the brands locations estimates in the table, I simulate the aggregated data by generating 1000 consumers' ideal-points for each of the 63 weeks, and simulate the forced switching data by assuming the existence of 1109 consumers with their first choices the same as the original dataset. The MH algorithm proposed in Chapter 3 is then used on these two simulated datasets, and the results reported below indicate that the true parameters can be recovered, with all true parameters within 95% credible interval.

Estimates of Utility Function Parameters from 2D model

		Posterior Mean (2.5%-97.5% credible interval)	
Price sensitivity	β	5.73 (5.65, 5.81)	
Quality coefficients	γ_1	-0.75 (-0.97, -0.52)	
	γ_2	2.32 (2.18, 2.46)	
Heterogeneity Parameters for Ideal Points	$\overline{w_1}$	0.95 (0.84, 1.06)	
	σ_1^2	0.47 (0.37, 0.57)	
	$\overline{w_2}$	-0.1 (-0.27, 0.07)	
	σ_2^2	0.25 (0.135, 0.365)	
Brands Locations		First Dimension	Second Dimension
	Pepsi	1.39 (1.26, 1.47)	1.16 (0.97, 1.35)
	Diet Pepsi	0.34 (0.22, 0.46)	0.38 (0.33, 0.47)
	Mountain Dew	-0.33 (-0.49, -0.18)	-0.63 (-0.70, -0.56)
	Pepsi Free	0.03 (-0.06, 0.164)	-0.03 (-0.09, 0.02)

Diet Pepsi Free	0.37 (0.27, 0.46)	0.23 (0.17, 0.29)
Lemon-Lime Slice	-0.29 (-0.47, -0.03)	-0.93 (-0.99, -0.83)
Diet Lemon-Lime Slice	-0.32 (-0.61, -0.1)	-0.77 (-0.88, -0.61)
Cherry Coke	-0.62 (-0.89, -0.40)	-0.73 (-0.87, -0.59)
Diet Cherry Coke	-0.19 (-0.53, 0.06)	-0.78 (-0.88, -0.61)
Coke	1.26 (1.07, 1.34)	1.45 (1.33, 1.57)
Diet Coke	0.41 (0.32, 0.48)	0.62 (0.56, 0.74)
Caffeine-Free Coke	0.33 (-0.02, 0.68)	-1.27 (-1.40, -1.18)
Caffeine-Free Diet Coke	0.57(0.42, 0.66)	0.46 (0.40, 0.56)
Sprite	-0.10 (-0.17, 0.009)	-0.30 (-0.35, -0.25)
Diet Sprite	-1.04 (-1.18, -0.80)	-0.09 (-0.23, 0.02)
Mr. PiBB	-2.29 (-2.4, -2.12)	0.02 (-0.1, 0.14)
Dr. Pepper	-0.51 (-0.62, -0.42)	0.19 (0.12, 0.28)
Diet Dr. Pepper	-0.70 (-0.90, -0.45)	-0.15 (-0.29, -0.02)
7-Up	1.49 (1.24, 1.64)	-0.38 (-0.46, -0.31)
Diet 7-Up	0.96 (0.74, 1.25)	-0.58 (-0.66, -0.50)
Other	-0.78 (-0.90, -0.67)	2.14 (1.93, 2.35)
Deviance	7188.8773	