

Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration

Zhenghui Sha
University of Arkansas

Veronica Saeger
Los Alamos National Laboratory

Mingxian Wang and Yan Fu
Ford Motor Company

Wei Chen
Northwestern University

ABSTRACT

For achieving viable mass customization of products, product configuration is often performed that requires deep understanding on the impact of product features and feature combinations on customers' purchasing behaviors. Existing literature has been traditionally focused on analyzing the impact of common customer demographics and engineering attributes with discrete choice modeling approaches. This paper aims to expand discrete choice modeling through the incorporation of optional product features, such as customers' positive or negative comments and their satisfaction ratings of their purchased products, beyond those commonly used attributes. The paper utilizes vehicle as an example to highlight the range of optional features currently underutilized in existing models. First, data analysis techniques are used to identify areas of particular consumer interest in regards to vehicle selection. Consumer responses related to the reasons for/against the purchase of a specific vehicle model form the basis of extracting these areas of interest. Second, satisfaction rating variables are leveraged to characterize and quantify those identified areas of interest. Several methods of handling missing values of rating variables are developed to impute the data sets in supporting the choice modeling. Finally, multinomial logit models are constructed using engineering attributes, customer demographics as well as the rating variables to quantitatively assess customer preferences, evaluate the feasibility of different imputation methods, and predict customers' choices. The results show that the inclusion of satisfaction ratings improves model fit and predictive capabilities.

CITATION: Sha, Z., Saeger, V., Wang, M., Fu, Y. et al., "Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration," *SAE Int. J. Mater. Manf.* 10(3):2017, doi:10.4271/2017-01-0243.

1. INTRODUCTION

Modern market consists of full of customers demanding variety and customization in products, whereas, producing products with large variety is always hampered by the costs involved. Mass customization is one way of approaching this issue. Salvendy [1] defines mass customization as "*producing goods and services to meet individual customer's needs with near mass production efficiency*". Mass customization combines the advantages of both customized products and mass production cost reductions, therefore, benefits both manufacturers and customers.

One effective way of realizing mass customization is to start with a common platform, and then adding additional features on top of such a platform. This describes the fundamental idea in the *product family design* [2], which manifests through the processes of creating product variants based on a common product platform architecture but with

different optional features, which are open to customers to choose. Vehicle is a good example, with optional features, such as GPS, rear camera, sun roof, leather seats, etc. in a car model defining various trim levels so that customers can choose according to their own needs and preferences.

Therefore, one of the most important problems in the product family design is to determine what optional features shall be added. Wielinga and Schreiber [3] define such a problem as *configuration design* - the search for an assembly of components from a predefined set that satisfies a set of requirements and obeys a set of constraints. Among different engineering/non-engineering requirements and constraints, customer demand is at the top of the six success factors for mass customization [4] and therefore, the key to a successful product configuration. That means an enterprise wishes to know a specific configuration would favor customer's demand in a market before the

design is adopted for final production. To this end, one area of investigation is to analyze customers' preferences to different product optional features. Quantitatively understanding such preferences would allow enterprises to assess the impact of features and feature combinations on the purchasing behaviors of customers. That information would be also helpful to enterprises that are attempting to optimize their configurations, as they would have a better understanding on the features that customers are truly looking for. It is worth noting that our work does not aim to fundamentally address the question of what is the best combination of features offered to one product. The analysis performed in this study still assumes fixed features, not configurable. Yet our focus is on analyzing customers' preferences to these features so that designers can assess whether a feature is important enough to be added while doing product configuration.

Uncovering consumer preference is a well-defined problem in market science, and has been paid much attention in engineering design literature as well because of the strength that the Decision-Based Design (DBD) methodology [5] has provided beyond traditional design and optimization. Among various customer preference analysis approaches, discrete choice analysis (DCA) is a widely utilized tool for quantitative assessment. DCA uses customer attributes (i.e. age, gender, income) and product attributes (i.e. power and gas mileage for vehicles) to predict customer purchase behavior in terms of a choice probability. Existing research in design community encompasses analyzing preference inconsistencies in customer choice of "green" products [6], preferences for aesthetic forms [7], the use of latent class analysis for modeling heterogeneous customer preferences [8], and many others. A variety of discrete choice models (DCM), including multinomial logit [9, 10], probit [11], nested logit [12], and mixed logit models [13, 14], have been introduced in design literature in order to involve demand forecast in design decisions.

In support of the product family design, Li and Azarm [15], and Michalek et al. [16, 17] have integrated demand models in the design of product families. But, these approaches do not systematically determine the number of products in the optimization formulation and simply ignore the commonality among product variances. In addressing these limitations, Kumar et al. [18] present a market-driven product family design (MPFD) framework. The MPFD leverages nested logit model and helps to make decisions on product line positioning, commonality, and optimal configuration of design variables for each product in the family based on engineering and manufacturing feasibility and a demand model that predicts market performance. Despite a wide variety of application of DCM in the automotive design context, other practical applications to make an optimal design and system engineering, which maximizes utilities of decision-makers, have been demonstrated for dial-readout bathroom scales [17], and large-scale complex systems such as Internet [19] and Air transportation systems [20].

In design literature, existing DCA typically model how customers make decisions among products which have distinct key features, but rarely analyze customers' preferences to products' optional features. Moreover, existing studies often use alternative-related product

attributes as input variables because these variables can be obtained from product catalog and can be acquired before the choice decisions are made. There are few studies that leverage customers' reviews and subjective feelings, such as satisfaction ratings and evaluations, which are often acquired after the decisions are made, to support the choice modeling. He et al. [21] incorporated the rating variable into the DCA, but the method of handling missing values in rating variables may double count the influence of customers' demographics in the DCA, thereby introduce correlations and may induce inaccurate estimation results. In addressing these issues, it demands new means of handling the missing rating problem and investigate how the additional of rating variables can help improve the DCA for gaining more insights on customers' preferences to product optional features.

In this paper, we aim to address two research questions (RQ). **RQ1**: how can customers' responses and comments to product optional features be used to *qualitatively* reveal customers' preferences and help identify the key inputs of DCA. **RQ2**: how can customers' satisfaction ratings to the optional features be incorporated into DCA to *quantitatively* assess customer' preferences in supporting product configuration. Answering these two questions is challenging due to the fact that, first, data about revealed preferences to the optional features is difficult to obtain. Such data limitation is also the reason why few studies are carried out so far. Second, the survey data covering customers' satisfaction ratings and responses to optional features is often large and messy, that needs pre-analysis to reduce the dimension of independent variables dataset. Finally, the missing value issue (i.e., customers are only asked to provide ratings to their purchased product but not to all their considered alternatives) and ownership bias issue (i.e., customers rate their purchased products higher due to a sense of ownership or investment in the product) in customer's satisfaction ratings are additional challenges hindering the application of rating variables in DCA. More detailed discussion on this topic will be elaborated in [Section 4.3](#).

In this regard, this paper presents a data-driven approach to estimate customers' preferences to support product configuration. The dataset used in this study covers fine-grained information of nearly 50,000 individuals' responses to their purchased products. Especially, the customers' responses of positive points and negative points to product are utilized to help identify the key input variables for DCA. In processing such a large volume of data, we integrate our previous study that uses joint correspondence analysis [22] and network analysis method [23] to reduce the data dimensions and cluster the vehicles in different communities such that our scope can be focused on a specific market segment (see [Section 2.2](#) for detailed description about the selected communities). We also develop methods based on parametric distributions to address the missing values issue and ownership bias issue simultaneously in satisfaction ratings to enhance the practical application of DCA in engineering design. Our study is conducted in a vehicle design context, and a case study is performed based on the data of 2013 China auto market. However, these methods can be generally applied in different design contexts in addition to the presented vehicle design case study.

The paper is structured as follows. In Section 2, the data source is introduced and the data selected for analysis in this study is presented. In Section 3, the descriptive analysis on customers' response to their purchased products is presented, from which we identify the key variables that shall be included in the DCA. In Section 4, we present the developed imputation methods of addressing the issue of missing satisfaction ratings in DCA. The models based on different imputation methods are compared and evaluated in this section too. In Section 5, the closing thoughts are presented and future work is outlined.

2. DATA SET AND SELECTION OF DATA FOR ANALYSIS

2.1. Data source

The data set used in this study is the New Car Buyers Survey (NCBS) that was conducted by an independent market research company. In order to guarantee that respondents had enough driving experience in the new car to accurately assess its qualities, while still being considered new car buyers, the respondents must have owned the car for between 6 and 32 weeks. The data consists of 49,921 new car buyers' responses over 389 unique vehicle models of the year 2013 in China's auto market. The survey had questions covering a variety of topics, including respondent demographics (e.g., the household income, age), vehicle information (e.g., the make and model, price, engine size, fuel consumptions), and customer perceived vehicle characteristics (e.g., youthful, sophisticated, business oriented). The respondents were also asked to list the car they purchased, the main alternative car they considered, and other cars they had considered before making the purchase decision.

2.2. Selection of Data for Analysis

Totally, there are 871 variables in the data set. This study chooses the list of optional features, the positive and negative points of the car purchased, the reasons for purchasing the car, and the reasons for not purchasing the main alternative car. Table 1 shows the checklist of 24 optional features that the respondents need to check if they have the features in their cars.

After identifying the variables to be explored, the next consideration is what cars to analyze. In our previous study, Wang et al. [22] performed the product community analysis and grouped the 389 car models into seven communities. In this study, three (*communities* 3, 6 and 7) out of the seven communities are chosen. The reasons for choosing these communities are two-fold. First, among the three communities, *community* 3 contains the average mid-sized sedans, *community* 6 is luxury mid-size sedans, and *community* 7 is comprised of minivans. These communities are different enough from one to another to provide some differentiation in the variables. The second reason for the communities chosen is size. Two of the communities have around 20 car models and the other has around 30. Compared to other communities, for example the *community* 2 which has around 175 car models, these three communities are much more manageable.

Table 1. The checklist of vehicles' optional features

Index	Features
1	DVD Player With Screens In The Back
2	Park Assistance System
3	Lane Assistance System
4	MP3/ 4/ iPod/ USB Socket
5	Air Conditioning
6	Sunroof/ Full Glass Roof/ Panoramic Glass Roof
7	Cruise Control
8	Bluetooth
9	Navigation System - Fixed
10	Navigation System - Portable
11	Tow Bar
12	ESP
13	Equipment For Fixing/ Loading Items Carried
14	Reduction Gear
15	Central Door Locking With Remote Control
16	Car Alarm
17	Stability Control System
18	Airbag: Passengers Side
19	Side Airbags
20	Rear Air Bags
21	Leather Seats/ Trim
22	Alloy Wheels
23	Anti-Lock Braking System (ABS)
24	Electric Front Windows

Table 2. Represented car models in communities being analyzed

Communities	Representative Models	Group Co-considered Rankings	Group Purchasing Rankings
<i>Community</i> 3: Common Mid-Size Sedans	VW Magotan	1	2
	Chevrolet Malibu	2	5
	Honda Accord	3	10
	VW Sagitar	4	1
	Toyota Camry	5	6
<i>Community</i> 6: Luxury Mid-Size Sedans	BMW 5 Series	1	2
	Mercedes E Class	2	3
	Audi A6L 2012	4	1
	Toyota Crown	5	7
	Lexus ES	7	6
<i>Community</i> 7: Minivans	Honda Odyssey	1	9
	Buick GL8	2	1
	VW Touran	3	2
	Mazda 8	4	5
	Mercedes Viano	7	5

For analyzing the variables, both the group averages and a representative sample of the cars within each group are used. Table 2 shows the five representative cars of each community. To determine representative cars, the models within each community are ordered by number of models purchased. Then a list is created that ranks the models by the frequency that they are co-considered within the community. To clarify, only the cars in the community are considered for the co-consideration ranking. The frequency does not include any time that the model is the main alternative but the customer buys a car from another community. Representative cars are finally chosen for a combination of three filters: how high they are ranked in the purchasing order, how high they are ranked in the co-considered order, and how recognizable they are. The last filter is considered in order to make the groups more accessible, as there are some models

like the Honda Accord or Toyota Camry that people are very familiar with. It is important to include both most frequently purchased and most frequently considered but not purchased in order to identify why certain cars are repeatedly considered but not purchased.

3. DESCRIPTIVE ANALYSIS OF CUSTOMER RESPONSES TO OPTIONAL PRODUCT FEATURES

Customer's response variables reveal what features the customers are most interested. In this section, we present the descriptive analysis of customers' responses to vehicles' optional features, as shown in [Table 1](#). Different from most choice models which typically use demographic information and engineering attributes, the work presented in this section is focused on the identification of variables outside those categories but could impact customers' decisions. The idea is to apply descriptive statistics, such as averages and percentages, to multiple variables in order to gain insights into customers' purchasing behaviors.

The analysis is decomposed to focus on several segments of the data set. This enables more in-depth investigation due to a decrease in data size. Specifically, prevalence of features is determined, as well as availability. Customer praised and criticized attributes are analyzed along with purchasing reasons to highlight overlapping or conflicting attributes, and to pinpoint the most relevant features in supporting the DCA. In the following subsections, we present the details of the variable structure and the analysis of these response variables.

3.1. Preferences Analysis with Customers' Responses

3.1.1. Response Data Structure

There are three different sets of responses, viz. *actual response*, *detailed categories* and *broad groups*. For each set, there are four different types of comments. The first one indicates the positive points that customers provide to describe their newly purchased vehicles, the second one indicates the negative points, the third one lists their reasons for purchasing the vehicles they selected, and the fourth one lists the reasons they did not purchase their main alternative vehicle.

Table 3. Examples of different types of comments

Type of comments	Examples of Comments
Positive points	• "Road holding, stability in cross winds, cornering, handling" • "General quality of the interior, including fit and finish, interior fittings, dials, instruments"
Negative points	• "Not roomy, not spacious, poor legroom, poor headroom" • "Poor acceleration, unresponsive engine, slow pick-up"
Reasons for buying	• "Value for money." • "Size of boot, good luggage space, payload capacity (height/ weight/ volume)."
Reasons for not buying	• "No discount offered, lower discount offered compared with new car." • "Too large, too big for requirements."

The *actual response* contains what the respondent actually said when asked to provide the comments to their vehicles. Examples comments of positive points and negative points are shown in [Table 3](#). Then the responses are grouped into detailed categories, which are then grouped into 11 broad groups by the survey company. These categories are Comfort, Design (Ergonomics), Driving Characteristics (DC), Others, Price, Quality Related Aspects (QRA), Safety, Size, Specification/ Accessories (SA), Technical Aspects (TA), and Dealer (Initial Purchase and After Sales Service). [Table 4](#) shows an example of the detailed categories of comments in the broad group. The column is checked if the detailed category is present for a specific type of comments, i.e., the positive and negative points, the reasons for buying and not buying the cars. For a list of the detailed categories, please refer to [\[24\]](#).

Each respondent can provide up to ten comments for each type of the four comments. For example, if the respondent only gave three comments the other seven cells for the comments in the data would simply be "NaN". The decision is made to treat all responses as equal. This means that no differentiation is made whether the comments were the first one given or the tenth one given. We also decide to merely total the number of comments and calculate the percentages from there due to the number of responses given. For example, for *community 3* there are 6953 respondents, but only one of those gives one comment. Half of the respondents give at least three comments and then it drops off steeply, with only 2% of respondents giving at least six responses and only two respondents giving 10. In the [Section 3.1.2](#) and [Section 3.1.3](#), we performed the descriptive analysis on the positive points and negative points of the customers' comments.

Table 4. Detailed categories of comments in the price group

Detailed Category	Positive Points	Negative Points	Reasons for Buying	Reasons for Not Buying
Discount on New Car			✓	✓
Economy in General	✓		✓	✓
Fuel Consumption	✓	✓	✓	✓
Maintenance Cost	✓	✓	✓	✓
Price	✓	✓	✓	✓
Price of Options		✓		
Resale Value	✓		✓	✓
Tax Cost		✓		
Terms of Payment			✓	✓
Trade in Price			✓	
Value for Money	✓	✓	✓	✓

3.1.2. Positive Points of New Car

As the examples shown in [Table 3](#), the positive points describe the advantages of the vehicles that customers purchased. We calculate the percentage of positive points in each broad group for communities 6 and 7. It is found that the largest groups are Design and Driving