1-1-2012

Decision Support Models For The External Variety Of Configurable Products

Erkan Isikli
Wayne State University,

DECISION SUPPORT MODELS FOR THE EXTERNAL VARIETY OF CONFIGURABLE PRODUCTS

by

ERKAN ISIKLI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for degree of

DOCTOR OF PHILOSOPHY

2012

MAJOR: INDUSTRIAL AND SYSTEMS ENGINEERING

Approved By:

Advisor Date

DEDICATION

In the memory of my grandmother,

Who has vanished physically, but has not broken off her spiritual bond with me.

I owe her a lot for teaching me courage, compassion, and love.

All that I am or ever hope to be, I owe to her.

I am grateful to God for every minute I spent with her in the first 10,199 days of my life.

The game is finally over, Gülot. Or, has it just started?

Well, you know better than me.

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr. Ratna Babu Chinnam for his guidance throughout the preparation of this dissertation. I greatly appreciate the time and effort he dedicated to helping me complete my PhD, as well as his advice over the past five years.

I sincerely appreciate all the feedback and experience Dr. Boris Baltes and Gint Puskorius shared with me. They helped me a great deal with their wisdom and constructive criticisms and provided me a lot of suggestions when reviewing my work. I would also like to thank Dr. Attila Yaprak, who is always kind and understanding, for his recommendations. His door was always open to me. It was also an honor to have Dr. Leslie Monplaisir and Dr. Alper E. Murat on my committee. I am so grateful to have such a distinguished and qualified committee.

I feel obliged to highlight the sacrifice and endless support of my family. I cannot find the words to express my gratitude to my parents. I am eternally thankful for their financial and emotional support, especially in the last year of my study. I would not be able to get my degree without their love and prayers.

Last, but not the least, I would like to thank Nezir Aydın for his great support and companionship during this tough journey. The life of a graduate student is not easy. I cannot thank Nezir enough for being an excellent friend since my arrival in the United States.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

BLP	:	Berry, Levinsohn, and Pakes
CA	:	Cluster Analysis
FA	:	Factor Analysis
HR	:	Hazard Rate
LC	:	Locational Choice
ML	:	Mixed Logit
MNL	:	Multinomial Logit
MNP	:	Multinomial Probit
NL	:	Nested Logit
PCA	:	Principal Component Analysis
SA	:	Survival Analysis
SKU	:	Stock keeping unit

CHAPTER 1: INTRODUCTION

1.1. Motivation

In any industry, deciding which product variants to offer is a difficult problem, especially for sales and marketing departments as this decision can impact every other function within the company. The product variety problem is a double-edged sword. On the one hand, economies of scale and manufacturing costs dictate offering a small number of configurations as the cost of manufacturing flexibility to produce a large variety of products is significant and product complexity can have negative effects on lead times and quality. On the other hand, customer dissatisfaction costs suggest offering a suitable number of options to avoid customers from “walking away” (e.g., customer will switch to a competitor’s product).

Even though the cost of lost consumers is a big problem in every industry, the manufacturers of durable goods pay attention to it more than others due to high profit margins in their markets. Conlon and Mortimer (2010) state that consumers consider product availability very important, and Batchelor (2001) reports that a significant amounf of U.S. car buyers switch to another brand when they are not offered the product variants they are looking for. Thus, the firms try hard to satisfy consumer needs by offering a large number of product variants.

External variety comes at the cost of greater internal complexity. More recently, manufacturers have introduced some coping strategies such as decreasing the number of alternatives to offer (e.g., through options “bundling” and limiting the number of options), or depending on supply chain strategies such as “delayed product differentiation (postponement)” (*see* Pil and Holweg, 2004), by outsourcing larger modules and components to direct suppliers (Kohlberger & Gerschberger, 2012) to reduce internal complexity. For instance, after Ford announced that orderable vehicle configurations for 2009 model vehicles would be cut, on average, by half, the 2009 Ford F-150 was offered theoretically – only – in nine million combinations, a 90% reduction compared to the previous year (Wilson, 2008). However, almost twenty years before this dramatic change, Ford Fiesta had been offered theoretically in 27 million variants in Europe (Batchelor, 2001).

From an operations point of view, the abundance of product variants affect the cost of manufacturing and supply chain complexity; from a retailing point of view, it might increase inventory costs; from a marketing point of view, it might lead to choice overload (also known as cognitive complexity). However, as the degree of variety decreases, customer dissatisdaction costs increase; therefore, the profit margins and the percentage of lost sales might increase. As seen in Figure 1, the manufacturers are interested in finding the optimum variety level that maximizes the benefit of variety and minimizes the cost of variety. To the best of our knowledge, there is not an analytical way that has been followed by the industry or proposed in the literature to determine the level of variety that maximizes the net benefit. On the other hand, we should note that an ideal level is specific to the products, the markets, and the industries.

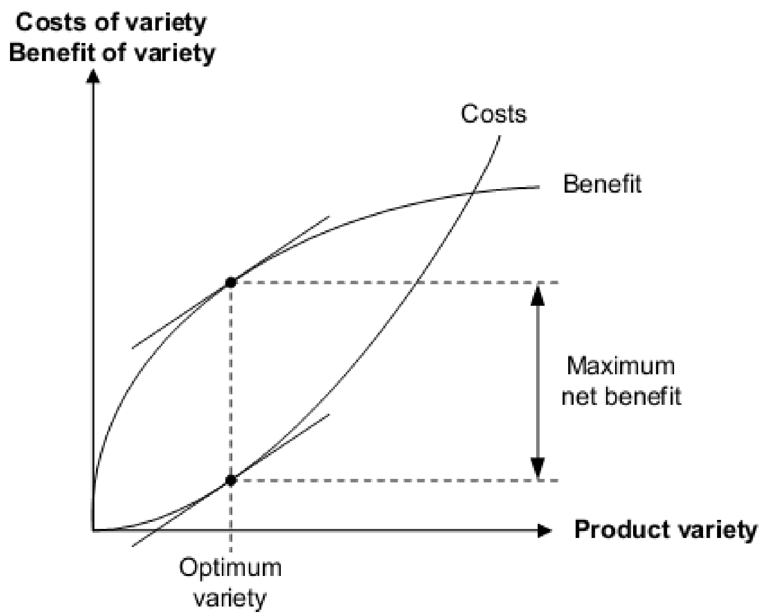

Figure 1: Conceptual description of costs and benefits associated with product variety (Rathnow, 1993; Blecker, 2003)

1.2. Background

1.2.1. Challenges in Managing Product Variety

Variety is often valued by consumers, at least to some degree, so it is possible to see many variants of a product in almost every industry. *Aussiebum*, which was founded as a men's swimwear manufacturer in 2001, today offers six different products (swimwear, loungewear, underwear, leisurewear, sportswear, surfwear) totaling more than 400 stock keeping units (SKUs), and the total number of underwear styles in their product line is increasing every day. It is even possible to see lots of varieties in small businesses such as local coffee shops (Caribou, Starbucks, etc.) and restaurants.

Many durable goods such as computers and automobiles are also customizable to an extent, which leads to heterogeneity in consumer preferences (i.e. consumers start choosing different variants because they can, not necessarily because they need them). Over the decades, manufacturers and retailers have tried very hard to satisfy consumers' preferences as much as possible since they thought more variety would lead to competitive advantage and higher market share, and in return they could dictate price easily (Wu, 2007). In 2006, for instance, Pontiac advertised that it could offer more combinations of options and accessories for its G6 series than the number of drivers in the U.S. (Ferguson & Donndelinger, 2010). There were, then, approximately 203 million drivers in the U.S., but only 300 thousand Pontiac sales were expected. Likewise, in 2008, it was claimed by Volvo that its new car C30 could be customized in five million ways (DriveChicago.com), which, in turn, sold only 39,966 units worldwide (Media Volvo, 2011).

The US auto industry is changing every day, trying to improve technology and product offerings based on customer preferences. An automaker can offer, theoretically, millions of different product configurations to gain a larger share in the market; however, offering more options is a burden for assemblers due to managerial problems and manufacturing complexity (Fisher *et al*., 1996; Wu, 2007). It is also problematic on the consumer side. It has been long debated that having the ability to choose among many alternatives increases the quality of life (Markus and Schwartz, 2010); however, beyond a certain point, consumers are likely to get overwhelmed with too much choice instead of feeling more freedom. This choice overload often leads to a decrease in satisfaction (Iyengar and Lepper, 2000).

The problem of product variety should be approached differently when durable products are in consideration. Consumers may feel confronted when choosing between different types of beverages at a coffee shop, but any dissatisfaction in this situation would not be devastating. When cars, laptops, and refrigerators are in consideration however, it could be more difficult for consumers to make a decision among too many alternatives. Thus, it is significant for OEMs to find a set of variants that are most profitable.

1.2.2. Definitions of Product Variety

Product variety is defined by Wu (2007) as “the number of different products or stock keeping units in a product line.” In the literature, various disciplines have introduced their own views on product variety. For example, in marketing, researchers are mostly interested in the distinction between “perceived” and “actual” variety (Kahn and Wansink, 2004), while in manufacturing, researchers classify product variety in three groups: fundamental variety, parts variety, and peripheral variety (Abdelkafi, 2008). Throughout this study, we will follow the terminology used by Anderson & Pine (1997), in which variety is broadly defined as “external” and “internal” variety. External variety, which is the focus of this dissertation, is the variety seen by the consumer, while internal variety is the variety experienced when manufacturing the product.

External variety can be defined as “a form of complexity arising from either a large number of product families or a large number of alternative configurations within single families” (Batchelor, 2001). It may increase manufacturing costs (Lancaster, 1990), inventory costs (Pil and Holweg, 2004), and the amount of time consumers spend to make a purchase. External variety has two types: useful (positive) and useless (negative). Offering differrent body style alternatives is an example of the former case as a typical auto buyer would be interested in that, while offering over abundance of interior trim choices (Elgard & Miller, 1998), different styles of steering wheels, or an excessive range of exterior paint are examples of the useless variety as they would not appeal to a typical auto buyer. Table 1 summarizes the discipline-specific definitions of product variety.

Table 1: Different definitions of product variety with examples from the automotive industry (Batchelor, 2001; Kahn and Wansink, 2004).

Type of Variety	Definition
Useful Variety	The level of variety demanded by customers e.g., a range of trims and functions on a car seat
Useless Variety	Variations in the product design not associated with meeting customer requirements (do not directly impact upon the functionality) e.g., a range of different fixtures and fitting procedures in each seat derivative
Fundamental Variety	The level of variety associated with the range of different platforms, models, and body styles
Peripheral Variety	The level of variety associated with the number of combinations of options provided (deals with all the different options available per model).
Parts Variety	The number and type of parts required to produce different models.
Perceived Variety	The form of variety that is a result of the evaluation/assessment of the actual variety by consumers
Actual Variety	The number of distinct items in the assortment.

1.3. Scope and Research Objectives

We are aiming to design and develop decision support models that can help managers make micromarketing strategic decisions regarding configurable products by answering questions about configuration selection, variety determination, and feature bundling. In a sense, we are preparing the appropriate environment for operational models to be estimated/computed quickly and constructed under realistic assumptions. We are not presenting an operational model in this study; however, our approach falls under marketing engineering, which is defined by Lillien & Rangaswamy (2007) as “the systematic translation of data and knowledge into tools used for decision support.” Our research is targeted at configurations in a single product category that can incrementally change over time; however, the proposed framework can also be extended to the case of multiple product categories.

In today’s automotive industry, manufacturers have been introducing more product variety due to the increasing variety seeking customers, changes in energy prices and environmental regulations (Staeblein *et al*., 2011). When shifting from mass production to mass customization, auto manufacturers should set the degree of variety at a level that would provide both a competitive edge and create acceptable costs. However, this level can be determined only theoretically (Abdelkafi, 2008); there is no analytical model suggested in the literature to solve this problem effectively.

The number of product configurations increases exponentially as the choice sets become extensive and the number of components gets large; a problem which leads to complex and expensive configuration design processes and operationally suffered sales configurators (Salvador & Forza, 2007). The addition of a new configuration may satisfy the needs of a lost customer, but it may not necessarily increase sales. As Salvador and Forza state, consumers may start selecting a configuration because they can, not because they specifically need it. If the number of alternatives is increased freely without considering this fact, the design process suffers and no significant incline in sales is experienced. As illustrated in Figure X, Salvador and Forza (2007) suggest taking four steps when reducing the cognitive complexity of the customers of configurable products. Our framework intends to support the last two steps in this scheme.

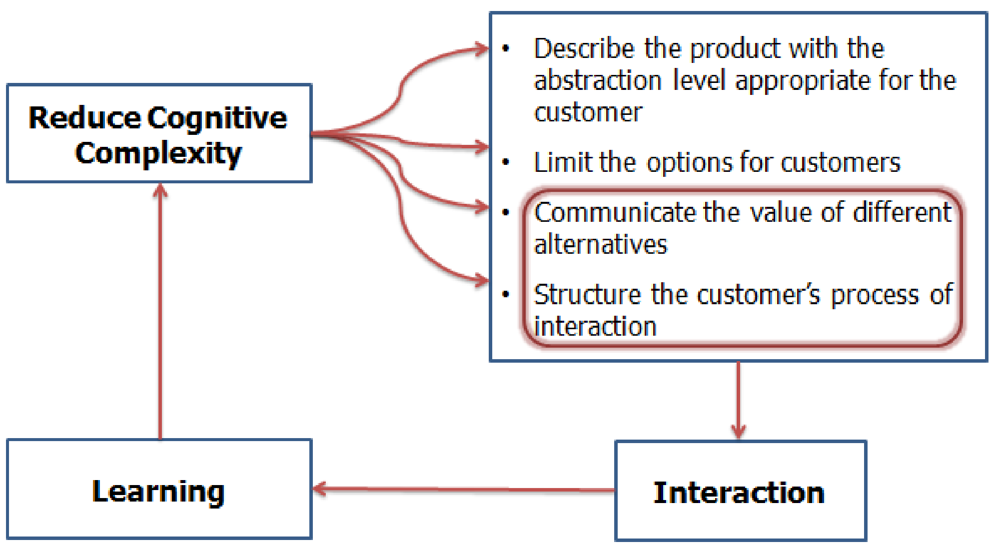

Figure 2: Key Principles for Sales Configurations Design (Salvador & Forza, 2007)

Pil and Holweg (2004) show that there is no real correlation between the level of variety offered and the units of cars sold in the European auto market in 2002. Peugeot, which was the market leader in sales with 596,531 units of its *206* model, offered only 1,739 configurations. In contrast, Mercedes offered more than three septillion ($3x10^{24}$) configurations of its E-Class model, but it sold only 157,584 units.

Thus, finding the best way of building configurations and setting the price accordingly can help the design process.

Perception of product variety differs across auto markets worldwide, even across regions within the same country. For instance, American and European automakers are not similar in offering and managing variety as relatively higher variety is almost a must in Europe. In Germany, Ford S-max, a minivan produced by the Ford Motor Company for the European market, was offered in 64 standalone options, whereas its closest competitor Volkswagen Touran was offered in 80 standalone options.

This study focuses on the U.S. auto market, but it can also be adapted for other markets that mostly operate in a build-to-stock environment. In addition, it is not an industry-specific study, and it is possible to generate similar models in every industry where configurable products exist and the market (segment) is well-defined. Decision makers in the auto, aircraft, and computer industries, where multi-dimensional product differentiation is common (Feenstra, and Levinsohn, 1989), can benefit from our approach. Since our proposed model can also be used to determine the effects of marketing actions such as cannibalization, it is convenient to use when making micro-marketing strategies; however, it should not be used as an operational model. On the other hand, even though we are not explicitly modeling the assortment problem, our models consider assortment structure implicitly and can address issues about the construction of assortments well.

Our research questions are summarized as follows:

- What is the most effective way of reducing too many decisions on product configurations (that can run into thousands if not hundreds of thousands or even millions) to a smaller number of specifications without compromising much of the important information captured before building operational models (e.g., inventory and demand models)?
- How do we effectively identify attributes that add value to consumers' perceptions about configurations?
- How do we identify the most suitable assortment structure in terms of variety that can increase sales and profitability?

The remainder of this dissertation is organized as follows: Chapter 2 reviews the product variety literature, Chapter 3 develops an experimental framework and discusses the proposed methodology, Chapter 4 discusses the experimental results, and Chapter 5 concludes with summary remarks and directions for future research.

CHAPTER 2: LITERATURE REVIEW

One of the most critical decisions manufacturers and retailers need to make in today's world is to determine the appropriate degree of variety. Researchers from various disciplines have tackled this problem, and each of them has employed the approach that would fit their needs best and developed the models that could serve their needs best. Even though there are studies that isolated the product variety problem and looked at it from a small angle, the views of manufacturers, consumers, and relatilers have been blended well in wholesome studies on product variety.

In this section, we provide a general review of the most relevant literature that helped us structure our study by focusing on three main streams of research: *consumer choice modeling*, *demand planning*, and *assortment planning*. These streams are not somewhat intertwined, however, since studies from different streams have been used as benchmarks for each other and have influenced one another.

2.1. Introduction

It is essential to manage variety in terms of revenue, manufacturing costs, more accurate demand forecasts, and long-term effects of consumer loyalty, in today's globalized economies. Different categorizations of product variety can be found in the literature. From a marketing point of view, there is actual variety and perceived variety (Kahn and Wansink, 2004), or attribute-based variety and temporal variation (Chintagunta, 1999), whereas from a manufacturing point of view, there are model-mix variety, options variety, and parts variety (MacDuffie *et al.*, 1996), or from a general point of view, there are internal and external variety (Pil and Holweg, 2004).

Market-oriented variety is concerned with satisfaction of consumer preferences so that market share can be increased, whereas manufacturing-oriented variety is mostly concerned with economies of scale (Corrocher and Guerzoni, 2009). In this work, our focus, external variety, can be seen as a form of actual variety.

Product variety has been studied by various researchers from distinct areas of research such as marketing science, operations management, assortment planning, category management, and inventory management. Table 2 provides a general summary of the related literature from different perspectives. Based on the primary foci of researchers, it is also possible to categorize the related literature as consumer-oriented research and cost-oriented research. The former is mostly interested in satisfying diverse consumers as much as possible (for an extensive review of this literature, *see* Ramdas, 2002), whereas the latter mostly focuses on reducing the manufacturer's burden (Fisher *et al.*, 1999). One can add psychosocial research to the above classification as a third category; however, in this review, we will consider it as a branch of consumer-oriented research. Interested readers should refer to Markus and Schwartz (2010), Ha *et al.* (2009), Schwartz (2004), and Iyengar and Lepper (2000), works that discuss the behavioral aspects of choice behavior and the concept of choice in detail.

Table 2: A General Look at the Literature

Focus	Description	Approach
Manufacturer's side	Developing strategies to manage external variety of Rover 75 (Batchelor, 2001); manufacturing complexity (Fisher et al., 1996; Wu, 2007); an analytical approach to manage the complexity and risk associated with external product variety in the auto industry (Kamrani & Adat, 2008)	Conceptual, Simulation
Marketer' side	As variety increases, market share increases (Wu, 2007), demand increases (Kahn, 1998), supply chain costs increase (Fisher and Ittner, 1999)	Conceptual, Statistics, Econometrics
Consumer's side	Having ability to choose among many alternatives increases the quality of life (Markus and Schwartz, 2010); beyond a certain point, it leads to choice overload (Iyengar and Lepper, 2000)	Behavioral, Statistics
Retailer's side	Assortment planning: Finding an optimal assortment (Kök *et al*., 2008)	Optimization, Data mining

Consumers value variety as they are willing to find the exact product in their mind when they are in the market; this variety seeking behavior makes companies satisfy consumers' needs by offering more customizable products. However, offering too much variety often leads to an increase in demand variability, complexity of manufacturing operations, forecast errors, inventory, and shortages, which can partially, or together, result in "market mismatch costs" (Fisher *et al.*, 1996; Fisher, 1997; Ramdas, 2002). Ramdas (2002), for example, shows that the extent to which a company is willing to customize depends on both its target market and the internal and supply chain capabilities of the company.

Ramdas (2003) identifies the variety related decisions within the company as shown in Figure 3. Pricing, packaging, and options offered are the internal decisions that mix with external parameters to determine sales volume.

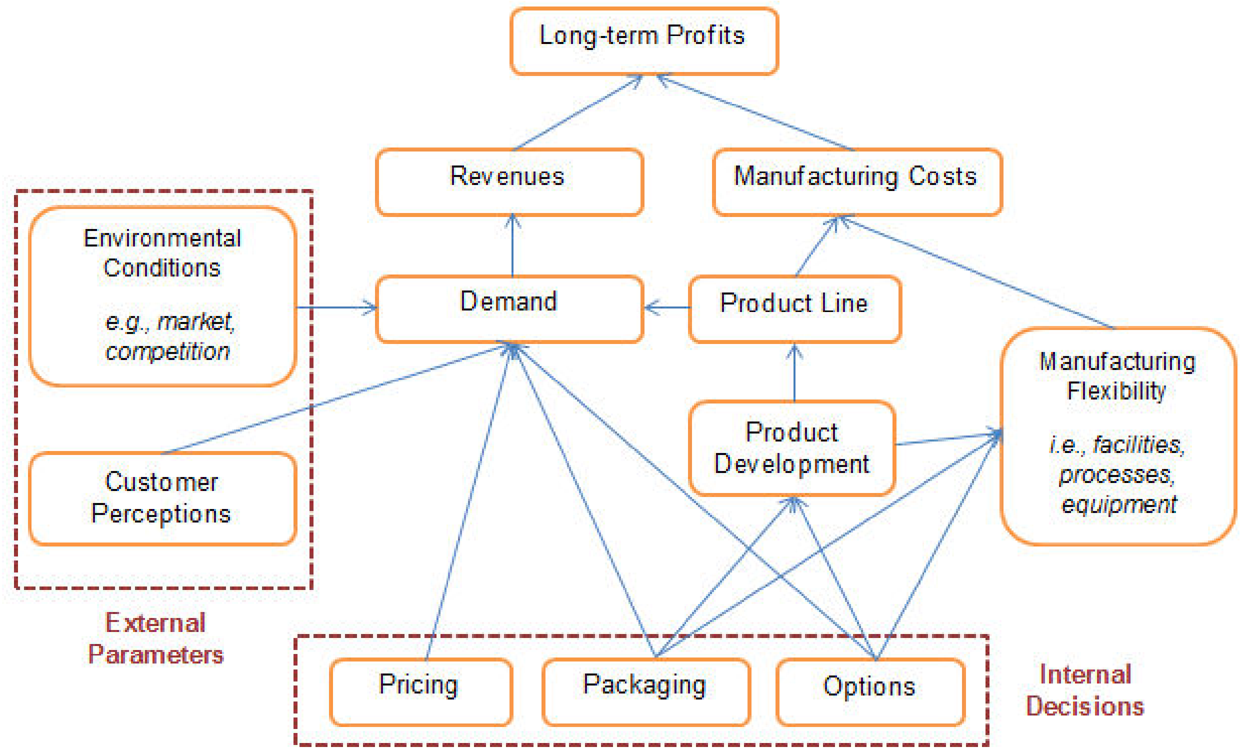

Figure 3: Variety related decisions (Ramdas, 2003)

2.2. Choice Modeling

For most of us, everything in life begins with a choice. Even though it seems simple at first, making a choice can be burdensome. Behavioral researchers of choice argue that when the number of alternatives is increased substantially, decision makers become overwhelmed as a result of cognitive complexity, and then they become dissatisfied in the end. Cognitive complexity (choice overload) is the difficulty in understanding the difference between the product variants offered (Scheibenne *et al.*, 2010).

In his remarkable book, *The Paradox of Choice (2004)*, psychologist Barry Schwartz also argues why people may not be better off when the number of alternatives is extensive. Consumers like the idea of finding the product that matches exactly what they have in mind. However, their decision making process may suffer when they are confronted by a great deal of choices. To an extent, having many alternatives to choose from is lucrative; however, as the number of choices grows further, consumers will become overwhelmed. Hence, it is important to converge to a break-even point in many instances where consumers think that the number of choices improves the quality of their life and that they are better-off, but are not overwhelmed.

As it is problematic for consumers to choose from a huge set of alternatives, it is often difficult for retailers and manufacturers to decide on how many configurations or which options to offer. Even though Berger *et al.* (2007) show that a high level of variety can provide companies a competitive advantage in the market, some researchers oppose to this argument (Ramdas, 2002).

Researchers have long debated the most of the drivers of choice. The most popular way of modeling it is to construct a utility function based on preference orderings. As discussed in detail in Anderson *et al.* (1992), a seminal reference on foundations of choice theory, the two well-known models of utility-based consumer choice are locational choice (see Lancaster, 1990) and discrete choice (*see* McFadden, 1986). In the former, consumers are assumed to perceive products as bundles of attributes, each product is defined by its location in attribute space, and consumers derive utility for each product from these attributes, where products are assumed to be equal in quality and price. The latter assumes that consumers try to maximize their surplus rationally, and estimate demand as a function of attributes, which drive consumer preferences.

A serious drawback of the locational choice (LC) models is that they consider only metric (continuous) attributes (Gaur and Honhon, 2006), which is a seldom found case as attribute-based product descriptions are commonly used in today's practice. However, they are more flexible in modeling substitution compared to discrete choice models. The most popular discrete choice model in the literature is the multinomial logit (MNL) model. However, it suffers from the well-known independence of irrelevant attributes (IIA) property, which, in simple words, assumes that introduction of an outside alternative does not change the odds of choosing two inside alternatives. This bans researchers from differentiating between initial choice and substitution (Kök and Fisher, 2006). To overcome this problem, McFadden (1986) introduced nested logit (NL) and mixed logit (ML) models to the literature, which make more realistic substitution assumptions and relax the IIA property [For more information on discrete choice models, see Anderson *et al.* (1992); for a recent review on extensions to MNL, see Train (2009)].

Both models (i.e., location choice and discrete choice) have been accepted in the literature based on their specific strengths. In a LC model, the total demand for an assortment covering the entire attribute space does not depend on the number of products in the assortment. Conversely, as variety increases, MNL demand increases. A detailed comparison of MNL and LC is provided by Kök and Fisher (2006). Another discrete choice model, multinomial probit (MNP), was used and modified by Hruschka (2007) to combine heterogeneity across households with a highly flexible deterministic utility term.

Another stream of research that is embedded in discrete choice models is present on choice set formation and screening rules, where consumers are assumed to first generate a reduced set of alternatives and then make a choice among these alternatives. Most of the studies in this stream have used scanner panel data functioning at the product level. Manrai and Andrews (1998) discuss alternative models and their extensions in assessing consumer choice processes. More recently, Levav *et al.* (2012) report the results of an extensive experimental study on the relationship between consumer search and the size of choice sets. Using different experimental settings, the authors explore how adaptive consumers are in making decisions when the choice sets are altered. They conclude that consumers search deeper when the choice set gets extensive. Gillbride and Allenby (2004) also provide deep insights on the estimation of consideration set models.

More recently, a new research stream has been started by Farias *et al*. (2011), which proposes a new data-driven approach when modeling consumer choice with limited information in which a non-parametric approach to both Amazon.com DVD sales data and a synthetic transaction dataset is applied.

2.3. Demand Estimation Under Product Unavailability

The majority of the demand planning literature has focused on estimating demand when consumers are offered the full choice sets. However, as Stefanescu (2009) noted, since unrealized demand is usually not known to the analysts, only observed sales can be used when estimating demand. Product availability information is a serious issue that should be addressed in cases on which we focus, thus we will cover only those studies that took product unavailibility and substitution behavior into account.

The previous literature on product substitution focuses extensively on fast moving consumer goods, such as food and beverages sold in super markets or in vending machines rather than durable goods such as automobiles, home appliances, etc. For instance, Bruno and Vilcassim (2008) extend the model proposed in Berry, Levinsohn, and Pakes (1995), which is often referred to as BLP, accounting for varying levels of product availability (without observing the set of available products in the store) at the aggregate level so that the demand estimates are not biased. They apply their framework in the British chocolate confectionery industry. Soft goods typically have such high service rates that substitution probabilities can be estimated easily by studying systems where only one product is missing from a full assortment. This is one of the advantages of studying consumer products. On the other hand, it is never easy to study a market where the full assortment of hard goods is offered in the store at any time. Albuquerque and Bronnenberg (2009) modeled demand for automobiles at the dealer level taking heterogeneity in consumer preferences into account and by assuming that the probability of purchase depends on dealer characteristics and geographic distance between consumer and dealer locations. Unfortunately, the authors are not able to model product unavailability explicitly in their remarkable study. However, as Stefanescu (2009) notes, if the stock-out effect is not considered when modeling demand, the forecasts for unavailable products are negatively biased.

2.4. Assortment Planning

As mentioned before, manufacturers and retailers have been searching for an answer to one of the most critical questions when making strategic decisions: What level of product variety should be offered? Inferring customer preferences and responding accordingly with updated product offerings plays a central role in a growing number of industries, especially for companies that manufacture configurable products. Of course it is not always possible for customers to find a variant of a configurable product due to retailers' capacity constraints. Thus, the retailers also want to know which products to include in the assortment in order to minimize the percentage of lost customers.

The assortment problem is defined for retail settings as "allocating space to items in a category" by Anupindi *et al.* (2009). It requires finding a set of products that should be discarded from the assortment in order to maximize revenue or profit. The nature of the problem is suitable to adopt an optimization framework, and as explicit advancement in software technologies has been achieved, the assortment planning literature has a rich stream called "assortment optimization" today. In this review, we primarily cover those studies in the literature that incorporate consumer choice models into an operational model in which the main objective is to build effective assortments (choice sets); however, we also cover some of the marketing-oriented research that aims to find the optimal choice set (assortment). Interested readers should refer to Kök and Fisher (2006) for a recent and more comprehensive review on this topic.

Yucel *et al*. (2008) consider assortment optimization in the case of consumer-driven substitution. The authors employ an exogenous demand model where level of substitution is limited to three. This assumption is in line with previous research where the ability to substitute has been allowed up to a certain level (Anupindi *et al*., 2009; Kusiak, 2007; Kök and Fisher, 2007; Gaur and Honhon, 2006). Their model is designed for the retail industry, recognizing supplier selection, shelf space constraints, and poor quality procurement. They stress on the consequences of neglecting substitution behavior of consumers by providing computational experiments, and suggest that retailers understand this behavior before making operational decisions. The authors also conclude that customer-driven substitution, supplier selection, and shelf space limitations should be considered when the aim is to generate efficient assortments.

Similar to Yücel *et al.* (2009), Kök and Fisher (2007) also assume an exogenous demand model in which either assortment-based (substitution that occurs when a product is not in the assortment) or stock-out-based substitution (substitution that occurs from a product being in the assortment but is temporarily out of stock) is valid. The authors emphasize the importance of capturing the willingness of consumers to substitute in an assortment planning framework. The formulation they propose for effective demand consists of three components: original demand, substitution demand, and unmet demand. They use a heuristic approach to determine the best assortment, and suggest that products with higher demand or higher margin should be included in the assortment first. This finding is in line with *efficient assortment strategy*, which has been recognized by many retailers and dictates retailers to abandon low-selling products (Cachon and Kök, 2007). However, the conclusion Kök and Fisher (2007) reach is somewhat different than the one in Gaur and Honhon (2006) as the latter claim that the optimal assortment does not have to include the best-selling product in every instance.

Even though the article by Kök and Fisher (2007) is a cornerstone in the assortment planning literature, the assumption of exogenous demand may be less restrictive in case of perishable goods compared to the case of durable goods. Gaur and Honhon (2006) is an innovative, yet limited study, which assumes that consumer preferences are not affected by the assortment structure. The authors assume that customer preferences are dependent on price exogenously, which can be restrictive in many real-life situations.

In a recent study, Anupindi *et al.* (2009) adopt a constrained integer programming model to find the optimal assortment a retailer carries considering disutility of consumers when their favorite product is not available. They find no significant change in consumers' variety perception when low selling items are eliminated, but favorite items of consumers are present.

In an attempt to estimate the demand for SKUs using historical sales and to find the optimal assortment, Fisher and Vaidyanathan (2009) adopt a locational choice approach. Fader and Hardie (1996) do not study the assortment problem, but also prefer SKUs in modeling consumer choice with MNL. They emphasized that it is challenging to define a set of attributes that can completely capture the variability among SKUs while using as few attributes as possible. Gaur and Honhon (2006) provide a generalization of locational choice models in case of stochastic demand, non-uniform consumer preferences, and inventory costs when building assortments.

Following a semi-parametric approach, Anupindi *et al*. (1998) develop a model for customer arrivals and purchase behavior in the context of retail vending that explicitly allows for product substitution and considers consumers' probability of walking away when there is a stock-out. The authors utilize information of stock-out occurrence and cumulative sales of all goods up to stock-out from inventory tracking systems and derive Maximum Likelihood Estimates of the demand parameters.

Nagarajan and Rajagopalan (2008) study optimal inventory policies for substitutable products. Two products with negatively correlated demands are considered for a numerical study where the level of substitution between items in a category and demand variation at the aggregate level is not high, but service levels are.

Chen and Plambeck (2008) analyze Bayesian optimal inventory level and myopic inventory levels to learn about substitution probabilities as well as customers' willingness to wait. This study is important in the context of systems with multiple, interacting products where customers generally substitute among multiple alternatives when their first choice is not available, and estimates on probability of substitution among multiple products is beneficial for inventory management.

Similar to the proposed methodology here, Kusiak (2007) proposes a *k*-means weighted clustering approach to determine key product configurations integrated with a sorting algorithm and an integer programming model based on sales data, and obtains a migration model in which consumer preferences are reflected in the configurations. The goal is to maximize customer coverage with the minimum number of prime configurations. The author tries to select prime configurations so as to make them as distant from each other as possible, the distance measure being calculated by the number of options to migrate and the difference in the amount of money to migrate.

Cachon and Kök (2007) explore multiple merchandise categories and basket shopping consumers in a duopolistic setting where retailers choose prices and variety levels in each category and consumers make store choices between retail stores and a no-purchase alternative based on their corresponding utilities. They demonstrate that category management never finds the optimal solution in terms of assortment and provides less variety and higher prices than optimal. The proposed model evaluates basket profits using point-of-sale data and is supported with a numerical study in which the loss in profits arising from category management is significant.

Kök with Fisher (2007) analyze demand estimation and assortment optimization under substitution. The authors propose a two stage algorithmic process to compute the best assortment for each store in which first the parameters of substitution behavior and demand are estimated, and then, an iterative optimization heuristic for solving the assortment problem is used. Their methodology is applied at a supermarket chain in the Netherlands; they claim that their model provides more than a 50% increase in profits.

Marketing-oriented research suggests that shrinkage of assortments does not necessarily decrease sales (Boatwright & Nunes, 2001; Iyengar and Lepper, 2000; Broniarczyk, Hoyer & McAlister, 1998). In a study examining how consumers perceive variety of an assortment, Hoch *et al.* (1999) use a measure of dissimilarity between product pairs and find that assortments with greater dissimilarity (perceived variety) satisfies consumers better. Similarly, van Herpen and Pieters (2002) queried the significance of two measures (entropy between products and disassociation of attributes) on the perception of variety. Sampaio *et al*. (2009) provide a comprehensive review on consumers' stock-out behavior, and identify fundamental variables that affect consumer responses using an experimental design. They try to answer if consumers substitute when their first preference is not available, or postpone their purchase, or simply walk away.

Brijs *et al.* (1999) present a case study where they suggest making assortment decisions based on association rules. In another data mining study, Wong and Fu (2005) improve the modeling approach in Brijs *et al.* (1999) by introducing the cross-selling effect; however, the authors cannot control for stock-outs and promotional effects. Chen and Lin (2007) employ a data mining approach to solve the product assortment problem using frequent item sets. They obtain association rules to find out which products are purchased by consumers at the same time using 6,568 transactions.

In summary, this literature review on product variety and consumer choice behavior shows that even though so much research has been done on the product variety problem, this area has still too much potential for new approaches. Especially data-driven approaches and real life applications are still necessary.

CHAPTER 3: RESEARCH DESIGN

Even though all manufacturing firms face competing objectives in determining which product configurations they will build, this problem is especially troublesome for manufacturers of configurable products, which sell a complex product with many options, resulting in a very large buildable configuration space.

The related literature is full of studies that analyze the product variety problem in different settings (Iyengar and Lepper, 2000; Batchelor, 2001; Kamrani & Adat, 2008; Kök *et al*., 2008). However, most of these models become impossible to apply in the real world, especially when the focus is on the product rather than on the brand, the market, or the industry. Table 3 summarizes some studies critical to our research.

Table 3: A Second Look at the Literature

Studies		**Objectives**
Some Conceptual Ideas	Markus and Schwartz (2010), Ha *et al.* (2009), Schwartz (2004), Iyengar and Lepper (2000)	Choice concept, behavioral aspects of consumer choice behavior, cognitive complexity
	Gillbride and Allenby (2004), Levav *et al*. (2012), Manrai and Andrews (1998)	Choice set formation, consumer search, screening rules
Some Data-Driven Approaches	Yunes *et al*. (2007)	Reduce the number of configurations without upsetting customers or sacrificing profits
	Kusiak *et al*. (2007)	Capture prime configurations
	Farias *et al*. (2011)	Model consumer choice with limited information

In this study, we offer a framework, benefiting from both quantitative and qualitative mehtods, that will help reduce the number of alternatives for a configurable product in a proper way when building statistical or operational models to develop and guide micromarketing strategies.

While the methods might be relevant to several industries and markets, our primary focus will be on the automotive original-equipment-manufacturers (OEMs). Given the significant differences between major automotive markets (e.g., generally speaking, North American dealers mostly sell from stock carried at the dealerships whereas in Europe and Japan there is often little inventory at the dealerships and the customers are more used to build-to-order strategies) and our relationship with U.S. Automotive OEMs, we will be focusing on the U.S. market, and in particular, the cross-over utility vehicle (CUV) market segment. The market share of CUVs has increased substantially since the early 2000s, and no significant decline in sales is expected in the near future. CUVs are especially popular nowadays (in comparison to sports utility vehicles (SUVs)) (Sema, 2009) as gas prices passed the psychological limit of $4 per gallon of gas. Due to relatively healthy profit margins, such decision support models in this market segment are crucial for any OEM targeting the segment.

3.1. Basic Terminology and Assumptions

We define a *configuration* as a combination of options (Mittal and Frayman, 1989), a *product variant* as a combination of features at the product or subsystem level, and *similarity sets* as clusters of configurations with similar attributes and demand patterns. Note that throughout this study, an *option* is referred to a product attribute (characteristic) that has only two levels (e.g., absence or presence of moon roof in a vehicle). Finally, we define an *assortment* as the configuration breadth in the store.

We make a couple of assumptions in order to have flexibility when designing our framework. Some of these assumptions are necessary for the demand estimation module; the others are made for general purposes. First of all, adopting Lancaster's product definition, we assume that a configurable product is a combination of attributes (options), and the utility of each customer is derived by the observed characteristics of the product (Lancaster, 1971). Each consumer ranks the configurations based on his/her taste and preferences, and each configuration is attracted by consumers at some level. We also assume that each consumer chooses the configuration from the available assortment that maximizes his/her utility, and purchases only one unit (We are proposing a decision support tool for configurable products, so it is unlikely to observe a customer purchasing more than one unit of such goods at a time).

Durable goods, especially the high-priced ones such as automobiles, are almost always sold through customer substitutions (also called "diversions" in practice). The number of possible configurations a manufacturer can build increases exponentially as a function of number of options and features. For instance, when there are 16 binary options offered by an automobile manufacturer, a customer can customize that product, in theory, in 65,536 (i.e., 2^{16}) distinct ways. However, due to limited channel inventory capacity (at individual dealers and within local markets), only a small proportion of the possible configurations are available at any time. Thus, we assume that consumers have flexible preferences. However, heterogeneity of consumers is confined since the focus of research is a specific product rather than a product line or a market segment. One can easily extend this framework for the case of multiple product categories by simply employing a utility approach that takes heterogeneity of consumer preferences into account explicitly.

Our framework breaks a regional (national) market down into more manageable parts, so we assume that each of these sub-regions have distinct demand characteristics and substitution between sub-regions is negligible. This is also common practice by many OEMs, and hence, is a very reasonable assumption. A practitioner, who wants to build aggregate models at the national level, should definitely consider this assumption – and if necessary, revise it – and select the demand estimation method accordingly. We should note that substitution is assumed to be consumer-driven, meaning it is either assortment-based or stock-out based; we cannot distinguish between the two.

We also assume that the product “variants” – or as we will call them “similarity sets” – are horizontally differentiated. Our framework focuses on a single product, so this assumption is not strong since consumers cannot easily distinguish product configurations in terms of quality. However, when multiple product categories (or even more than one brand) are considered, this assumption is not valid due to increasing external variety caused by consumers’ subjective preferences (i.e., perceived variety).

We also assume that a product’s design evolves gradually over time, but its features (e.g., options) maintain their relevance (context) over extended product generations/iterations (e.g., while it is typical to see new vehicle designs being introduced in the U.S. once every 4 to 6 years for most segments, many of the features such as the availability of a “moon roof” or “leather seats” holds true for decades if not longer). Hence, the proposed methodology, which aims to help decision makers understand and "mine" consumer preferences for product configurations and features from historical datasets (e.g., product availability/sales patterns for current/similar models in different markets), is expected to provide meaningful and actionable insights in planning future products and their configuration assortments.

3.2. Methods and Procedures

Different from other domains, marketing research is mostly focused on determining the variables that drive choice. Thus, when the number of attributes defining an alternative is relatively large, incorporating product attributes into demand models is a challenge most practitioners face. One possible way is to follow a Lancasterian approach (Anderson *et al.*, 1992) and keep the attribute space finite so that the number of alternatives is somewhat limited. Then, demand is characterized in terms of a finite number of "dimensions" as utility for a product is a function of attributes. Another possible approach is to assume that the number of characteristics is larger than the alternatives offered. We will stick to the former; however, more about the latter approach can be found in Sandeep *et al.* (2008). Sandeep *et al*. (2008) reveal the possibility that choice may not be explained within standard utility theory and suggest the use of alternative techniques such as hierarchical elimination by aspects (HEBA) in some cases.

Since we are analyzing external variety, we define configurations based on attributes. Nevertheless, in case of 30 binary attributes, one needs to search for more than one billion scenarios, and even if this explosion had been overcome, it would have been impossible to realize at least one transaction for each configuration considering the relatively small number of potential consumers. On the other hand, any statistical model using product attributes as indicator variables could suffer from multicollinearity (i.e., correlations between the indicator variables).

Since the abundance of different configurations to consider has a negative effect on the efficiency of an operational model; interaction between product features restrict the use of statistical models; and binary product attributes restrict the use of statistical models, we need to preprocess data in order to avoid potential obstacles. For these reasons, in the first stage, we create new variables denoting the *relative mix rates* and use principal component analysis (PCA) as a dimension reduction technique due to its flexibility in the presence of high multicollinearity, and re-describe configurations based on dimensions retrieved from this analysis by dichotomization, clustering, or subjective grouping. Then, we benefit from text mining techniques to validate the quantitative results.

We also propose the use of consumer reviews on auto blogs and other websites to check the validity of dimension reduction obtained by the PCA. In the second stage, we propose consumer preference / demand modeling using the most appropriate models over the space of these reduced PCA dimensions.

In the framework we propose, we first run PCA using the relative mix rates of each vehicle option to describe the dimensions on which configurations should be defined. These dimensions are validated by employing text-mining techniques on a secondary dataset, which includes online consumer reviews. After completing the validation step, the dimensions are named as properly as possible, and they are dichotomized. Reliance on core dimensions reduces the resolution of treatment of configurations to so-called “similarity sets” (marginal attributes/features will be out of consideration). For example, vehicle options and accessories such as floor mats and spare tire covers may no longer be in consideration. This preprocessing step is needed since the predictive power of almost every logit-like demand model suffers from the abundance of alternatives in the choice sets. Another reason is to avoid having too many instances where perfect substitutes are present.

We know that in perishable goods markets consumers often substitute (Anupindi *et al.*, 2009), but we do not have enough evidence to make a similar assumption in durable goods markets. In a case like ours, at the product configuration level, almost every alternative would be substitutable with one another if we employed an attribute-based approach. This leads to infinitesimal probabilities, which are pseudo since consumers may not perceive different configurations as different product variants. Following this procedure, we control the substitution that is not driven by consumers to an extent, and we deal with only consumer-driven substitution in the demand model.

Figure 4 illustrates our overall solution framework with pre- and post-processing steps. In order to achieve long term profitability, firms can handle product variety with an integrated model that consists of five modules:

a. Options development: Mostly handled by marketing and production departments,
b. Configuration selection: Determining the buildable configurations,
c. Variety determination: Determining the extent of different configurations that should be offered to the customers,
d. Bundling: Determining whether creating coexisting options enhance the demand of the product,
e. Pricing: Determining the most effective price for the product.

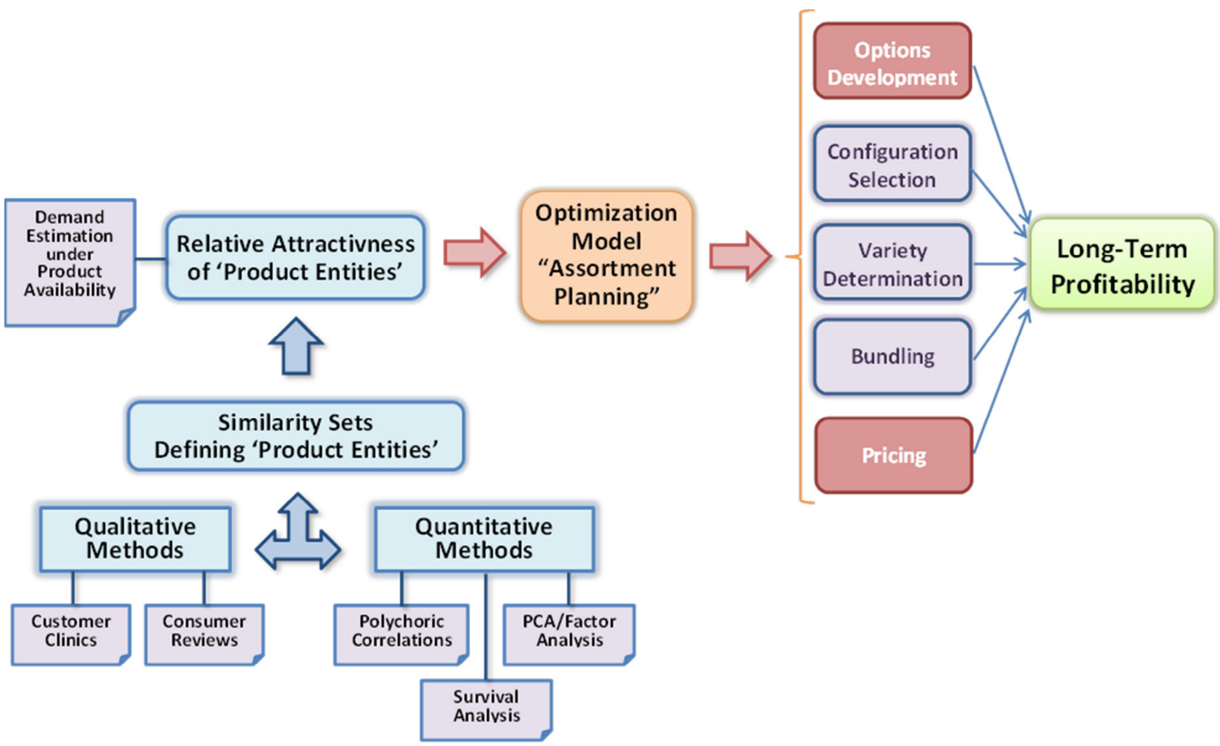

Figure 4: Overall Framework for Managing External Variety of Configurable Products

The reduction of external variety is a strategic decision that should not be made unless production, marketing, and other departments within the company cannot have a consensus on it. Our main objective is to group product configurations within a framework that benefits from both qualitative and quantitative methods in order to prevent further operational models of inventory management or demand estimation suffer from the abundance of product variants offered. Note that one needs to understand the dynamics of the market or the market segment in order to input the optimization model with the right information.

In this study, we are not exploring through which processes consumers went when estimating the relative attractiveness of similarity sets. Thus, we are not able to employ approaches such as consumer information processing (Thompson and Hamilton, 2006); we realize the sale of a product variant after the consumers have seen the ads, heard about the promotion, consulted with their family and friends, and checked their finances.

3.2.1. Preprocessing

Nowadays it is more common to see attribute-based product descriptions since it has become standard in product configurators (Salvador and Forza, 2007). However, consumers use well known attributes rather than the new and novel attributes when screening alternatives (Gillbride & Allenby, 2004), and they select groups of features together based on their needs and price points. This creates dependencies between features that should be leveraged to improve the product offering and forecasting. In order to avoid multicollinearity in such situations, one could employ Principal Component Regression (PCR) or Partial Least Squares

Regression (PLSR). If the main objective is data dimensionality reduction, as in our case, then Principal Component Analysis (PCA) or Factor Analysis (FA) can be employed.

PCA and FA ask the same basic question: Is it possible to duplicate the correlation or covariance matrices by using fewer inputs than the number of original variables? They lead to similar results if the dataset is good enough and if the research is reasonable (if the factor structure is strong). However, there are also some fundamental differences between these two techniques. PCA exploits the total variance (the correlation matrix), whereas FA uses only the shared variance (the covariance matrix). The former is convenient to use as it does not make any assumptions on the distribution of the data, whereas multivariate normality is sought when using the latter.

The basic assumption we make in this step is that what dealers ordering is somewhat correlated to what customers want. The estimation is done globally, which helps vehicle similarity set to have consistent meaning over time, but has the inability to account for trends and inter-temporal variation. This weakness of global estimation can be overcome with demand estimation model, which captures the trends and changes in customer preferences.

Traditional PCA cannot deal with binary variables as Spearman's correlation is not appropriate to use with discrete variables. Even though this obstacle can be overcome by using tetrachoric or polychoric correlations when reproducing the correlation matrix, estimation of these correlations requires maximization of a likelihood function, thus convergence many not be achieved in some cases. Besides, it is often time consuming as it involves tedious calculations. Instead, we adopted a patented approach by Ford Motor Company (Puskorius *et al.*, 2012) to transform the original discrete variables into continuous variables so that PCA can be run without hesitation. The formulation for the transformation is given in Equations 5 and 6:

$$x_{ik} = \delta_{ik}^{x} - \frac{1}{|S|}\sum_{j \in S} \delta_{jk}^{x}$$

where δ_{ik}^{x} a binary variable that denotes the absence or presence of the option *x* on vehicle *i* in period *k* and *S* denotes the assortment in period *k*. We also created the conjugates of each x_{ik} for practical reasons:

$$\widetilde{x}_{ik} = 1 - x_{ik}.$$

Thus, the nth principal component associated with the options $x_{ik}, y_{ik},$ and z_{ik} is denoted by

$$PC_{ik}^{n} = a_1 x_{ik} + a_2 y_{ik} + a_m z_{ik}$$

An illustration of the encoding is given in Figure 3. Inventory mix rates are calculated for each option (column means) and then column means are subtracted from the value of the corresponding option so that conventional factor analysis methods can be used.

	SE	SEL	LIMITED	SPORT	Row Mean
	1	0	0	0	0.25
	1	0	0	0	0.25
	0	1	0	0	0.25
	0	1	0	0	0.25
	0	1	0	0	0.25
	0	0	1	0	0.25
	0	0	1	0	0.25
	0	0	0	1	0.25
Column Mean	0.25	0.375	0.25	0.125	

	SE	SEL	LIMITED	SPORT	Row Mean
	0.75	-0.375	-0.25	-0.125	0
	0.75	-0.375	-0.25	-0.125	0
	-0.25	0.625	-0.25	-0.125	0
	-0.25	0.625	-0.25	-0.125	0
	-0.25	0.625	-0.25	-0.125	0
	-0.25	-0.375	0.75	-0.125	0
	-0.25	-0.375	0.75	-0.125	0
	-0.25	-0.375	-0.25	0.875	0
Column Mean	0	0	0	0	

Figure 3: The Encoding Patented by Ford Motor Company

3.2.2. Re-configuration

In this step, we transform the configuration space to a reduced product variant space using quantitative and/or qualitative data. This process starts with the employment of three approaches when describing the product variants: dichotomization, clustering, and subjective judgments. In the former approach, after determining the number of principal components that should be used, the component scores are dichotomized based on the corresponding means; under clustering, the component scores can be clustered using unweighted pair-group average as the linkage rule. In both of these approaches, we are trying to find how much any two configurations resemble one another.

A reasonable next step would be to validate the quantitative results with text mining techniques and name the dimensions obtained based on the classification of attributes (threshold, central, variety-enhancing) as proposed by Sanchez (1999). Our primary interest here would be variety-enhancing attributes, which account for product variety perceived by consumers.

3.2.3. Demand Potential

Our proposed approach first categorizes the principal components and create two-level dimensions to reduce the number of product variants to consider, then the dimensions are referred to as product “core features” and their combinations form “similarity sets” (consistent with the language from some OEMs). When estimating demand, these entities are taken into consideration instead of configurations.

In some cases, due to unavailability of some configurations, sales of available configurations might increase. This type of substitution is considerably high in auto markets. Thus, vehicle configurations would be perceived more substitutable than real if one used a demand model that does not consider product unavailability. Since vehicles are high priced products, we assume that customers would be willing to substitute for only a limited number of options when they cannot find the specific vehicle configuration for sale. This assumption is in line with the related literature. For instance, Smith and Agrawal (2000) assume that customers substitute only once; Kök (2003) show that effective demand in case customers are assumed to substitute three times can be approximated by changing the parameters of their single-attempt-substitution model; and Yücel *et al.* (2009) suggest that assuming the number of substitutions customers would make to be three is reasonable since substitution probabilities become smaller as the number of times customers can substitute increases. Note also that we are doing a region-based analysis, thus the substitutions are highly likely to occur within the region.

Since we needed the most proper demand estimation model in order to evaluate the relative attractiveness of each similarity set, we covered a great deal of studies in the related literature. The model that could help us should have worked when only product characteristics were known and have taken the product unavailability and substitution into account. The methods such as well-known BLP (Berry *et al.*, 1995) and its likes, which are based on consumer-level discrete choice models, cannot be used when practitioners have limited data sources. Therefore, we follow the demand estimation procedure provided in Vulcano *et al*. (2012), which employs the expected maximization (EM) method with the assumption of a multinomial logit (MNL) model of consumer preferences and a nonhomogeneous

Poisson model of consumer arrivals over multiple time periods. Their methodology is line with that proposed in Anupindi *et al*. (1998) as they both relate the rate of consumer arrivals to choice probabilities; however, the latter cannot model these probabilities within a utility framework.

Following the notation in Vulcano *et al*. (2012), we define the probability of a vehicle from similarity set *j* being purchased as

$$\pi_j(X,v) = \frac{v_j}{\sum_{k \in X} v_k + 1} \ ,$$

and the probability of customer walking away as

$$\pi_0(X,v) = \frac{1}{\sum_{k \in X} v_k + 1}$$

where ***X*** denotes the choice set at point of purchase and ***v*** denotes the preference vector. These probabilities are calculated using the manufacturer's market share, *s*,

$$s = 1 - \frac{1}{\sum_{k=1}^{m} v_k + 1}$$

and the transaction data in which the choice sets at each period are provided.

Vulcano *et al*. (2012) breaks observed sales into two parts: primary demand, and substitute demand. We are mostly interested in the substitute demand as we use it to decide whether the definitions of similarity sets are appropriate or not. Customers either substitute out of vehicle similarity set *q* when no vehicle from *q* is available or substitute in vehicle similarity set *q* when their primary choice is not available. In the former case, the substitute demand for vehicle similarity set *q* in period *t*, $\hat{W}_{qt}$, is defined by Equation (1), where $q \notin X_t \cup \{0\}$ and $\boldsymbol{z_t}$ is the vector of observed sales in period *t*, in the latter case, $\hat{W}_{qt}$ is defined by Equation (2), where $q \in X_t$ and z_{qt} is the observed sales of vehicle similarity set *q* in period *t*.

$$\hat{W}_{qt} = -\frac{v_q}{\sum_{k=1}^{m} v_k + 1} \cdot \frac{\sum_{l \in X_t} v_l + 1}{\sum_{l \in X_t} v_l} \cdot \sum_{l \in X_t} z_{lt} \qquad (1)$$

$$\hat{W}_{qt} = \frac{\sum_{l \notin \{X_t \cup \{0\}\}} v_l}{\sum_{k=1}^{m} v_k + 1} \cdot z_{qt} \qquad (2)$$

This economic consumer choice model is easy to use and understand, and it helps decision makers in grouping different configurations without losing too much information about external variety. Other than market share parameter, the model only needs information on product availability and transaction time (date). We claim that if the substitutions between different similarity sets are relatively low, then the new product definitions can be used without hesitation.

3.3. Assortment Structure

In the marketing literature, there have been a number of studies devoted to measuring the variety of an assortment. Hoch *et al*. (1999) is one of the pioneering studies that focused on building efficient assortments based on a variety index. Entropy, which was first proposed by Swait and Adamowicz (2001) for the complexity of choice sets, measures the degree of the dissimilarity between the products in an assortment. The authors state that when the assortment is full of equally attractive alternatives, the entropy is at its maximum. Thus, choice gets more complex for consumers when they are offered choice sets with high entropy. Note that the first use of entropy as a measure dates back to late 1940s. Shannon (1948) used this measure to quantify “the amount of information of a set of objects” (Fasolo *et al*., 2009).

After calculating the choice probabilities of each observation (configuration, product variant, or similarity set), we follow the footsteps of Swait and Adamowicz (2001) and evaluate the quality of an assortment in each period using the entropy measure

$$H(X) = H(\pi_X) = -\sum_{j=1}^{J} \pi(x_j) \log(\pi(x_j))$$

which is based solely on the choice probabilities, $\pi(x_j)$.

The variety of an assortment can be assessed following a product-based approach or an attribute-based approach. In this study, we are employing the former approach since our demand model is estimated using similarity sets, product variants, or configurations. Thus, we are not able to estimate purchase probabilities for attributes. However, one can easily modify the framework to have insights on the attribute space.

Traditionally, the choice probabilities, $\pi(x_j)$, are estimated using a logit-like discrete choice model such as MNL, NL, and ML. However, in our case, the choice probabilities are estimated using and EM-based demand model proposed by Vulcano *et al*. (2012). The difference is that in most of these discrete choice models, the heterogeneity of consumer preferences is taken into account explicitly; however, the EM-based demand model can only account for that implicitly. Nevertheless, overall conclusions would not suffer from this difference.

We should note that this entropy measure has been employed by many researchers (Fasolo *et al.*, 2009; Vermeulen *et al*., 2011), and it is accepted as one of the best ways to measure choice complexity (Duquette, 2010). Recently, Kessels *et al*. (2006) adopted this measure for a Bayesian design of choice. Rather than using it as a measure of complexity, in this study, we are linking this measure to the assortment effectiveness (*e.g.*, the number of products sold in a period). The main reason why we are hesitating to make conclusions about consumers' choice complexity is the implicit utility framework employed in our demand estimation model. We are not explicitly modeling utility taking consumer heterogeneity into account as aforementioned.

The entropy measure helps us to gain insights about the relationship between the variety (quality) of an assortment and the number (percentage) of products sold at a given time. Using this measure, the decision maker can determine whether the size or variety of an assortment affects the sales in a time period. It also sheds some light on consumers' buying and variety seeking behavior. Note that when the products in the assortment are equally attracted to the consumers, the entropy is at its maximum. However, when a product with a very high purchase probability (a dominating product) is included in the assortment, then the entropy measure gets so close to its minimum.

3.4. Survival Analysis

Survival (duration) models estimate the time-to-event using a set of explanatory variables, which can be either fixed or time-varying. These models are always preferred to an OLS model as they do not impose the basic assumptions of the OLS such as multivariate normality, linearity, and heteroscedasticity among explanatory variables (Tabachnick and Fidell, 2007).

There are mainly three approaches to survival analysis: parametric, semi-parametric, and non-parametric. In our study, we employ Cox proportional hazards model (Equation 5), which is semi-parametric and relates hazard rates to a log-linear function of the explanatory variables of interest.

$$\boldsymbol{h}(t) = \boldsymbol{h_0}(t)\boldsymbol{e}^{(\boldsymbol{x}'\boldsymbol{\beta})} \tag{5}$$

Cox PH model has non-parametric baseline hazard; this is why it is often preferred to parametric models, in which the distribution of the hazard function needs to be determined a priori. Cox PH model also has an efficient partial likelihood estimation procedure that makes it more convenient to use.

The explanatory variables we use in our survival analysis are the mix rate factor scores obtained using PCA, and in Eq.5, t_i denotes the time a configuration spends on lot in week k. Note that t_i is always positive and smaller than or equal to 8.

Cox PH model allows for discrete explanatory variables; however, using product attributes (options) as covariates to estimate “survival time” might lead to biased results in cases like ours. An unemployed person’s probability of re-employment does not depend on the set of all the unemployed people in consideration. In other words, a person does not stay unemployed since another person becomes employed or a new person becomes unemployed. However, in our case, the absence or presence of a configuration at a time point might affect the time to purchase of the other configurations available at that time. This is why we are using the mix rate coding in order to estimate days a configuration spends in store.

3.5. Concluding Remarks

Our methodology, illustrated in Figure 5, can be employed both as a data-driven and a knowledge-driven decision support tool as it can answer “what-if” questions using quantitative models and captures the qualitative knowledge with the help of consumer reviews.

In case of durable goods, since the consumers often are not aware of what products exist in the marketplace, the substitution cannot be classified as stock-out or assortment-based (Kök *et al*., 2008). However, substitution between different configurations (product varieties) is still valid (present). Even though there is evidence in the literature that customers sometimes substitute up to a limited number of bits when their first choice is not available (Yücel *et al.*, 2009), we are not making any assumptions about the number of substitutions that can occur between different varieties/configurations. Our demand model assumes that if customers cannot find their first choice (ideal product), they either substitute or walk away. However, when customers substitute, we are not interested in the number of substitutions until the purchase is made.

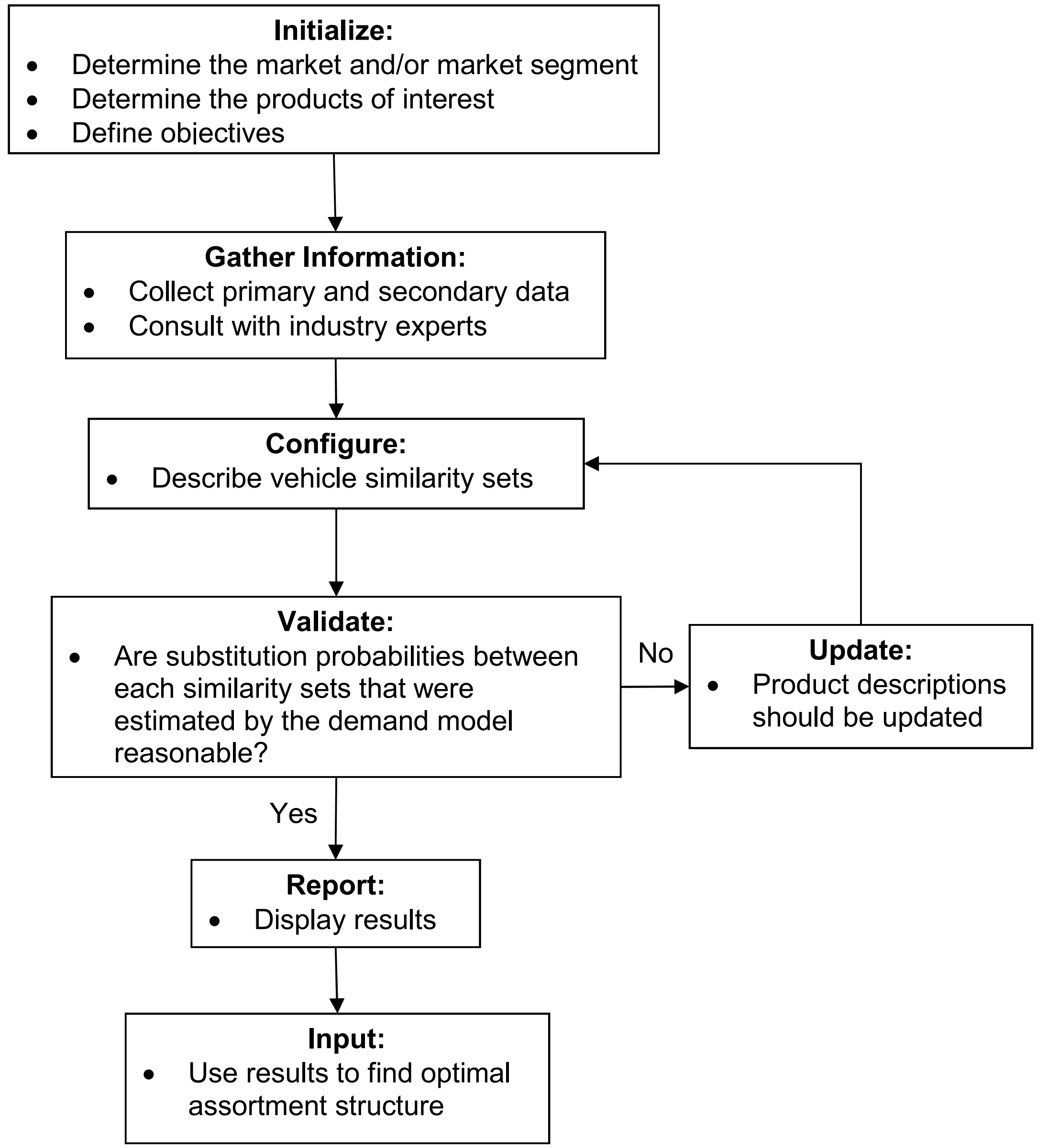

Figure 5: Flow Chart of the Proposed Framework

CHAPTER IV: RESEARCH SUMMARY AND EXPERIMENTAL RESULTS

4.1. Introduction and Data Description

We apply our framework in the midsize cross utility vehicle (CUV) segment of the U.S. automobile industry. The market is worth billions of dollars, and it has been growing since the mid-2000s. The leading brands are Toyota, Honda, Nissan, Ford, and General Motors. The nationwide economic crisis around 2008 increased the interest in more fuel-efficient vehicles and the incline of oil prices following the recession led consumers from SUVs to CUVs. Having gained 83% market share in 2011, CUVs are expected to continue dominating the U.S. SUV market (Mintel, 2012).

The primary data provided by our collaborating OEM covers daily transactions from August 2010 to March 2012 on the product of focus, which is one of the leading models in the segments. The firm divides the national market for the product into 17 regional markets. We observe unit sales, price, options, and availability of each product configuration. The dataset was from a perpetual inventory system, i.e. every time a purchase was made, the dealer knew how many vehicles were available and which configurations were out of stock. It provides information about the alternatives that were available at the point of purchase, but were not purchased. However, it does not include any information on consumer characteristics. We know the list prices for the configurations, but not the transaction prices. Crafton and Hoffer (1980) state that the relationship between the list prices and the transaction prices is market determined, but they show that the former can be used to explain the latter as consumers start negotiating with the dealer based on list prices.

As a secondary data source, we use consumer reviews retrieved from main auto blogs and consumer web sites, which help practitioners to learn more about consumers' views about the product in consideration. We apply text mining techniques to this information to find out post-purchase behavior and perceptions of the consumers. This is important since comments and reviews on such web sites may have an effect on consumers. We are benefiting from text mining techniques to retrieve and extract information on consumer behavior. Note that this is an exploratory analysis

4.2. Data Preprocessing

We employ our methodology in one of the largest markets in the U.S.: New York. We apply our framework on the first 32 weeks of the data, and run the analysis for each 8-week period separately. One reason to do this is the carryover effect. Using a regression model, we find that the variation in supply at time t can be explained significantly by sales at time t-4, but it cannot be explained by sales in t-1, t-2, or t-3. Equations 1-4 below show the results of the four separate regression models estimated to explain the variation in supply with sales by changing the length of time lag. S_t denotes the supply, whereas Q_t denotes the sales volume at time t; p value shows the probability of accepting that the beta coefficient is equal to zero when it is not. The beta coefficient was found insignificant for time lag smaller than 4.

$$S_t = 311.95 - 0.4545 \cdot Q_{t-4} \qquad R^2 = 13.6\%;\ p < 0.05 \qquad (1)$$
$$(33.28) \quad (0.1687)$$

$$S_t = 287.98 - 0.3134 \cdot Q_{t-3} \qquad R^2 = 6.5\%;\ p = 0.077 \qquad (2)$$
$$(34.49) \quad (0.1733)$$

$$S_t = 258.31 - 0.1526 \cdot Q_{t-2} \qquad R^2 = 1.6\%;\ p = 0.388 \qquad (3)$$
$$(34.70) \quad (0.1753)$$

$$S_t = 237.76 - 0.0494 \cdot Q_{t-1} \qquad R^2 = 0.2\%;\ p = 0.788 \qquad (4)$$
$$(34.72) \quad (0.1757)$$

Another reason for repeating the analysis for each sub-period is to reduce the effect of possible trends, seasonality, and changes in consumer behavior. Even though focusing on a market segment somewhat limits the degree of change in consumer preferences, trends are still not easy to capture. In some seasons (periods), the manufacturer may expect more sales compared to other ones, and this is usually a result of macroeconomic changes. Albuquerque and Bronnenbeg (2009) study at the quarter level when estimating demand assuming that the market conditions do not differ for the consumers who purchase a car in the same quarter. Similarly, we use 8-week periods when running our analyses in order to reduce the effect of changes in macro-economic conditions.

We treat vehicle options as product characteristics to define product configurations; we use binary coding when describing product configurations. Even though information on exterior paint and customer type was given in the original data set, we should note that we do not use exterior paint in product description and only consider the build-to-order vehicles in our analysis. The descriptive statistics of the options and MSRP are given in Table 4 below for the New York data set.

Table 4: Mean values of the options and list prices

Variables		**New York**
Price	MSRP	35,443 USD
Drive	FWD	31.269%
	AWD	68.731%
Body Style	SE	9.732%
	SEL	51.807%
	Limited	33.985%
	Sport	4.476%
Design: Trim Type	Charcoal Black	60.714%
	Leather	77.243%
Design: Others	Moonroof	66.070%
	Floor Mats	43.643%
	Headlamps	27.307%
	Roofrack	17.740%
	Ambient Package	77.243%
	Trailer Tow	9.778%
Design: Wheels	Premium Wheels	42.396%
	Trim Level 1	34.407%
	Trim Level 2	9.163%
	Trim Level 3	51.358%
	Trim Level 4	5.072%
	Tires 1	9.154%
	Tires 2	61.833%
	Tires 3	23.941%
	Tires 4	5.072%
Technology	Satellite Radio	97.762%
	Speed Control	11.191%
	Sync (Touch)	55.458%
	Blis	27.096%
	Navigation Center	50.257%
	Driver's Package	31.802%
	Comfort Group	38.782%
	Rearcamera	88.800%
Number of Observations		*10,902 observations*

4.2.1. Principal Component Analysis

Employing the encoding described in Chapter 3 and using varimax rotation, we obtained the principal components that are partially provided in Table X. Varimax rotation, which is an orthogonal rotation, was done only to obtain a reasonable interpretation. The total variance explained does not change after varimax rotation, but the variance explained by each component might. We should note that rotating the factor space is not necessary. It is not different than rotating a mirror to look nicer. Table 5 partially shows which options load on which components in the periods under consideration after varimax rotation.

Table 5: A Partial View of the Loadings of Some Options across Different Periods

Component	Period 1	Period 2	Period 3	Period 4
2	BLIS (-0.5) VISION PKG (-0.5)	BLIS (-0.5) VISION PKG (-0.5)	BLIS (-0.5) VISION PKG (-0.5)	BLIS (-0.5) VISION PKG (-0.5)
4	SEL (-0.3929) LIM (0.4635)	SEL (0.3577) LIM (-0.4110)	SEL (0.3690) LIM (-0.3926)	SEL (0.4482) LIM (-0.4691)
10	FWD (0.7071) AWD (-0.7071)	FWD (0.7071) AWD (-0.7071)	FWD (0.7071) AWD (-0.7071)	FWD (-0.7071) AWD (0.7071)
18	SE (0.8770) SEL (-0.5067)	SE (0.8280) SEL (-0.5419)	SE (0.8340) SEL (-0.5304)	SE (0.8827) SEL (-0.4321)
19	SPORT (0.9310)	LIM (-0.3542) SPORT (0.9200)	LIM (-0.3893) SPORT (0.9066)	LIM (-0.3355) SPORT (0.9257)

Based on Table 5, in Period 1, if there is a positive increase in the relative mix rates of both SEL and LIM vehicles, the scores of Component 4 also increases. For Periods 1-3, we can express the component associated with Component 10 as follows:

$$Comp10_i = 0.7071FWD_i - 0.7071AWD_i$$

Note that since $FWD_i = -AWD_i$, $Comp10_i = 2(0.7071)FWD_i$.

Option bundling, which is used to build predetermined sets of options so that external variety is controlled, is already being used extensively in the auto industry. In our case, the manufacturer also utilizes this strategy. For instance, *blis* and *vision package* options are always observed together in the original data. Thus, the loadings given in Table 5 is a byproduct of the encoding used since PCA after varimax rotation detects the natural bundling of the options.

4.3. Defining Vehicle Similarity Sets

Vehicle similarity sets are inputs for the demand and the optimization models that could be built to select the most profitable configurations. For each configuration, vehicle similarity sets are constructed by taking into account the interactions between configurations based on relative mix rates of options. Vehicle similarity sets are the most crucial part of the framework since we assume that substitutions occur within similarity sets.

Our approach is similar to the customer migration model proposed in Yunes *et al*. (2007), which aimed to determine the potential configurations a customer can purchase based on a few parameters such as the commonality factor, the set of fixed features, the first choice probability, etc. However, adopting a data-driven approach, we are primarily trying to find a way to group the configurations built for a region so that a reasonable number of product variants are taken into consideration; Yunes *et al*. (2007) suggest John Deere & Co. the use of customer migration lists within an optimization framework to decrease the size and complexity of production lines. Dichotomization and clustering approaches are used separately to extract the common information of the car options so that the practitioners do not suffer from the curse of dimensionality in further analyses. In any of these approaches, we do not claim to know why correlations between inputs exist, i.e., no inferences on causality are made.

4.3.1. Consumer Reviews

We are using consumer reviews not because the quantitative analyses are incomplete. Such qualitative models and industry expertise can provide better insights about the problem. Hogarth has a remarkable quote that can explain our intuition better: When driving at night with your headlights on, you do not necessarily see too well; however, turning your headlights off will not improve the situation (Hogarth, 1987). Besides, if the number of alternatives of a configurable product is abundant, it will be implausible for consumers to consider all options (car features) to make a choice (Adamowicz *et al*., 2008). Even though cars are durable goods, it is not always the case that consumers will consider all vehicle features at or before the point of purchase. Thus, consumer reviews are used to complement the qualitative methods. They provide good insights since they can help the practitioners to see significant differences based on important features, not only based on accessories. Other marketing models such as conjoint analysis or consumer clinics, which focus on stated preferences instead of revealed preferences, can also be employed. Thus, the use of consumer reviews is only one way of learning more about consumers and markets and a different angle to look at the problem to use in the decision making process.

According to the consumer reviews on the product of focus collected from online vehicle information data sources such as Cnet.com, Edmunds.com, and Kbb.com, consumers value all-wheel-drive configurations, especially the ones whose body style is Limited or SEL. They are also interested in most of the technology aspects such as the Sync system, Navigation Center, Rearcamera, and Blis. Configurations with leather seats and premium wheels are also appreciated. We collect consumer reviews for these reasons.

4.3.2. Dichotomization Approach

Before running the PCA, we created as many variables for each vehicle as the number of times they were observed during the first 32 weeks of the analysis. After obtaining the factor scores, we averaged these scores over the number of weeks each vehicle was available in order to maintain consistency. Then, dichotomization was done based on these new variables, which we will call *average relative mix rates*. We created one binary variable for each original factor, which takes the value of 1 if the average relative mix rate is positive, and zero, otherwise. Figure 6 illustrates the dichotomization approach for the vehicle with ID=187.

ID	Week	Factor Score 1	Factor Score 2	Factor Score 3	Factor Score 4
187	9	1.12838	-1.21321	-1.32430	-1.32298
187	10	1.12116	-1.20551	-1.31614	-1.31506
187	11	1.14121	-1.22521	-1.33701	-1.33896
187	12	0.98057	-1.06035	-1.16548	-1.16465
	Means	1.09283	-1.17607	-1.28573	-1.28541
	Grand Means	0.00000	0.00000	0.00000	0.00000

ID	Dimension 1	Dimension 2	Dimension 3	Dimension 4
187	1	0	0	0

ID	Binary Code
187	1000

ID	Vehicle Similarity Set
187	XX

Figure 6: Defining the vehicle similarity set for a vehicle

Since we selected 6 principal components for the analysis, we could obtain at most 64 vehicle similarity sets. The final number of similarity sets used in the analysis was 46. The percentages of each option in each vehicle similarity set are summarized in Appendix A.

The demand model we employ looks at market response at the individual level and estimates the purchase probabilities (level of attractiveness) of each similarity set of which sum equals the market share. The model needs the market share information a priori in order to calculate the purchase probabilities. In our study, we set it at 20% after consulting with industry experts.

The primary demands of each vehicle similarity set at different time periods following the dichotomization approach are shown in Figure 7. Primary demand corresponds to sales volume in almost each case, which means that demand due to substitution is quite low. This indicates that vehicle similarity sets are distinctive from each other; substitution is not frequently seen between vehicle sets, but it can be high between vehicles within the same similarity set.

Figure 8 shows how preference weights of each vehicle similarity set change over time after dichotomization is employed. A similarity set's relative attractiveness may change by time, which is expectable considering seasonality effect and demand fluctuations. We should also note that preference weights are assumed to be dependent of the consumer choice set at the point of purchase.

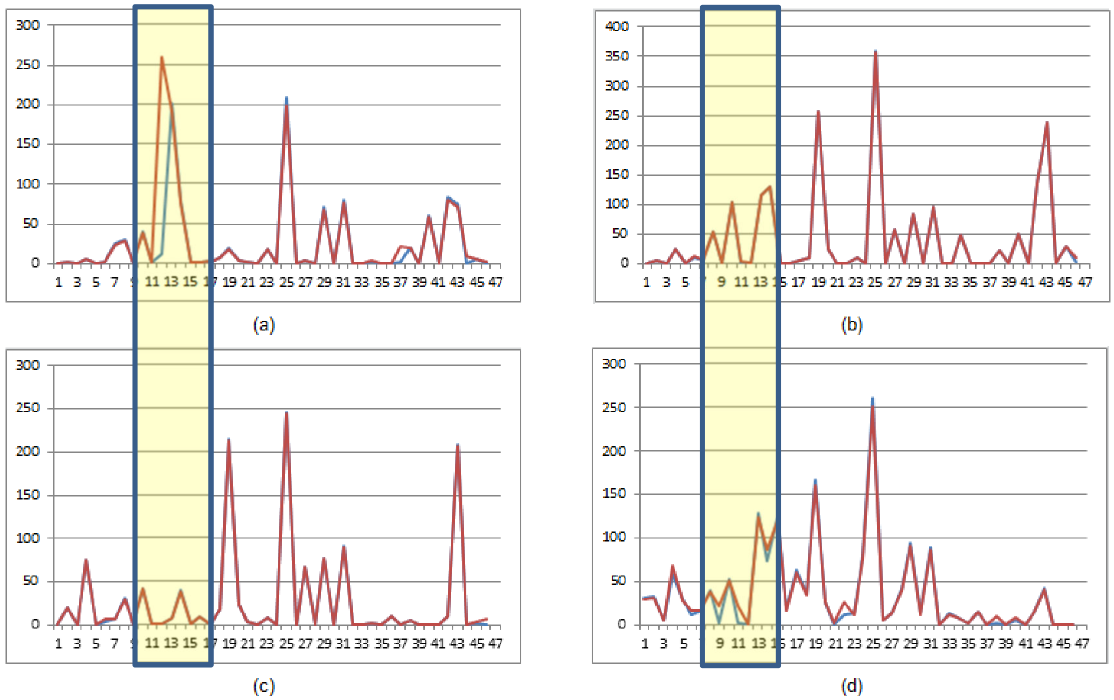

Figure 7: Sales (blue legend) versus primary demand (red legend) in (a) Weeks 1-8, (b) Weeks 9-16, (c) Weeks 17-24, and (d) Weeks 25-32

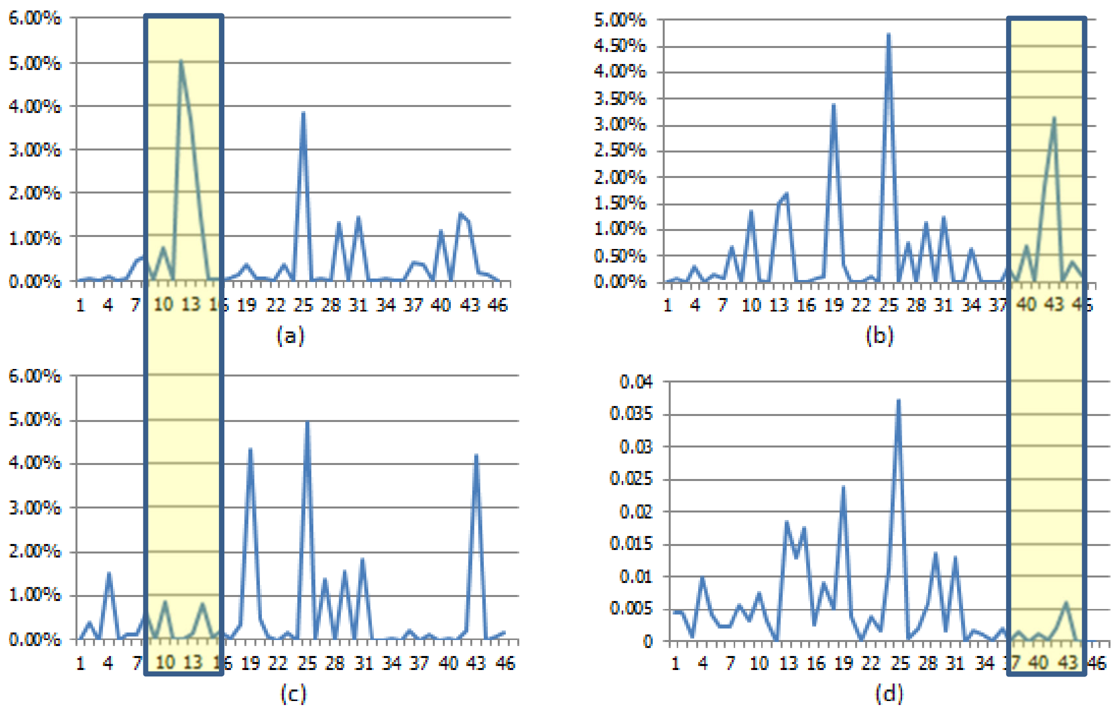

Figure 8: Relative attractiveness of each similarity set in (a) Weeks 1-8, (b) Weeks 9-16, (c) Weeks 17-24, and (d) Weeks 25-32 (Dichotomization)

Figures 7 and 8 illustrate that the level of attraction of a vehicle similarity set is not constant over time. This might be due to the changes in the market environment or in the marketing mix such as promotion and advertising, or due to other factors. The twelfth vehicle similarity set is found very attractive in the first 8 weeks of the study period; however, it does not appear to any consumers in the following periods. There was no vehicle from the similarity set #12 available after the second week of the study period. All 11 vehicles from this set were sold very fast (within the first two weeks), so the 12th preference weight was found to be very high. This is not a flaw of the estimation procedure; on the contrary, it helps us to see how ignoring product availability information could affect the conclusion. It also shows us how important it is to run such an analysis in sub-periods separately instead of running one analysis for the overall period. Table 6 below shows the time each vehicle from similarity set #12 spent on lot and their list prices. The average “days on lot” is 6.64 days (less than a week) and the average list price is $ 38,107.56. This set, in which all the vehicles have Limited body style, headlamps, moonroof, rearcamera, satellite radio, leather seats, ambient package, drivers package, and navigation center, is well distinguished. It is also composed of mostly all-wheel drive vehicles. Note that most of these options were also appreciated in consumer reviews.

Table 6. The days on lot and list prices of the vehicles in the 12th similarity set

Days on Lot	MSRP (in USD)
7	36,239.29
3	37,922.29
4	37,922.29
5	37,922.29
11	37,922.29
11	37,922.29
6	37,986.29
11	38,683.29
3	38,747.29
3	38,747.29
9	39,168.29

The 19th similarity set, one of the largest sets obtained after dichotomization, attracted too many customers in the last three sub-periods, which is not surprising considering its similarity to the 12th set. Table 7 compares the mix rates of each option in these two similarity set.

Table 7: The structures of the vehicle similarity sets #12 and #19

	12	19
FWD	9%	7%
SE	0%	0%
SEL	0%	0%
LIMITED	100%	99%
SPORT	0%	1%
Charcoal Black Trim	73%	53%
Sync (Touch)	0%	0%
Floormats	36%	59%
Headlamps	100%	96%
Moonroof	100%	98%
Rearcamera	100%	100%
Roofrack	0%	1%
Satelliteradio	100%	100%
Speed Control	0%	69%
Trailertow Pkg	0%	15%
Leather	100%	100%
Premium Wheels	0%	1%
Ambient Pkg	100%	100%
Blis	0%	100%
Comfort Grp	0%	0%
Drivers Pkg	100%	99%
Navigation Center	100%	99%
Trim Level 1	0%	6%
Trim Level 2	0%	0%
Trim Level 3	100%	94%
Trim Level 4	0%	0%
Tires 1	0%	0%
Tires 2	100%	93%
Tires 3	0%	7%
Tires 4	0%	0%

We should also note that the vehicles from the 19th set started to arrive more frequently starting with the fourth week of the study period. Until then, the vehicles from the 12th set were already sold. The average days on lot for the vehicles in the 19th set is around 77 days (almost 2 months), thus the preference weights of these two sets indicate the difference in days on lot, which denotes how fast a vehicle leaves the dealer.

According to Figures 7 and 8, note also that the 25th vehicle similarity set along with the 14th, 29th, 31st, 43rd sets attracted too many customers in all of the four sub-periods.

Figure 9 gives a three-dimensional look at the preference weights of the change in each similarity set changing between 8-weekly periods. The *x*-axis denotes the sales of each similarity set. Likewise, Figure 10 illustrates how preference weights changed with time based on the percentage of front-wheel drive vehicles present in the similarity sets. Figures 11, 12, 13, and 14 are the similar graphical representations when the *y*-axis is selected to denote the SE, SEL, Limited, and the Sport body styles, respectively.

Based on Figure 9, it is apparent that even the level of attraction of the best-selling vehicles can change significantly over time. Figure 10 shows that vehicle similarity sets with more all-wheel drive vehicles were found to be relatively more attractive. A similar conclusion cannot be made easily using the figures for the body styles; however, we can clearly see that consumers were more attracted to the vehicles that do not have an SE body style.

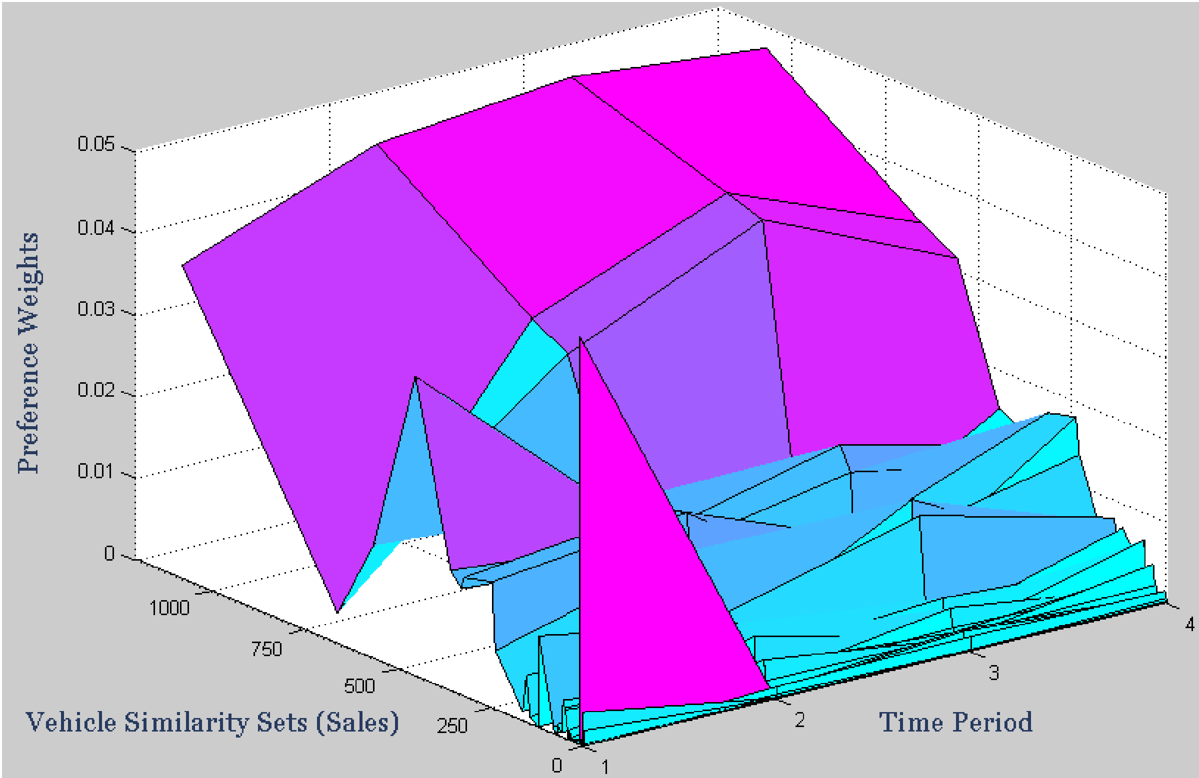

Figure 9: Change in preference weights with time as sales indicating the vehicle similarity sets (Dichotomization)

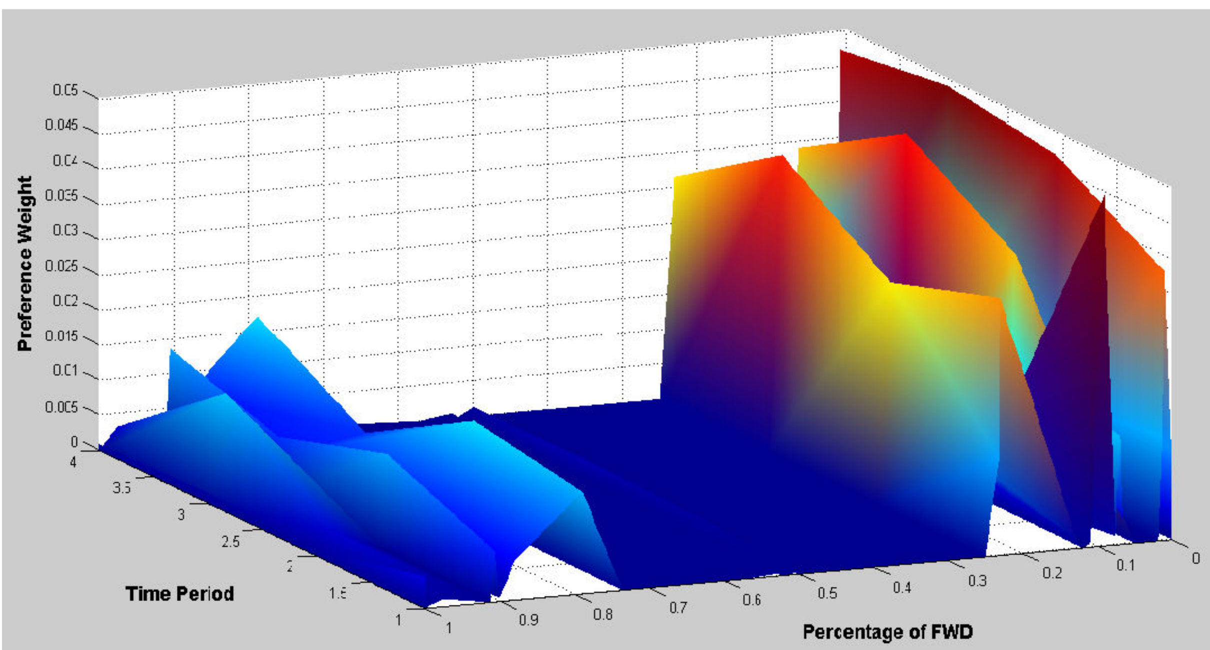

Figure 10: Change in preference weights by time with percentage of front-wheel drive indicating vehicle similarity sets (Dichotomization)

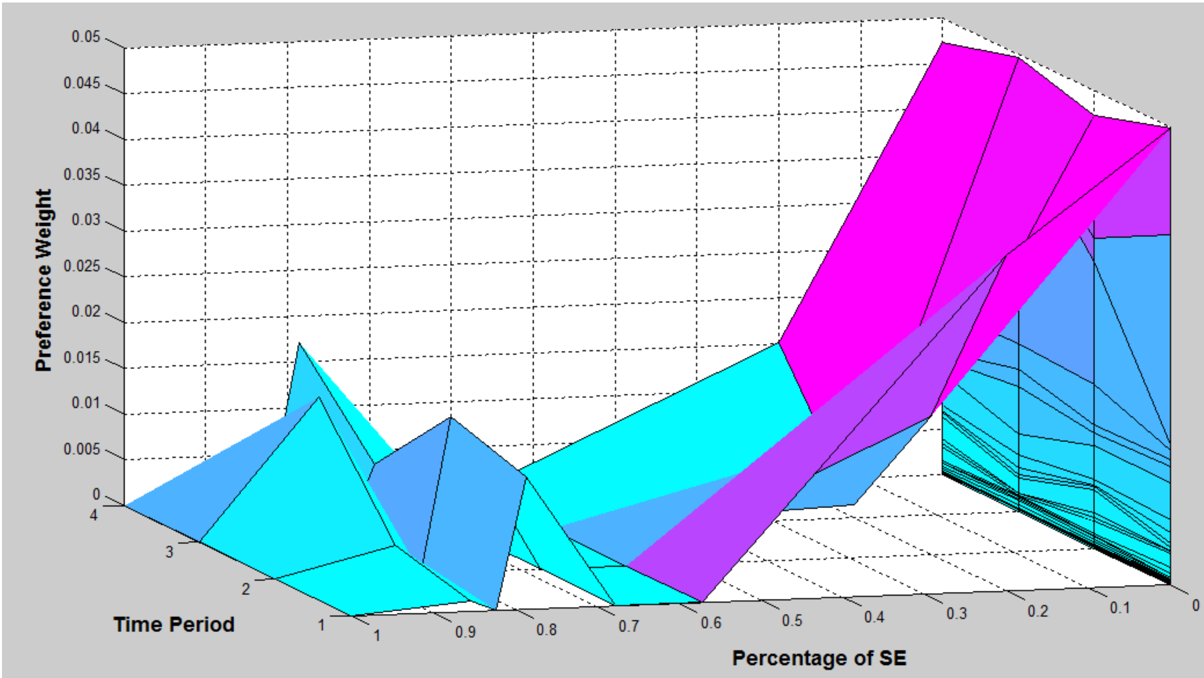

Figure 11: Change in preference weights by time when percentage of SE vehicles indicating the structure of vehicle similarity sets (Dichotomization)

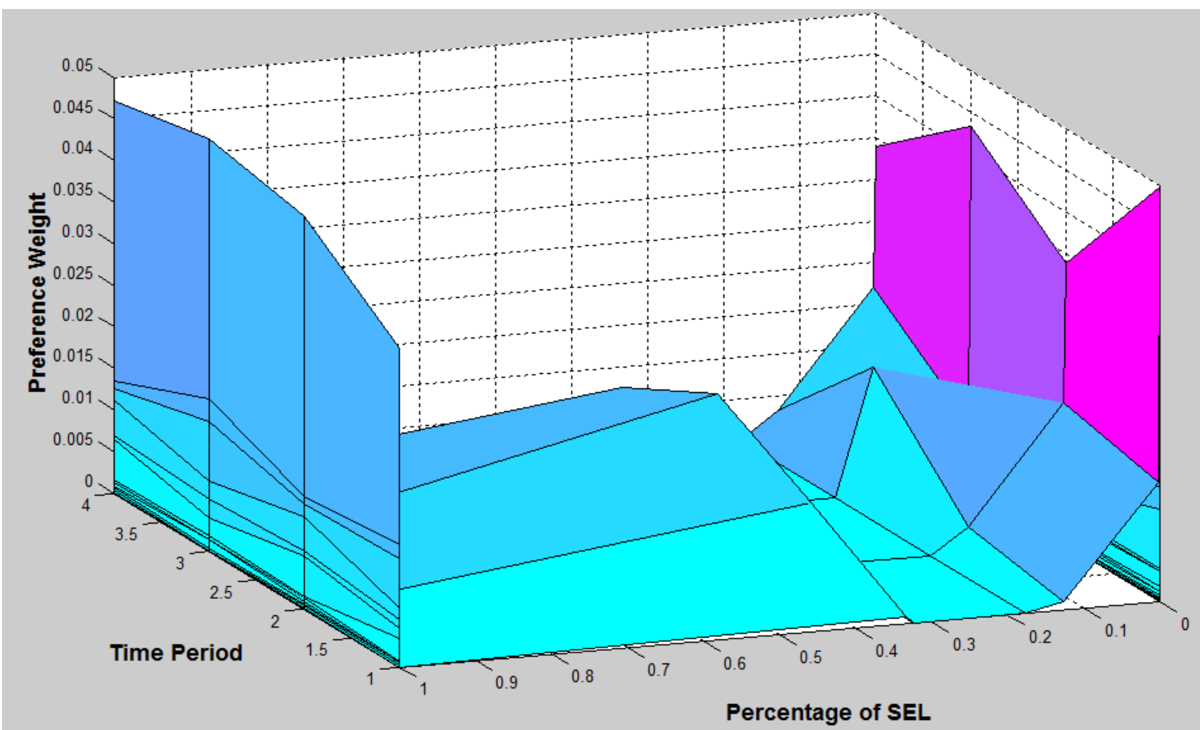

Figure 12: Change in preference weights by time when percentage of SEL vehicles indicating the structure of vehicle similarity sets (Dichotomization)

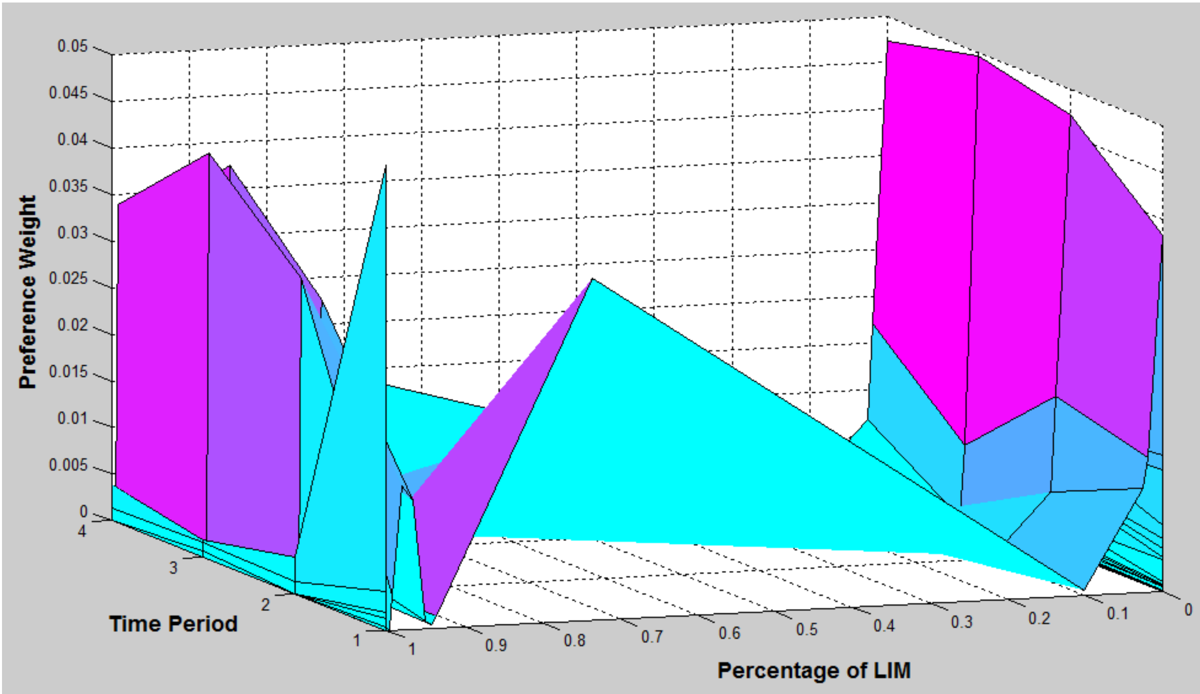

Figure 13: Change in preference weights by time when percentage of Limited vehicles indicating the structure of vehicle similarity sets (Dichotomization)

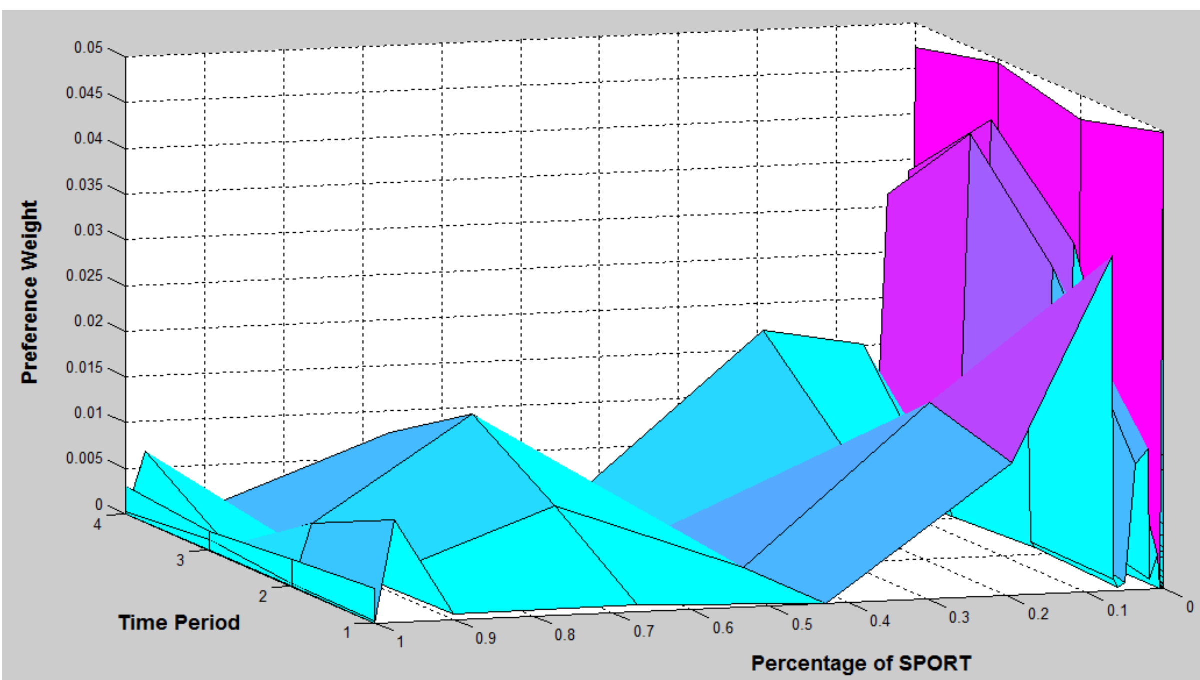

Figure 14: Change in preference weights by time when percentage of Sport vehicles indicating the structure of vehicle similarity sets (Dichotomization)

4.3.3. Clustering Approach

As an alternative to the previous approach, instead of using the relative mix rates of options, we try grouping the configurations based on their normalized list prices taking the number of weeks they spent on lot into account. Thus, we define $\overline{\overline{MSRP_{ik}}}$ for each observation *i* in period *k* as

$$\overline{\overline{MSRP_{ik}}} = \left(\frac{MSRP_{ik} - E_{j \in S}(MSRP_{ik})}{Var_{j \in S}(MSRP_{ik})} \right)$$

when forming clusters.

Since we intended to employ a data-driven approach, we tried to let the data tell us which steps we should take when describing the similarity sets. Thus, we preferred using a hierarchical method to a non-hierarchical one such as *k*-means in which the value of *k* has to be chosen in advance even though hierarchical clustering algorithms need more computational resources.

In hierarchical cluster analysis, every data point starts as a cluster, and then clusters are combined into larger groups based on similarity measures (linkage rules). We are using unweighted pair-group average as the linkage rule when clustering the configurations as numerous Monte Carlo studies claimed that this method most often leads to the best solution. Thus, its use is highly recommended in practice. There is no significant test in cluster analysis that can help us whether or not our results are meaningful; however, since we have a priori information about the product definitions, we can greatly benefit from cluster analysis.

Using the average linkage rule, we formed clusters for each sub-period separately, and with the help of dendograms given collectively in Figure 15 below, we decided to use 12 clusters to estimate the preference weights for each cluster (vehicle similarity sets). Note that consistency between different sub-periods was satisfied before doing further analyses.

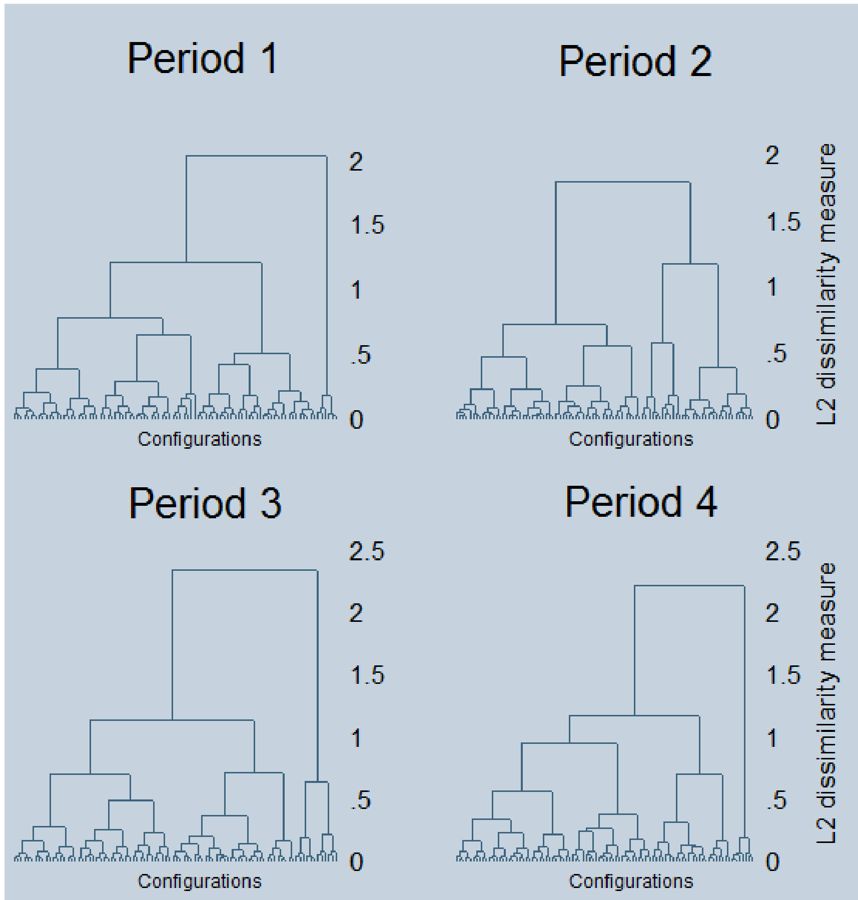

Figure 15. Dendograms from Each Time Period

As seen in Figure 16 below, excluding Cluster #4, #5, and #9 the sizes of the clusters are do not significantly differ.

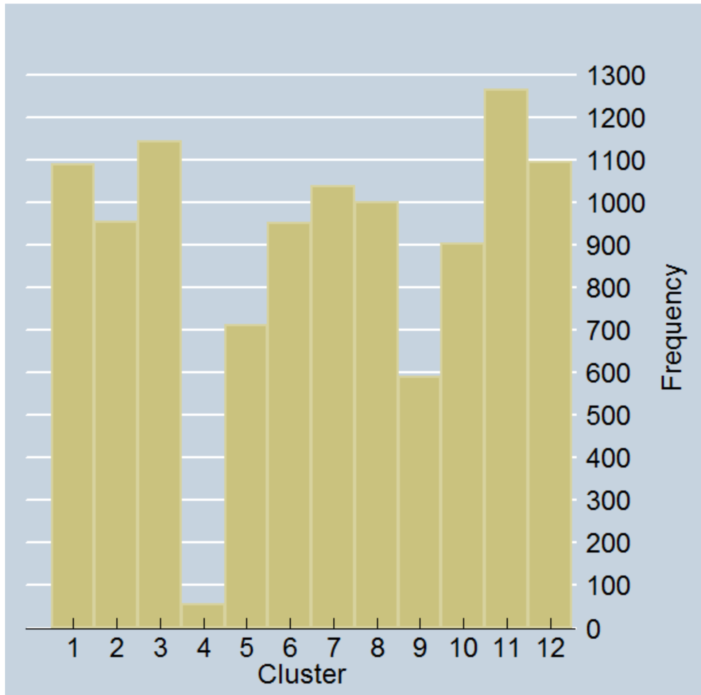

Figure 16: Number of Vehicles in Each Cluster

Table 8 below reports the summary statistics of the list prices of each cluster. According to the table, Cluster #7 is most likely composed of standard configurations, whereas Cluster #4 consists of high-end configurations.

Table 8: Descriptive Statistics of MSRP of Each Cluster

Cluster	Mean	Median	Std Dev
1	34122	34057	1149
2	39631	39894	775
3	35858	36046	894
4	41767	42013	708
5	38798	38996	803
6	32882	32541	1733
7	27971	28028	392
8	35013	35113	1050
9	37972	38140	868
10	39947	40176	887
11	36532	36593	1652
12	33663	33319	1635

We are not reporting the primary and substitute demand estimates as substitution between different clusters is minimal. This insight is also supported by Table 9 below.

Table 9: Performance measures for demand estimation using the clusters

Period	No. of Iterations	Total Primary Demand	Total Substitute Demand	Total Demand for No-Purchase	% of Lost Sales	Recapture Rate
1	7	948.454	2.546	0	0	0.26844%
2	6	1714.591	0.409	0	0	0.02388%
3	6	1109.474	0.526	0	0	0.04744%
4	6	1431.600	0.400	0	0	0.02791%

The preference weights of each cluster in each period are given in Figure 17. The most stable clusters in terms of attractiveness are Clusters #3, #5, and #12.

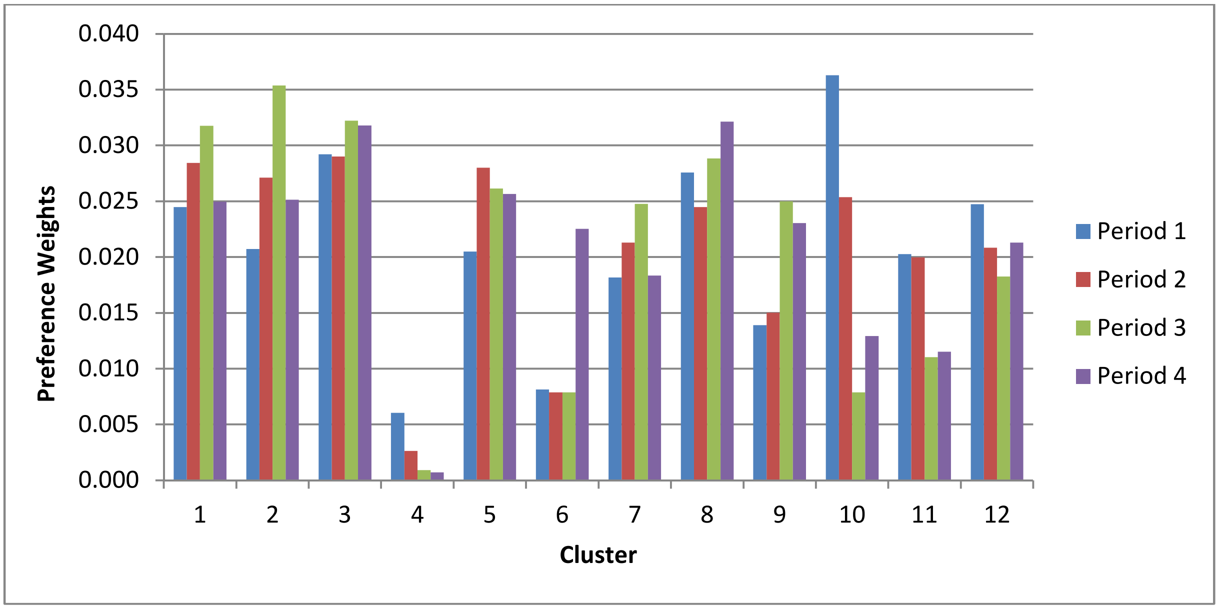

Figure 17: Preference Weights (Clustering Approach)

4.3.4. Subjective Selection

In the previous approaches, we relied on data when describing vehicle similarity sets; however, one might also want to re-configure product variants based on subjective judgments. We will prefer the terms *product variant* and *product configuration* throughout this subsection to describe vehicle similarity sets. The experimental results obtained when the full configuration space is used will be summarized right after the results for the limited information case.

Following the subjective approach, we define the product variants as shown in Table 16 in Appendix. We excluded the all-wheel drives with the SE body style from further analysis as it was not observed in the first 32 weeks of the study period. Full information approach uses all vehicle options (except paint and exterior trim level) and uses binary coding when describing product configurations.

Table 10 summarizes the preference weights estimated for each product variant in each period. In parallel with the consumer reviews and the results obtained from dichotomization approach, all-wheel drive vehicles with a body style of SEL or Limited are found to be consistently attractive among the four sub-periods. All the other product variants show some level of fluctuation as the period changes.

Table 10: Preference weights obtained using the subjective approach

Product Variant	Coefficient			
	Period 1	Period 2	Period 3	Period 4
FWD SE	0.0161	0.0195	0.0224	0.0158
FWD SEL	0.0248	0.0231	0.0171	0.0207
FWD Limited	0.0084	0.0129	0.0091	0.0079
FWD Sport	0.0010	0.0009	0.0010	0.0008
AWD SE	N/A	N/A	N/A	N/A
AWD SEL	0.0908	0.0852	0.0915	0.1087
AWD Limited	0.0885	0.0928	0.0898	0.0792
AWD Sport	0.0205	0.0156	0.0191	0.0170

As a benchmark to the models estimated so far, we repeated all the steps on the configuration space also. Using 26 binary options, we obtained 1,210 configurations that were built for the region under study. It is very inconvenient to show the detailed results of the full information case; however, in Table 1Y in Appendix A, we summarize the structures of the most attractive configurations in the first 32-weeks of the analysis. On the other hand, Table 11 shows the estimated preference weights of these configurations in each period.

Table 11: The preference weights of the most attractive configurations

Period 1		Period 2		Period 3		Period 4	
Configuration	**Pref. Weight**	**Configuration**	**Pref. Weight**	**Configuration**	**Pref. Weight**	**Configuration**	**Pref. Weight**
1092	0.69%	424	0.61%	275	0.62%	224	0.63%
416	0.52%	416	0.55%	227	0.56%	568	0.53%
690	0.51%	385	0.50%	424	0.51%	403	0.48%
1033	0.51%	227	0.47%	568	0.51%	424	0.44%
424	0.49%	216	0.43%	403	0.49%	416	0.41%
568	0.40%	224	0.42%	423	0.42%	803	0.36%
451	0.37%	403	0.41%	85	0.40%	220	0.35%
403	0.35%	568	0.37%	224	0.40%	385	0.34%
385	0.34%	536	0.34%	385	0.40%	312	0.34%
84	0.33%	220	0.30%	416	0.39%	702	0.34%

Table 12 summarizes the performance measures of the full information demand model. The percentage of lost sales, which is the percentage of time customers walked away due to configuration unavailability, is found to be acceptable in each period. The recapture rate, which shows the total percentage of customers that substituted the unavailable configuration they primarily demanded with an available configuration, is relatively low; however, this is quite reasonable considering the level of study. We estimated our models at the regional level, where the level of configuration unavailability is much smaller than the one at the dealer level; thus, the analyst should pay attention to this fact when interpreting the recapture rate.

Table 12: Performance measures for demand estimation
using the configuration space

Period	No. of Iterations	Total Primary Demand	Total Substitute Demand	Total Demand for No-Purchase	% of Lost Sales	Recapture Rate
1	263	1230.98	103.38	326.93	26.56%	8.40%
2	147	2138.17	162.69	497.54	23.27%	7.61%
3	115	1397.23	112.39	339.61	24.31%	8.04%
4	154	1777.01	141.02	410.84	23.12%	7.94%

4.4. Analyzing Assortment Structures

Assortment analysis can provide greater insight on the data sets contained information about vehicle availability at the dealer level. It may not say much at the region level; however, it can provide better insight when it is done at the dealer level. One could consider adding trend or supply as additional covariates when modeling the percentage/number of cars sold. We do not believe that it would be a good idea to comment on whether the consumers are utility maximizers or not since the level of analysis is not sufficiently deep.

The relationship between the number (percentage) of cars sold and the entropy measure defined in Chapter 3 is given in Figure 18 (Figure 19) for full information. The maximum number of cars sold (321) when the entropy measure of the regional assortment is used was approximately 12.694. This is also the value of the entropy measure when the percentage of cars sold was at maximum (20.34%). Note that it is the 10th week of the analysis when the entropy measure was at this level. Figure 18 (Figure 19) suggests that entropy and number (percentage) of cars sold have a non-linear relationship (flat).

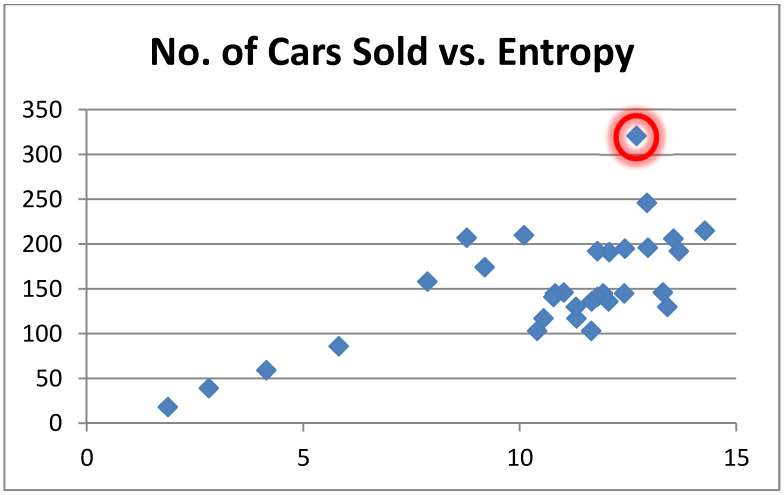

Figure 18: The relationship between number of cars sold and entropy of an assortment (Full Information)

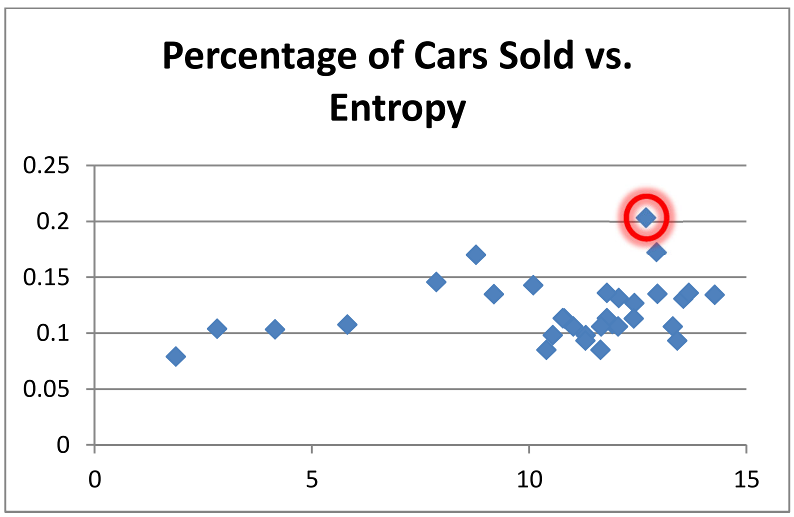

Figure 19: The relationship between percentage of cars sold and entropy of an assortment (Full Information)

The relationships between these measures when the subjective approach was followed are illustrated in Figures 20 and 21 below. Both the percentage of cars sold and the number of cars sold are at their maximum when the entropy was measured as 225.761 (regional assortment in the 10th week).

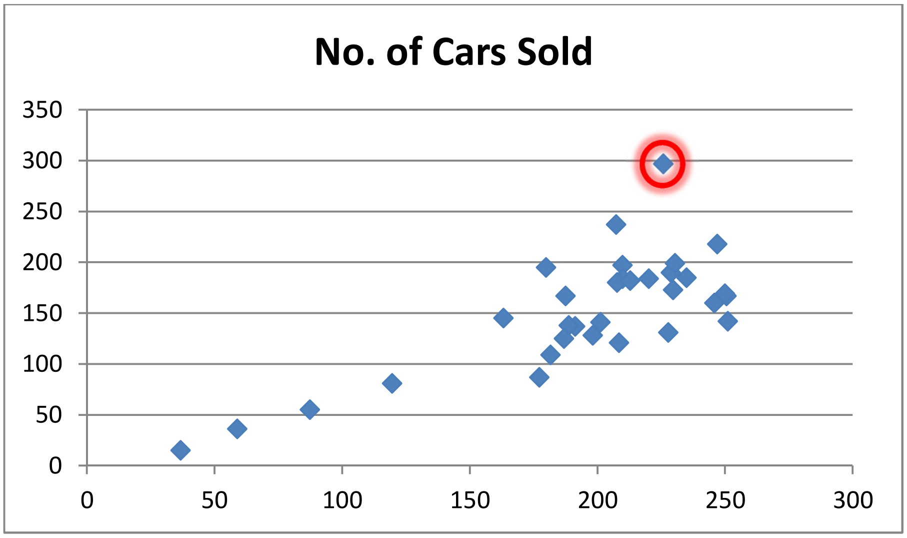

Figure 20: The relationship between number of cars sold and entropy of an assortment (Subjective)

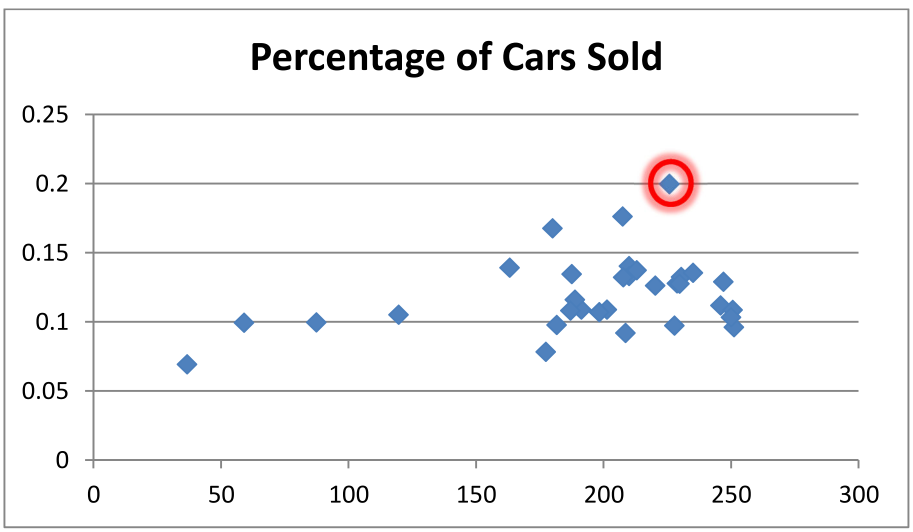

Figure 21: The relationship between percentage of cars sold and entropy of an assortment (Subjective)

The two approaches (subjective judgment and full information) both emphasize that the assortment structure in the 10th week is the best in terms of number of vehicles sold. The high entropy in this week shows that the alternatives were closer to each other in terms of relative attractiveness. Table 13 shows the number of each product variant (described following the subjective judgment) available in the 10th week; Figure 22 illustrates how many cars were available of each configuration in this week. Figure 22 shows that the configurations with AWD SEL (#416 and #424) and the configuration with FWD SE (#85) were supplied the most. This finding is also supported by Table 13.

Table 13: The assortment structure in the 10th week (Subjective Selection)

Product Variant	Number of Cars Available
FWD SE	252
FWD SEL	165
FWD Limited	86
FWD Sport	6
AWD SE	0
AWD SEL	462
AWD Limited	503
AWD Sport	92

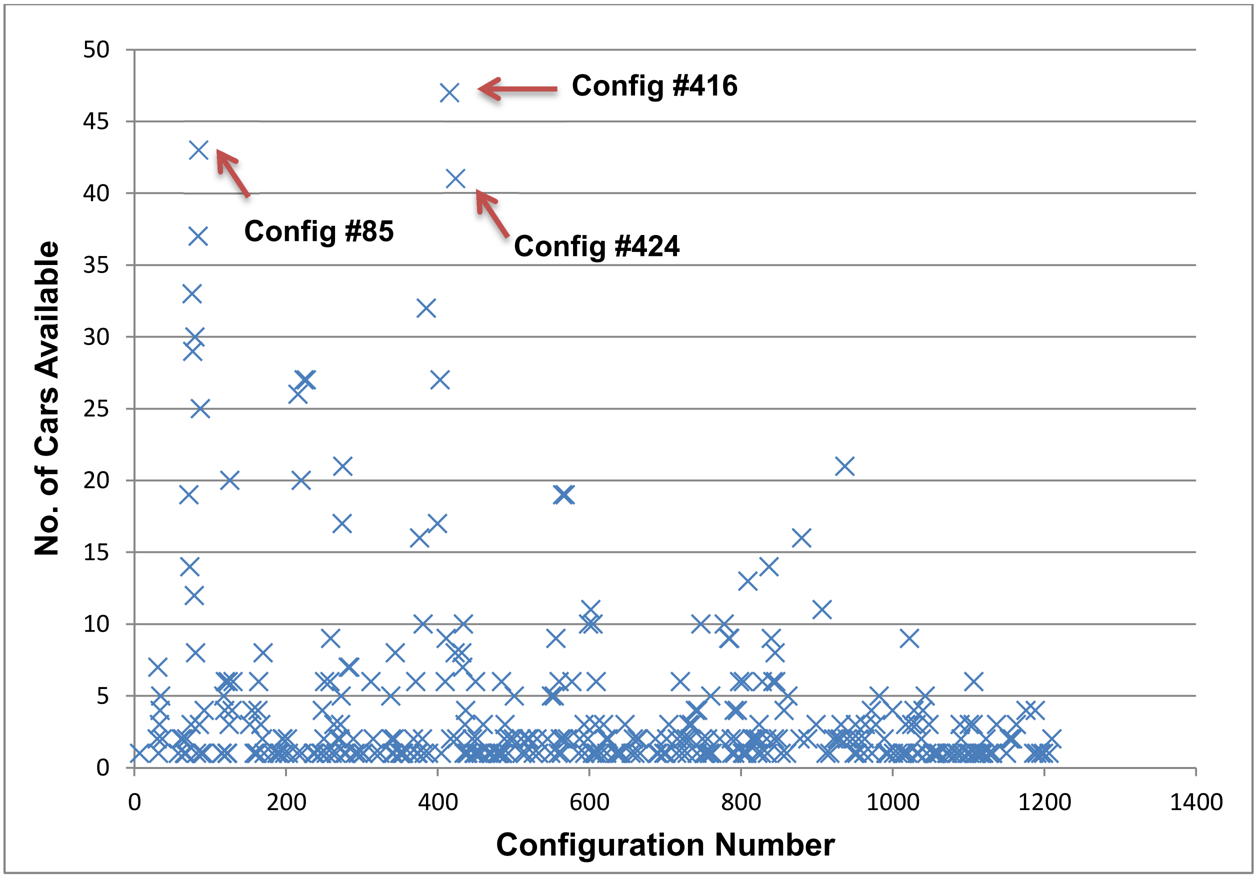

Figure 22: Number of Cars Available per Configuration in the 10th Week

Figure 23 provides a closer look to the configurations #84, #416, and #424. As shown in Table 11 above, configurations #416 and #424 did sell very well in all sub-periods. They were in the same product variant group when subjective judgment approach was employed. These two configurations only differ on the *floor mats* option. Since most consumers would not insist on having or not having this option, it is not surprising to find that they are substitutable.

Figure 23: A Partial View of Three Best Selling Configurations

	Configuration		
Option	84	416	424
drive	FWD	AWD	AWD
body	SE	LIM	LIM
charcoalblacktrim	1	1	1
sync (touch)	1	0	0
floormats	0	0	1
headlamps	0	1	1
moonroof	0	1	1
rearcamera	0	1	1
roofrack	0	0	0
satelliteradio	1	1	1
speedcontrol	0	1	1
trailertowpackage	0	0	0
leather seats	1	0	0
premium wheels	1	0	0
ambient pkg	0	1	1
blis	0	1	1
comfort grp	0	0	0
drivers pkg	0	1	1
navigation center	0	1	1
tires1	1	0	0
tires2	0	1	1
tires3	0	0	0

4.5. Survival Analysis

The relative mix rate gives us an idea about the absence/presence of an option on a vehicle and the mix rate of that option in the region in a week. Most of the time, the real mix rates of options are not equal to 0% or 100%, so a vehicle has a positive relative mix rate if it has an option that is not present on all the vehicles in a given week. When we sum the relative mix rates of an option of all the vehicles in a week, we obtain 0. Moreover, the relative mix rates are always the elements of the interval [-1,1]. If the absolute values of the relative mix rates in a week are close to each other, then there is a theoretically perfect balance between the absence and the presence of the option.

Since there is collinearity between the relative mix rates of different options, we use PCA to avoid multicollinearity related problems. PCA also helps us work with fewer variables when estimating the hazard rates. Note that the factor scores are linear combinations of the relative mix rates, so the factor scores in this context are indices that indicate the relative mix rates of bundles.

Figure 24 below illustrates the hazard rates (HRs) estimated weekly for each sub-period using these six factors. Note that in survival analysis, the change in survival conditions is assumed to be fixed over time (Tabachnick and Fidell, 2007). This is the primary reason why we estimate days on lot weekly.

When more than one observation fails at a certain time, which is almost always the case in practice, the partial likelihood function needs to be approximated (Jenkins, 2008). The most common techniques used to estimate the survival parameters are the Breslow Method, the Efron Method, the Exact Partial Likelihood Method, and the Exact Discrete Partial Likelihood Method. We are reporting the results of the Breslow Method in Figure 24, but we must note that the results did not significantly deviate from the ones we are reporting.

According to Figure 24, the HRs did not fluctuate substantially until the 17th week. In the first 8-week period, the HRs of all the factors are close to each other across the weeks, except the first week. In periods 1 through 3, the hazard ratios of the component mostly associated with FWD was found smaller than 1. Thus, a one-unit increase in δ_{ik}^{FWD} (denotes whether vehicle i is a front-wheel drive or not) leads to a $1-\frac{1}{|S|}$ increase in the relative mix rate of FWD, and a $1.442\left(1-\frac{1}{|S|}\right)$ increase in the component FWD/AWD. Let $\alpha = 1.442\left(1-\frac{1}{|S|}\right)$, which is always positive. Then, the baseline hazard rate is $\exp(\alpha\beta_{FWD})$ times as large as in the baseline scenario, indicating a smaller hazard, and hence, an increase in the expected duration (days on lot). In other words, if the vehicle was a FWD, its expected days on lot would be longer.

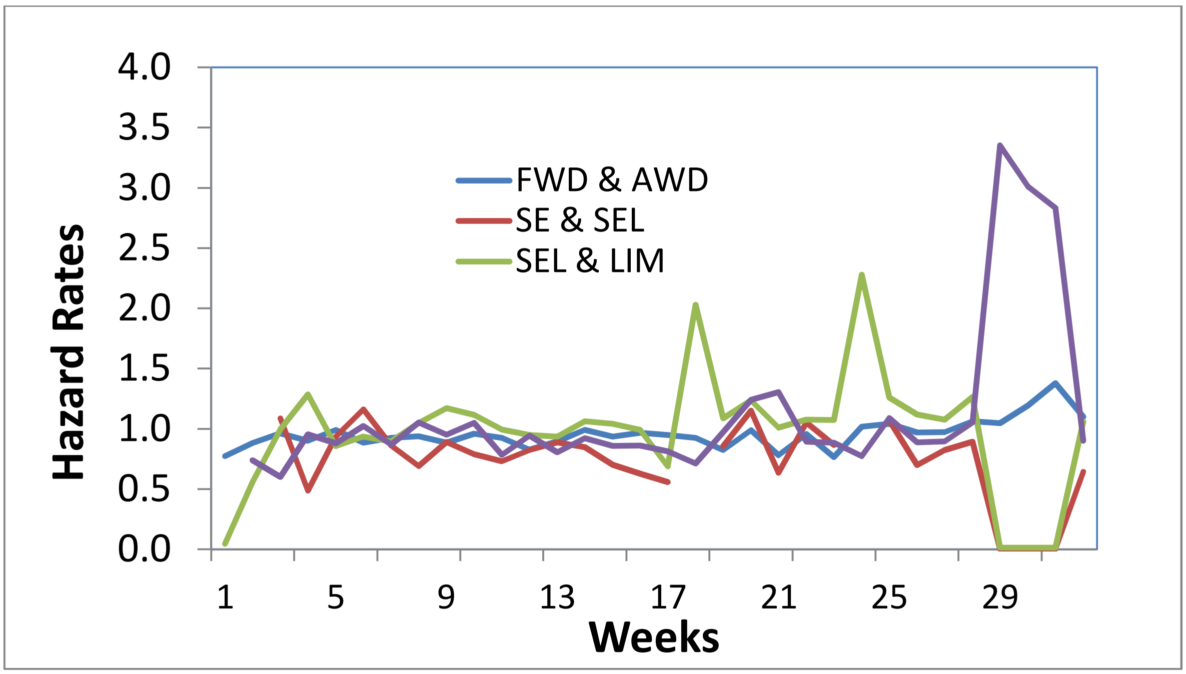

Figure 24: The Principal Component Based Hazard Rates across Time

4.6. Concluding Remarks

We used a univariate approach to demand modeling; i.e., we built a model separately for a single product without accounting for competition. However, in case the product of interest is correlated with other products within or outside the brand (or product family) of focus, the demand model can easily be altered as multivariate in order to better estimate relative attractiveness of each product variant. If the product variants are described properly using the data set in hand, it is not vital to study at disaggregate level or at aggregate level. On the other hand, since we did our analyses in 8-week periods, we reduced the degree of inter-temporal dependence of the product variants. In other words, if there was a seasonal effect or a trend, we partially accounted for that. We also accounted for product dependence of demand by collecting highly substitutable products in the same set.

A utility based demand model can also be used in our case; however, such a model should not fail to consider heterogeneity in consumer preferences. If one had represented the potential car buyers by a single agent, he/she must have shown that the distributions of days different cars spent on lot were similar (when controlling for marketing variables). Some studies in the assortment planning literature failed to take this into account by using locational choice models of demand, in which consumers are assumed to have similar preferences. Such a representative approach should not be employed in case of high-priced configurable products such as automobiles.

Although sophisticated demand estimations can be easily made by today's cutting-edge computers, most demand models are inapplicable in the case of configurable products as they suffer from the abundance of alternatives that should be considered. As the number of product variants increases, they begin converging slowly; in some cases, they do not converge at all. In our experiments, we had at most 1,210 configurations; however, in real life, there might be tens of thousands of configurations even at product level (e.g., Ford F-150). This is why shrinkage of the configuration space is essential.

We focused on a single market segment assuming that the environment is static and there is no competition (or the effect of competition is singled out). Thus, if the decision maker does not take competition into account, her only focus should be short-run profit as our models cannot be used as a standalone approach when the goal is maximizing long-profits. To evaluate marketing actions and improve profit margins or market shares, the decision makers should definitely take competition into account (employing the optimization model). However, note that when competition is considered, the configuration space may substantially increase.

Note also that the vector of preference weights is the exponential of the product of beta weights and the vector of product attributes: $\boldsymbol{v} = e^{\boldsymbol{\beta}'\boldsymbol{z}}$. In our case, the median of the list prices can be used as an attribute to estimate the corresponding beta coefficient.

The MNL model is embedded in the demand estimation model we benefit from. The literature underlines one of the disadvantages of using discrete choice models and their alike as the inability in considering more than one product at a time. However, in case of durable and configurable goods, practitioners are not affected by this disadvantage since the possibility of a consumer buying two of such goods in one store visit is very low.

Note that the first 8 weeks is the transient phase of the sales season, so the results obtained in this period should be approached carefully. This is a possible reason why the 12th set was not observed after the first 8 weeks. However, future research should consider development of techniques for further dimension reduction of similarity sets.

CHAPTER V: CONCLUSION AND FUTURE WORK

The decision of what configurations of a product to offer is a difficult one for most marketing and sales departments. Economies of scale suggest smaller numbers of configurations to keep unit costs low. Conversely, realizing high sales requires a large number of configurations to ensure that customers are satisfied with the available selection. Retailers also need to know the right level of variety to make decisions regarding the assortment structure and inventory.

Existing decision support tools for managing external variety cannot support complex configurable products. Since the abundance of too many product variants to consider affects the efficiency of demand and assortment planning models, practitioners need a suitable technique to reduce the configuration space. In this dissertation, we propose a flexible framework and associated techniques to overcome this limitation. We develop decision support models that can help decision makers when they use operational models to manage the external variety of configurable products, and the type of analysis we propose can generate reliable and meaningful inputs to the futher steps of decision making (e.g., assortment optimization model).

Our study supports different levels of understanding and analysis. Due to its ability and flexibility, it can be employed under various marketing actions. It should be used primarily for marketing and sales; however, it can also guide product development-related decisions. This is not a retrospective study, and we believe that the insights resulting from the proposed methodology can shed some light on designing new products.

Our overall framework is quite logical and intuitive, but as Lillien *et al*. (2007) noted, “marketing engineering succeeds because of sophisticated managers, not sophisticated models.” Our proposed framework helps transform quantitative and qualitative data about a market segment into sub-decisions that can help improve strategic decisions about pricing, bundling, options development, and so on. The marketing literature is full of different models that can answer our main research questions: ‘Which configurations should be offered?’, ‘How many configurations should be offered?’, and ‘Which options should be bundled?’ For instance, stated preferences data are extremely valuable and can improve experimental results obtained in this study. However, one should note that we are proposing our framework as an additional tool to extract deeper understanding and explore what more we can learn.

To the best of our knowledge, this is the first attempt to develop a data-driven framework for the external variety of configurable products using historical data. We are contributing to the literature by showing what can be done with limited data (information) when making decisions in case of configurable products. Our framework would not improve when richer data sets were used; however, our experimental results could. For instance, the EM-based demand model we employ cannot account for heterogeneity in consumer preferences; therefore, the experimental results due to this limitation might be biased compared to the case where complete information on consumer demographics and characterists is available when modeling demand. However, this bias is not due to the proposed approach as the EM-based demand model can easily be replaced with a solely utility based demand model that can account for heterogeneity. The second contribution is providing a proper way to shrink the size of product (configuration) space by taking the interactions between product features (options) into account (i.e., determining super-options) and creating product entities (similarity sets). We are also contibuting to the literature by discussing how to employ survival analysis in order to argue about the effect of product attributes on the expected survival time

Some possible avenues for future research in the durable goods market could investigate the effect of sales person (e.g., the persuasive ability of most sales representatives in dealers) on the consumer’s decision making process, the use of discrete time duration models in case of configurable goods, and significance of environmental conditions (competition, market dynamics, etc.) and consumer perception on the assortment structure in case of configurable products.

We attempted to estimate how long products (configurations) remain unsold using a simple survival model. Future research should benefit from frailty models, which are used to estimate hazard rates in case there are omitted variables, as unobserved heterogeneity might be significant. Discrete time duration models and survival models treating mix rates as time varying covariates should also be considered as a benchmark.

Future research should benefit from the EM-based demand model employed in this study by considering all the products available in the market segment. In such a case, competition should definitely be taken into account and market share parameters should be set more carefully.

The insights gained for the present design of a product can have a substantial use when designing the future designs of the product. However, a more in-depth analysis is needed in order to make conclusions about a completely new product.

Endogeneity might be seen as a problem in our experimentation since transaction prices are unknown to us. Even though Crafton and Hoffer (1980) state that MSRP can be used in case transaction prices are not reported, future research should address endogeneity and try to overcome it.

APPENDIX: STRUCTURES OF VEHICLE SIMILARITY SETS

Table 14: The percentages of each option in each vehicle similarity set after dichotomization

	Vehicle Similarity Sets											
Options	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
FWD	52%	0%	91%	98%	26%	0%	100%	88%	50%	91%	67%	9%
SE	0%	0%	82%	85%	0%	0%	0%	0%	0%	0%	0%	0%
SEL	100%	100%	18%	13%	100%	100%	100%	100%	100%	100%	100%	0%
LIMITED	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
SPORT	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%
Charcoal Black Trim	66%	76%	44%	59%	59%	57%	45%	58%	50%	54%	0%	73%
Sync (Touch)	100%	100%	82%	85%	100%	97%	78%	91%	100%	92%	67%	0%
Floormats	39%	33%	22%	21%	45%	67%	45%	28%	50%	28%	0%	36%
Headlamps	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Moonroof	90%	71%	0%	2%	73%	77%	93%	38%	50%	49%	0%	100%
Rearcamera	100%	100%	0%	2%	100%	100%	100%	100%	100%	100%	100%	100%
Roofrack	4%	10%	53%	60%	15%	13%	3%	34%	50%	26%	100%	0%
Satelliteradio	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Speed Control	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trailertow Pkg	5%	8%	2%	1%	15%	10%	7%	12%	0%	12%	33%	0%
Leather	100%	100%	0%	2%	100%	100%	100%	0%	0%	100%	100%	100%
Premium Wheels	26%	56%	100%	100%	41%	63%	52%	69%	100%	45%	67%	0%
Ambient Pkg	100%	100%	0%	2%	100%	100%	100%	0%	0%	100%	100%	100%
Blis	0%	0%	0%	2%	46%	97%	0%	0%	0%	15%	33%	0%
Comfort Grp	100%	100%	0%	0%	100%	100%	100%	0%	0%	100%	100%	0%
Drivers Pkg	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	100%
Navigation Center	66%	30%	0%	1%	46%	67%	50%	0%	0%	20%	33%	100%
Trim Level 1	92%	5%	9%	1%	91%	3%	3%	32%	100%	15%	67%	0%
Trim Level 2	0%	0%	91%	98%	0%	0%	97%	68%	0%	85%	0%	0%
Trim Level 3	8%	95%	0%	2%	9%	97%	0%	0%	0%	0%	33%	100%
Trim Level 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 1	0%	0%	91%	98%	0%	0%	97%	68%	0%	85%	0%	0%
Tires 2	56%	2%	0%	2%	38%	7%	0%	15%	100%	1%	0%	100%
Tires 3	44%	98%	9%	1%	62%	93%	3%	17%	0%	15%	100%	0%
Tires 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 14 (Continued)

	Vehicle Similarity Sets											
Options	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
FWD	23%	79%	2%	2%	53%	8%	7%	11%	90%	100%	0%	5%
SE	20%	79%	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%
SEL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%
LIMITED	73%	2%	100%	100%	94%	94%	99%	95%	10%	0%	0%	0%
SPORT	7%	19%	0%	0%	6%	6%	1%	5%	90%	0%	0%	0%
Charcoal Black Trim	57%	68%	59%	57%	63%	65%	53%	59%	100%	69%	62%	51%
Sync (Touch)	14%	50%	0%	0%	0%	0%	0%	0%	0%	77%	0%	100%
Floormats	55%	30%	47%	48%	47%	40%	59%	56%	50%	46%	32%	40%
Headlamps	73%	1%	100%	100%	63%	49%	96%	67%	0%	0%	0%	0%
Moonroof	79%	19%	100%	100%	79%	73%	98%	77%	80%	0%	0%	65%
Rearcamera	80%	21%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%
Roofrack	14%	43%	0%	0%	6%	19%	1%	10%	10%	48%	62%	13%
Satelliteradio	94%	72%	100%	100%	100%	100%	100%	100%	100%	77%	100%	100%
Speed Control	57%	1%	0%	0%	0%	0%	69%	47%	0%	0%	0%	0%
Trailertow Pkg	17%	0%	22%	20%	9%	21%	15%	18%	0%	0%	11%	6%
Leather	80%	21%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%
Premium Wheels	27%	98%	0%	0%	6%	6%	1%	5%	90%	100%	100%	42%
Ambient Pkg	80%	21%	100%	100%	100%	100%	100%	100%	100%	0%	0%	100%
Blis	80%	21%	0%	0%	0%	0%	100%	100%	100%	0%	0%	0%
Comfort Grp	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%
Drivers Pkg	80%	20%	100%	100%	72%	67%	99%	94%	90%	0%	0%	0%
Navigation Center	80%	21%	100%	100%	74%	63%	99%	88%	70%	0%	0%	52%
Trim Level 1	0%	0%	0%	0%	42%	0%	6%	0%	0%	0%	0%	0%
Trim Level 2	20%	79%	0%	0%	5%	0%	0%	0%	90%	0%	0%	0%
Trim Level 3	80%	20%	100%	35%	53%	87%	94%	58%	10%	100%	100%	100%
Trim Level 4	0%	1%	0%	65%	0%	13%	0%	43%	0%	0%	0%	0%
Tires 1	20%	79%	0%	0%	5%	0%	0%	0%	90%	0%	0%	0%
Tires 2	80%	18%	100%	35%	60%	46%	93%	29%	10%	100%	100%	100%
Tires 3	0%	2%	0%	0%	35%	41%	7%	28%	0%	0%	0%	0%
Tires 4	0%	1%	0%	65%	0%	13%	0%	43%	0%	0%	0%	0%

Table 14 (Continued)

	Vehicle Similarity Sets											
Options	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
FWD	1%	74%	0%	2%	0%	12%	1%	0%	0%	0%	57%	0%
SE	0%	68%	0%	0%	0%	0%	0%	0%	0%	0%	57%	0%
SEL	100%	32%	33%	100%	100%	100%	100%	100%	100%	100%	0%	0%
LIMITED	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
SPORT	0%	0%	67%	0%	0%	0%	0%	0%	0%	0%	43%	100%
Charcoal Black Trim	61%	47%	84%	48%	50%	46%	65%	33%	70%	61%	81%	100%
Sync (Touch)	91%	68%	0%	100%	86%	100%	87%	0%	100%	100%	0%	0%
Floormats	34%	18%	49%	71%	51%	38%	32%	67%	11%	25%	19%	35%
Headlamps	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Moonroof	62%	0%	61%	86%	73%	50%	36%	0%	0%	6%	43%	87%
Rearcamera	100%	0%	67%	100%	100%	100%	100%	100%	100%	100%	43%	100%
Roofrack	19%	76%	26%	11%	17%	18%	34%	100%	48%	58%	52%	0%
Satelliteradio	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	43%	100%
Speed Control	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trailertow Pkg	12%	12%	7%	7%	10%	14%	9%	0%	15%	22%	0%	0%
Leather	100%	0%	67%	100%	100%	0%	0%	100%	0%	0%	43%	100%
Premium Wheels	47%	100%	100%	67%	64%	88%	82%	67%	74%	70%	100%	100%
Ambient Pkg	100%	0%	67%	100%	100%	0%	0%	100%	0%	0%	43%	100%
Blis	0%	0%	67%	100%	100%	0%	0%	0%	0%	0%	0%	0%
Comfort Grp	100%	0%	0%	100%	100%	0%	0%	100%	0%	0%	0%	0%
Drivers Pkg	0%	0%	62%	0%	0%	0%	0%	0%	0%	0%	19%	35%
Navigation Center	28%	0%	63%	78%	57%	0%	0%	0%	0%	0%	24%	77%
Trim Level 1	0%	12%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trim Level 2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trim Level 3	100%	88%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Trim Level 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 2	98%	79%	95%	100%	100%	96%	97%	33%	85%	80%	95%	100%
Tires 3	2%	21%	5%	0%	0%	4%	3%	67%	15%	20%	5%	0%
Tires 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 14 (Continued)

	Vehicle Similarity Sets									
Options	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**
FWD	0%	0%	0%	0%	25%	7%	23%	0%	13%	50%
SE	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
SEL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
LIMITED	100%	0%	0%	3%	0%	98%	96%	100%	100%	100%
SPORT	0%	100%	100%	98%	100%	2%	4%	0%	0%	0%
Charcoal Black Trim	50%	100%	100%	99%	100%	58%	58%	0%	56%	83%
Sync (Touch)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Floormats	50%	27%	60%	59%	50%	46%	38%	0%	64%	33%
Headlamps	100%	0%	0%	0%	0%	79%	44%	0%	0%	0%
Moonroof	100%	88%	100%	95%	50%	90%	66%	100%	36%	0%
Rearcamera	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Roofrack	0%	0%	0%	0%	0%	4%	16%	0%	28%	100%
Satelliteradio	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Speed Control	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trailertow Pkg	50%	0%	0%	1%	0%	22%	14%	0%	8%	33%
Leather	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Premium Wheels	0%	100%	100%	98%	100%	2%	4%	0%	0%	0%
Ambient Pkg	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Blis	0%	0%	100%	100%	0%	0%	0%	100%	100%	100%
Comfort Grp	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Drivers Pkg	100%	49%	100%	90%	100%	81%	54%	100%	54%	100%
Navigation Center	100%	55%	100%	93%	100%	81%	53%	0%	79%	0%
Trim Level 1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trim Level 2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Trim Level 3	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Trim Level 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Tires 3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Tires 4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 15: The definitions of the most attractive configurations

	configurations										
options	84	85	216	220	224	227	275	312	385	403	416
drive	FWD	FWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD
body	SE	SE	LIM	LIM	LIM	LIM	SEL	SEL	LIM	LIM	LIM
charcoalblacktrim	1	1	0	0	1	1	1	1	0	0	1
sync (touch)	1	1	0	0	0	0	1	1	0	0	0
floormats	0	0	0	1	0	1	0	0	0	1	0
headlamps	0	0	1	1	1	1	0	0	1	1	1
moonroof	0	0	1	1	1	1	1	1	1	1	1
rearcamera	0	0	1	1	1	1	1	1	1	1	1
roofrack	0	1	0	0	0	0	0	0	0	0	0
satelliteradio	1	1	1	1	1	1	1	1	1	1	1
speedcontrol	0	0	0	0	0	0	0	0	1	1	1
trailertowpackage	0	0	0	0	0	0	0	0	0	0	0
leather seats	1	1	0	0	0	0	0	0	0	0	0
premium wheels	1	1	0	0	0	0	0	0	0	0	0
ambient pkg	0	0	1	1	1	1	1	1	1	1	1
blis	0	0	0	0	0	0	0	0	1	1	1
comfort grp	0	0	0	0	0	0	1	1	0	0	0
drivers pkg	0	0	1	1	1	1	0	0	1	1	1
navigation center	0	0	1	1	1	1	0	1	1	1	1
tires1	1	1	0	0	0	0	0	0	0	0	0
tires2	0	0	1	1	1	1	1	1	1	1	1
tires3	0	0	0	0	0	0	0	0	0	0	0

Table 15 (continued)

	configurations									
options	423	424	451	536	568	690	702	803	1033	1092
drive	AWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD	AWD
body	LIM	LIM	SEL	SPRT	SPRT	LIM	SEL	SEL	SEL	SEL
charcoalblacktrim	1	1	0	1	1	0	0	1	0	1
sync (touch)	0	0	1	0	0	0	1	1	1	1
floormats	1	1	1	0	1	0	1	1	0	1
headlamps	1	1	0	0	0	0	0	0	0	0
moonroof	1	1	0	1	1	1	1	1	0	1
rearcamera	1	1	1	1	1	1	1	1	1	1
roofrack	0	0	1	0	0	0	0	0	0	0
satelliteradio	1	1	1	1	1	1	1	1	1	1
speedcontrol	0	1	0	0	0	0	0	0	0	0
trailertowpackage	0	0	0	0	0	1	0	0	1	1
leather seats	0	0	0	0	0	0	0	0	0	0
premium wheels	0	0	0	1	1	0	1	1	1	0
ambient pkg	1	1	1	1	1	1	1	1	1	1
blis	1	1	1	1	1	0	1	1	0	1
comfort grp	0	0	1	0	0	0	1	1	1	1
drivers pkg	1	1	0	1	1	1	0	0	0	0
navigation center	1	1	0	0	1	0	1	1	0	1
tires1	0	0	0	0	0	0	0	0	0	0
tires2	1	1	1	1	1	1	1	1	0	1
tires3	0	0	0	0	0	0	0	0	1	0

Table 16: The description of the product variants based on subjective selection

Variant Code	Description	Frequency
1	FWD SE	428
2	FWD SEL	435
3	FWD Limited	185
4	FWD Sport	13
5	AWD SE	0
6	AWD SEL	1,971
7	AWD Limited	1,843
8	AWD Sport	333

REFERENCES

Abdelkafi, N. (2008). *Variety-induced complexity in mass customization: Concepts and management*. T. Blecker, G. Q. Huang, F. Salvador (Eds.). Berlin: Erich Schmidt Verlag GmbH & Co.

Adamowicz, W., Bunch, D., Cameron, T.A., Dellaert, B.G.C., Hanneman, M., Keane, M., Louviere, J., Meyer, R., Steenburgh, T., and Swait, J. (2008) Behavioral Frontiers in Choice Modeling, *Marketing Letters*, Vol.19, 215-228.

Albuquerque, P. and Bronnenberg, B.J. (2009) Measuring the Impact of Negative Demand Shocks on Car Dealer Networks, *Marketing Science*, 31(1), 4-23.

Anderson, S.P, Palma, A., and Thisse, J.F. (1992) *Discrete Choice Theory of Product Differentiation*, Cambridge, MA: The MIT Press.

Andrews, R.L. and Manrai, A.K. (1998) Feature-based Elimination: Model and Empirical Comparison, *European Journal of Operational Research*, Vol.111, 248-267.

Anupindi, R., Gupta, S., Venkataramanan, M.A. (2009) Managing Variety on the Retail Shelf: Using Household Scanner Panel Data to Rationalize Assortments, in *Retail Supply Chain Management (eds. N. Agrawal, S.A. Smith)*, Springer Science-Business Media.

Anupindi, R., Dada, M., and Gupta, S. (1998) Estimation of consumer demand with stpck-out based substitution: An application to vending machine products, *Marketing Science*, 17(4), 406-423.

Batchelor, J. (2001) Engineering a Vehicle for World Class Logistics: From Paradox to Paradigm Shifts on the Rover 75, in *Human Performance in Planning and*

Scheduling (B.L. MacCarthy and J.R. Wilson, eds.), Taylor and Francis, London.

Berger, I.E., Cunningham, P., and Drumwright, M.E. (2007). Mainstreaming Corporate Social Responsibility: Developing Markets for Virtue. *California Management Review*, Vol.49, 132-157.

Berry, S., Levinsohn, J., and Pakes, A. (1995) Automobile Prices in Market Equilibrium, *Econometrica*, 63(4), 841-890.

Bettman, J.R., Luce, M.F., Payne, J.W. (1998) Constructive Consumer Choice Processes, *The Journal of Consumer Research*, Vol.25, 187-217.

Blecker, T, Abdelkafi, N, Kaluza, B, Friedrich, G (2003). Variety Steering Concept for Mass Customization, Working Paper No. 2003/04, University of Klagenfurt.

Boatwright, P. and Nunes, J.C. (2001), Reducing Assortment: An Attribute-Based Approach, *Journal of Marketing*, 65 (3), 50-63.

Broniarczyk, S.M., Hoyer, W.D., and McAlister, L. (1998) Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction, *Journal of Marketing Research*, 35 (May), 166-176.

Brijs T., Swinnen G., Vanhoof K., and Wets G. (1999) The Use of Association Rules for Product Assortment Decisions: A Case Study, in *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, San Diego (USA), August 15-18, 254-260.

Bruno, H.A. and Vilcassim, N.J. (2008) Structural Demand Estimation with Varying Product Availability, *Marketing Science*, 27(6), 1126-1131.

Cachon, G.P. and Kök , A.G. (2007) Category Management and Coordination in Retail Assortment Planning in the Presence of Basket Shopping Consumers, *Management Science*, 53(6), 934–951

Chen, M.C. and Lin, C.P. (2007) A Data Mining Approach to Product Assortment and Shelf Space Allocation, *Expert Systems with Applications*, Vol.32, 976–986

Chintagunta, P.K. (1999) Variety Seeking, Purchase Timing, and the "Lightning Bolt" Brand Choice Model, *Management Science*, 45(4), 486-498.

Conlon, C.T. and Mortimer, J.H. (2010) *Demand Estimation Under Incomplete Product Availability*, Working Paper (May 2010), Department of Economics, Harvard University, Cambridge, MA.

Corrocher, N. and Guerzoni, M. (2009) Product Variety and Price Strategy in the Ski Manufacturing Industry, *Journal of Evolutionary Economics*, Vol.19, 471–486.

Crafton, S.M. and Hoffer, G.E., 1980. Do Consumers Benefit From Auto Manufacturer Rebates to Dealers? *Journal of Consumer Research*, Vol.7, 211-4.

DriveChicago.com, 2008 Volvo C30 Review, retrieved on 29th April 2011 from http://www.drivechicago.com/reviews/review.aspx?review=1453

Duquette, E.N. (2010) *Choice Diifculty and Risk Perceptions in Environmental Economics* (Doctoral Dissertation). Retrieved from ProQuest (3435765).

Fader, P.S. and Hardie, B.G.S. (1996) Modeling Consumer Choice Among SKUs, *Journal of Marketing Research*, 33 (November), 442-452.

Farias, V. F., Jagabathula, S. and Shah, D. (2011), A non-parametric approach to modeling choice with limited data, *Management Science* (to appear).

Fasolo, B., Hertwig, R., Huber, M., and Ludwig, M. (2009) Makes the Difference Between Small and Large Real-World Assortments? *Psychology & Marketing*, 26(3), 254–279.

Feenstra, R. and Levinsohn, J. (1989) Distance, Demand, and Oligopoly Pricing, NBER Working Paper #3076, Cambridge, MA.

Ferguson, S. and Donndelinger, F. (2010) A Multidisciplinary Approach to Market Segmentation and Product Family Definition, in *13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference 13-15 September, Fort Worth, Texas*.

Fisher, M.L., Jain, A., and MacDuffie, P. (1996) *Strategies for Product Variety: Lessons from the Auto Industry, Designing the Firm* (eds. E. Bowman, B. Kogut), New York: Oxford.

Fisher, M.L. (1997) What is the Right Supply Chain for Your Product?" *Harvard Business Review*, 75(2), 105-116.

Fisher, M. and Ittner, C. (1999) The Impact of Product Variety on Automobile Assembly Operations: Empirical Evidence and Simulation Analysis. *Management Science*, Vol.45, 771–786.

Fisher, M., Ramdas, K., and Ulrich, K. (1999) Component Sharing in the Management of Product Variety: A Study of Automotive Braking Systems, *Marketing Science*, 45(3), 297-315.

Gaur, V. and Honhon, D. (2006) Assortment Planning and Inventory Decisions under a Locational Choice Model, *Management Science*, 52(10), 1528–1543

Gensch, D.H., 1987. A two-stage disaggregate attribute choice model. *Marketing Science*, Vol.6, 223-231.

Gillbride, T.J. and Allenby, G.M. (2004) A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules, *Marketing Science*, 23(3), 391-406

Hoch, S.J., Bradlow, E.T., and Wansink, B. (1999) The Variety of an Assortment, *Marketing Science*, 18(4), 527-546.

Hogarth, R. M. (1987) *Judgement and Choice* (2^{nd} ed.), New York, NY: John Wiley & Sons.

Hruschka, H. (2007), Using a Heterogeneous Multinomial Probit Model with a Neural Net Extension to Model Brand Choice, *Journal of Forecasting*, 26, 113–127.

Iyengar, S.S. and Lepper, M.R. (2000) When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6), 995-1006.

Jenkins, S.P. (2008) *Survival Analysis*. Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester.

Kahn, B.E. and Wansink, B. (2004) The Influence of Assortment Structure on Perceived Variety and Consumption Quantities, *Journal of Consumer Research*, 30 (March), 519-533.

Kessels, R., Goos, P., and Vandebroek, M. (2006) A Comparison of Criteria to Design Efficient Choice Experiments, *Journal of Marketing Research*, 43(3), 409-419.

Kohlberger, R. and Gerschberger, M. (2012) External Product Variety in Value Networks – A Case Study From an Austrian Premium Producer, in *Proceedings of the International Conference on Industrial Logistics (ICIL)*, Zadar (Croatia), June 14-16, 70-77.

Kök, A.G., Fisher, M.L., and Vaidyanathan, R. (2008) *Retail Supply Chain Management*, Eds. N. Agrawal and S. A. Smith, Kluwer Publishers

Kök, A.G. and Fisher, M.L. (2007) Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application, *Operations Research*, 55(6), 1001-1021

Kök, A.G. and Fisher, M.L. (2006) Assortment Planning: Review of Literature and Industry Practice, Working Paper (November 2006)

Kusiak, A., Smith, M.R., and Song, Z. (2007) Planning Product Configurations Based on Sales Data, IEEE *Transactions on Systems, Man, and Cybernetics—Part C: Applications And Reviews*, 37(4), 602-609.

Lancaster, K.J. (1971), *Consumer Demand: A New Approach*. New York: Columbia University Press.

Lancaster, K.J. (1990), The Economics of Product Variety: A Survey, *Marketing Science*, Vol.9, 189-206.

Lillien, G. L., Rangaswamy, A., and Bruyn, A. D. (2007) *Principles of Marketing Engineering*, State College, PA: DecisionPro Inc.

Manrai, A.K. and Andrews, R.L. (1998) Two-Stage Discrete Choice Models for Scanner Panel Data: An Assessment of Process and Assumptions, *European Journal of Operational Research*, Vol.111, 193-215

Markus, H.R. and Schwartz, B. (2010) Does Choice Mean Freedom and Well-Being? *Journal of Consumer Research*, 37(2), 344-355

McFadden, D. (1986) The Choice Theory of Market Research, *Marketing Science*, Vol.5, 275-297.

Media Volvo, retrieved on April 29th 2011 from https://www.media.volvocars.com/global/enhanced/en-gb/Media/Preview.aspx?mediaid=18624

Mintel US Market Report: SUVs and CUVs, April 2012.

Mittal, S. and Frayman, F. 'Towards a generic model of configuration tasks', in: Proc. of the 11th International Joint Conference on Artificial Intelligence (IJCAI), 1395–1401, 1989.

Osnos, E. (1997, September 27) Too many choices? Firms cut back on new products, *Philadelphia Inquirer*, D1-D7.

Pil, F.K. and Holweg, M. (2004) Linking Product Variety to Order-Fulfillment Strategies, *Interfaces*, 34(5), 394-403.

Puskorius, G., Goodman, B.R., Lenardon, M., Li, H., and Nance, B. (2012) *U.S. Patent No.8,214,313 B1*. Dearborn, MI: U.S.

Ramdas, K. (2003) Managing product variety: An integrative review and research directions, *Production and Operations Management*, Vol.12, 79–101.

Rathnow, Peter J. (1993) Integriertes Variantenmanagement: Bestimmung, Realisierung und Sicherung der optimalen Produktvielfalt, Goettingen.

Salvador, F. and Forza, C. (2007) Principles for Efficient and Effective Sales Configuration Design, *International Journal of Mass Customisation*, 2(1), 114-127.

Sampaio, M. and Machline, C. (2009) Consumer Response to Stockouts, POMS 20th Annual Conference, Orlando, Florida, U.S.A., May 1-4.

Sanchez, R. (1999) Modular Architectures in the Marketing Process, *The Journal of Marketing*, Vol. 63, 92-111.

Sandeep R. Chandukala, Jaehwan Kim, Thomas Otter, Peter E. Rossi, and Greg M. Allenby (2008) *Choice Models in Marketing: Economic Assumptions, Challenges and Trends, Foundations and Trends in Marketing*, Vol. 2, No. 2 (2007) 97–184

Scheibehenne, B., Greifeneder, R., and Todd, P.M. (2010) Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload, *Journal of Consumer Research*, 37(3), 409-425.

Schwartz, B. (2004) *The Paradox of Choice*: *Why More is Less,* HarberCollins E-books.

Sema e-News (2009), SUV and CUV Sales on the Rise, http://www.sema.org/node/5299

Shannon, C.E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, Vol.27, 379-423.

Smith, S.A. and Agrawal, N. (2000) Management of multi-item retail inventory systems with demand substitution, *Operations Research*, Vol.48, 50–64.

Staeblein, T., Aoki, K., and Tomino, T. External product variety in the automotive industry: An empirical analysis of German and Japanese OEM operations, in Proceedings of Gerpisa Coloquium, Paris, 2011.

Stefanescu, C. (2009) Multivariate customer demand: modeling and estimation from censored sales, Working Paper, European School of Management and Technology, Berlin, Germany, 2009, http://ssrn.com/abstract=1334353.

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

Swait, J. and Adamowicz, W. (2001) The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching, *Journal of Consumer Research*, 28(1), 135-148.

Tabachnick, B.G. and Fidell, L.S. (2007) *Using Multivariate Statistics* (5th ed.), USA: Pearson.

Thompson, D.V. and Hamilton, R.W. (2006) The Effects of Information Processing Mode on Consumers' Responses to Comparative Advertising, *Journal of Consumer Research*, Vol.32, pp.530-540.

Train, K.E. (2009) *Discrete Choice Methods with Simulation*, 2nd edition, Cambridge University Pres.

van Herpen, E. and Pieters, R. (2002) Research Note: The Variety of an Assortment: An Extension to the Attribute-Based Approach, *Marketing Science*, Vol.21, 331-341.

Vermeulen, B., Goos, P., Scarpa, R., and Vandebroek, M. (2011) Bayesian Conjoint Choice Designs for Measuring Willingness to Pay, *Environmental and Resource Economics*, Vol.48, 129–149.

Wilson, A. (February 18, 2008) Ford Cuts Equipment Combinations on '09s, *Automotive News*.

Wong, R.C.W. and Fu, A.W.C. (2005) Data Mining for Inventory Item Selection with Cross-Selling Considerations, *Data Mining and Knowledge Discovery*, Vol.11, 81–112.

Wu, L. (2007) Toward an Understanding of the Product Variety Decision: An Integrative Model, International Joumal of Management, 24(3), 499-504.

Yunes, T.H., Napolitano, D., Scheller-Wolf, A., and Tayur, S. (2007) Building Efficient Product Portfolios at John Deere and Company, *Operations Research*, 55(4), 615-629.

Yücel, E., Karaesmen, F., Salman, F.S., and Türkay, M. (2009) Optimizing product assortment under customer-driven demand substitution, *European Journal of Operational Research*, 199(3), 759-768.

ABSTRACT

DECISION SUPPORT MODELS FOR MANAGING EXTERNAL VARIETY OF CONFIGURABLE PRODUCTS

by

ERKAN ISIKLI

December 2012

Advisor: Dr. Ratna Babu Chinnam

Major: Industrial and Systems Engineering

Degree: Doctor of Philosophy

Marketers, retailers, and manufacturers have all been trying to know how much product variety is sufficient to satisfy consumer needs while keeping stocking and production costs at a reasonable level. Even though all manufacturing firms face competing objectives in determining the product variants to be built, this problem is especially troublesome for manufacturers of configurable products (e.g., automobiles) that sell a complex product with many options and features resulting in a very large buildable configuration space.

This study proposes a framework for analyzing external product variety in case of configurable products and provides a methodology to reduce the number of product variants that should be considered before building operational models to estimate demand or optimize assortments in case of configurable products. The main goal is reduction of the abundance of configurations offered by manufacturers as analytically as possible within a decision support framework rather than product assortment planning. The research proposes a number of methods and techniques to extract meaningful and actionable information on external variety when the data

sources are limited. The proposed methods are validated using multiple datasets from a large North American automotive original equipment manufacturer covering vehicles in multiple segments and U.S markets.

AUTOBIOGRAPHICAL STATEMENT

Erkan Isikli was born in Istanbul, Turkey on July 13, 1982, as the son of Nizami and Tulay Isikli. He received his bachelor’s degree in mathematical engineering from Istanbul Technical University (ITU) in 2004. Then, he started studying for his MBA at the same university, and in December 2005 he was accepted to the Industrial Engineering department of ITU as a research assistant. He received his master’s degree in January 2007, and soon after that he was admitted to the Ph.D. program in the Industrial & Systems Engineering Department at Wayne State University.

He decided to pursue his doctoral education at the Wayne State University, Detroit, Michigan, USA as a recipient of the fellowship granted by the Higher Education Council of Turkey in September 2007. In November 2012, he finished his study at this program with great encouragement and help of his colleagues and instructors. Upon graduation, he plans to work as an Assistant Professor at the Industrial Engineering Department of ITU.

Isikli has published and presented many proceedings at several conferences in different countries and is focused mainly on publishing quality journal papers related to his research of interest. He wants to specialize in constructing statistical models for real-world problems in various fields such as non-profit organizations, sports, environment, marketing, supply chain, and healthcare.