Proceedings of DETC'03
ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conference
Chicago, Illinois USA, September 2-6, 2003

DETC2003/DTM-48683

ENHANCING DISCRETE CHOICE DEMAND MODELING FOR DECISION-BASED DESIGN

Henk Jan Wassenaar
Integrated Design Automation Laboratory
Department of Mechanical Engineering
Northwestern University
Evanston, Illinois 60208-3111
wassenaer@yahoo.com

Wei Chen*
Integrated Design Automation Laboratory
Department of Mechanical Engineering
Northwestern University
Evanston, Illinois 60208-3111
weichen@northwestern.edu

Jie Cheng
Power Information Network
J.D. Power & Associates
Troy, Michigan 48098

Agus Sudjianto
Engine Engineering Analytical Powertrain
Ford Motor Company
Dearborn, Michigan 48121

ABSTRACT

Our research is motivated by the need for developing a rigorous Decision-Based Design framework and the need for developing an approach to demand modeling that is critical for assessing the profit a product can bring. Even though demand modeling techniques exist in market research, little work exists on product demand modeling that addresses the specific needs of engineering design in particular that facilitates engineering decision-making. Building upon our earlier work on using the discrete choice analysis approach to demand modeling, in this work, we provide detailed guidelines for implementing the discrete choice demand modeling approach in product design. The modeling of a hierarchy of product attributes is introduced to cascade customer desires to specific key customer attributes that can be represented using engineering language. To improve the predictive capability of demand models, we propose to use the Kano method for providing the econometric justification when selecting the shape of the customer utility function. A real (passenger) vehicle engine case study, developed in collaboration with the market research firm J.D. Power & Associates and Ford Motor Company, demonstrates the proposed approaches. The example focuses on demand analysis and does not reach beyond the key customer attribute level. The obtained demand model is shown to be satisfactory through cross validation.

Key words: decision-based design, demand modeling, discrete choice analysis, vehicle design, engine design, Kano method, hierarchy of attributes, customer utility

* Corresponding author

1 INTRODUCTION

Our research is motivated by the need for developing a rigorous engineering design framework (Hazelrigg, 1999) and the need for developing an approach to demand modeling that is critical for assessing the profit[1] a product can bring. Many existing engineering design methods, such as Taguchi's robust design (Phadke, 1989) and Design for Six-Sigma (Fowlkes and Creveling, 1995), constitute preference systems in which it is assumed that meeting customer satisfaction is the primary goal of design decision making. Such methods seldom consider the cost associated with adding or improving a quality feature. Decision-based design (DBD) is emerging as a new approach to engineering design that recognizes the substantial role that decisions play in design and in other engineering activities. The approach is developed to perform in design environments characterized by ambiguity, uncertainty and risk. It seeks to maximize the value of a designed artifact while considering the interests of both the producer and the end-users (Hazelrigg, 1999; Wassenaar and Chen, 2001).

Although there is great consensus that for a profit-driven company, the *value* of a product should be a measure of the *profit* it brings, there exist concerns on using profit as the single criterion because of the belief that profit seems too difficult to model. O*ne difficulty in modeling the profit is the construction of a reliable product demand model* that is critical for assessing the revenue, the total product cost, and eventually the profit.

[1] Profit is a result of accounting practices, which need not be related to engineering design such as depreciation. Therefore, with profit is meant net revenue, i.e., the difference between revenue and expenditure. The net revenue can be discounted to present value.

Even though demand modeling techniques exist in market research, little work addresses the specific needs of engineering design, in particular that facilitates engineering decision-making.

In market research, two major demand analysis techniques, Discrete Choice Analysis (DCA) and Conjoint Analysis are being used to capture customer choice behavior. *DCA is based on probabilistic choice models*, which were originated in mathematical psychology (see Thurstone 1927; Luce, 1959; Marshak, 1960; Tversky, 1972) and developed in parallel by economists and cognitive psychologists. DCA identifies patterns in choices customers make between competing products and generates *the probability* that an option is chosen. Conjoint Analysis (CA) (Green and Wind, 1975; Green and Srinivasan, 1978, 1990), is an approach that determines the preference structure (a multiattribute utility function) of (classes of) consumers, which can be used to predict the demand as function of the product features (attributes). Though conjoint analysis remains to be a popular approach, DCA is seen by some as an evolutionary improvement (Riedesel, http://www.action-research.com/compare.htm; Louviere, 2000).

Demand analysis in market research is limited, focused on feature upgrades, packaging, product placement, etc., and regularly not intended to guide engineering design decisions. Applications of DCA for demand estimation has been seen in transportation engineering (Ben-Akiva and Lerman, 1985), however, very little research on demand analysis exists in the field of engineering design. Li and Azarm (2000) presented an approach for estimating the market demand by comparing multiattribute utility values obtained by conjoint analysis. Cook proposed the S-model approach (Cook, 1997; Donndelinger and Cook, 1997), which is a Taylor expansion about a reference point where the value and price of all products are identical and equally accessible to all customers resulting in equal demand for each product. These two existing demand modeling approaches share several limitations. One major limitation is that uncertainty is not incorporated in customer utility and demand estimations. Predicting a customer's choice requires a more sophisticated approach such as *probabilistic choice theory* (Luce, 1959).

In our previous work (Wassenaar and Chen, 2001), we enhanced the DBD framework originally proposed by Hazelrigg (1998) and proposed a discrete choice analysis based approach to demand modeling. We found that the disaggregate demand models built from DCA use data of individuals instead of group averages, which enables the variation of characteristics of individuals to be captured more accurately and avoids paradox associated with group decision-making.

Even though we have demonstrated the usefulness of DCA for demand analysis in engineering design (Wassenaar and Chen, 2001), there are still many important research issues to be addressed before the method can be fully applied and used with confidence. First, more guidelines need to be provided to assist designers when applying the DCA approach to demand modeling that facilitates engineering decision-making. The modeling of a hierarchy of product attributes needs to be addressed and the procedure should enable designers to focus the demand survey on specific features of the product without harming the consistency of the demand analysis at the *system level* of a product. Secondly, a method for selecting the form of the customer utility function should be provided so that the construction of a demand model is based on sound econometric reasoning instead of simple (mathematical) model fitting. A mathematical model may fit the data well however may not capture the true underlying purchase behavior and therefore its value in predicting demand may be poor.

The outline of this paper is as follows: In Section 2, we provide the necessary background of Discrete Choice Analysis and the Kano method, which we propose to use for providing the econometric justification when selecting the shape of the customer utility function. In Section 3 of this paper we provide detailed guidelines for demand modeling that facilitates engineering decision-making by modeling a hierarchy of product attributes and cascading customer desires to specific key customer attributes that can be represented using engineering languages. We also present an approach for selecting the form of the customer utility function of the demand analysis to enhance the predictive accuracy. The proposed approaches are demonstrated in Section 4 using a real (passenger) vehicle engine design case study in collaboration with the market research firm J.D. Power & Associates and Ford Motor Company. However, in the case study, we focus on demand modeling rather than decision-based design optimization. The concluding remarks and direction for future research are presented in Section 5.

2 TECHNICAL BACKGROUND

2.1 The Decision-Based Design Framework

The flowchart of the DBD framework that links engineering design with business decision making and that we proposed as an enhancement to the framework proposed by Hazelrigg (1999) in our previous work (Wassenaar and Chen, 2001) is shown in Fig. 1. One major contribution lies in introducing Discrete Choice Analysis as a systematic approach to establish the relationship between the key customer attributes $\boldsymbol{A}$, the socioeconomic and demographic background $\boldsymbol{S}$ of the market population, time t, and the demand Q. We discern two different types of attributes in our approach, namely the engineering attributes $\boldsymbol{E}$ and the key customer attributes $\boldsymbol{A}$. The engineering attributes $\boldsymbol{E}$ are product properties that are of interest to a design engineer, represented as functions of design options $\boldsymbol{X}$ through engineering analysis. The key customer attributes $\boldsymbol{A}$ are product features (next to brand, price, and warranty) that a customer typically considers when purchasing the product. In a demand model $\boldsymbol{A}$ needs to be represented using engineering languages (see more details in Section 3.1). The arrows in the flowchart indicate the existence of relationships between the different entities (parameters) in DBD.

2.2 Background of Discrete Choice Analysis

DCA identifies patterns in choices customers make between competing products and generates *the probability* that an option is chosen. A key concept of DCA is the use of random utility (probabilistic choice theory) to address unobserved taste variations, unobserved attributes, and model deficiencies. A quantitative process based on *multinomial analysis* is used to generate the demand model. Random utility entails the assumption that the individual's true utility U consists of a deterministic part W and a random disturbance ε (see Eq. 1). The deterministic part of the utility can be parameterized as a function of observable independent variables (key customer attributes $\boldsymbol{A}$, socioeconomic and demographic attributes $\boldsymbol{S}$, and price P) and unknown coefficients β, which can be estimated by observing the choices respondents make (real or stated) and thus represent the respondent's taste, see Eq. 2. The β-coefficients and utility functions are indicated with the subscript n, representing the n^{th} respondent, the index i refers to the i-th choice alternative. There is no functional form imposed on the utility function W, i.e., W can be additive, multiplicative, quadratic, etc.

$$U_{in} = W_{in} + \varepsilon_{in} \tag{1}$$

$$W_{in} = f(A_i, P_i, S_n; \beta_n) \tag{2}$$

The probability that alternative 1 is chosen from a choice set containing two alternatives (binary choice) is then defined as the probability that the utility of alternative 1 exceeds the utility of alternative 2 or alternatively on the *probability that the difference between the disturbances does not exceed the difference of the deterministic parts of the utility*, i.e.

$$\begin{aligned} \Pr(1)\big|[1,2] &= \Pr(W_{1n} + \varepsilon_{1n} \geq W_{2n} + \varepsilon_{2n}) \\ &= \Pr(\varepsilon_{2n} - \varepsilon_{1n} \leq W_{1n} - W_{2n}) \end{aligned} \tag{3}$$

Methods such as logit (Ben-Akiva and Lerman, 1985; Hensher and Johnson, 1981) or probit (Daganzo, 1979; Hensher and Johnson, 1981) can be used to form a choice model that predicts the choice probabilities. The choice probability of the multinomial logit model is shown in Eq. 4, where $\Pr_n(1)$ is the probability that respondent n chooses alternative i.

$$\Pr_n(i) = \frac{e^{W_{in}}}{\sum_{l=1}^{J} e^{W_{ln}}} \tag{4}$$

Estimation techniques such as the maximum likelihood method and the method of least squares can be used to determine the β-coefficients such that the predictions of the model matches the observed choices as closely as possible. *The total demand for a particular design i is the summation of the predicted choice probabilities across the choice alternatives for the entire market population* (Ben-Akiva and Lerman, 1985).

2.3 Kano Method

Traditional market analysis often assumes that customer satisfaction is proportional to product performance, i.e., linear. The Kano method (Shiba *et al*, 1993), introduced in the late 1970s by Dr. Noriaki Kano of Tokyo Rika University, provides an approach to determine the generalized shape of the relation between product performance and customer satisfaction by classifying the customer attributes into three distinctive categories: must-be, basic, and excitive, see Fig. 2 (note: these terms may be named differently in various references). The three categories are described briefly, details regarding the classification process can be found in literature. *Must-be* attributes are expected by the customer and only cause dissatisfaction if not met, e.g., unusual engine noise. Customer satisfaction never rises above neutral no matter how good the engine sounds; however, the consumer will be dissatisfied if unusual engine noise occurs. Improving the performance of *basic* attributes increases satisfaction proportionally (i.e., linear), e.g., gas mileage (unless gas mileage is really bad). The *excitive* attributes increase satisfaction significantly for the reason that the customer does not expect them. For instance the oil-change interval, a (unexpected) long interval may be expected to significantly increase satisfaction.

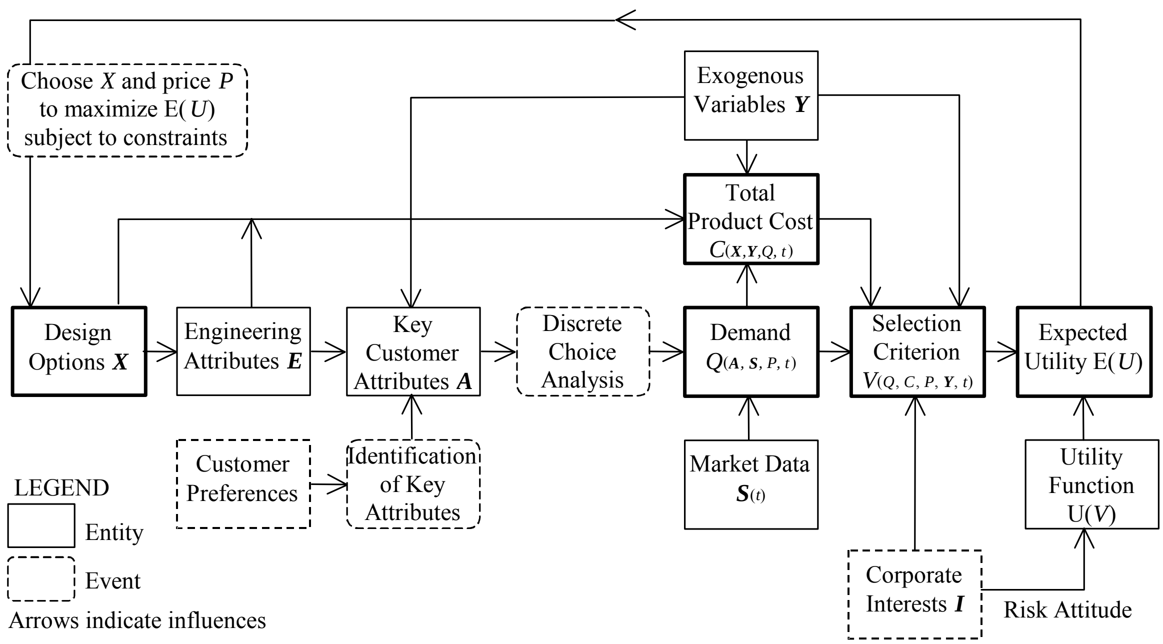

Figure 1. Decision-Based Design flowchart (Wassenaar and Chen, 2001)

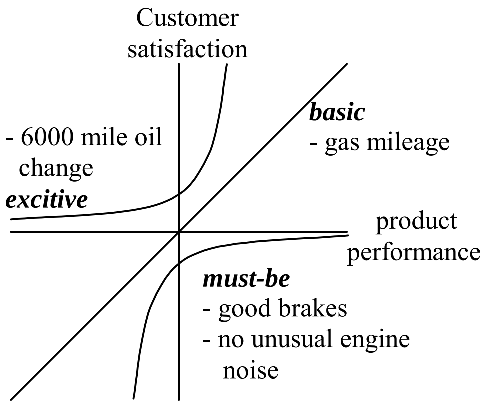

Figure 2. Kano Diagram of Customer Satisfaction

The Kano method is frequently used in the literature to rank order the attributes according to their importance to determine where to focus the engineering effort to maximize customer satisfaction in an attempt to ensure the product's success in the market. However, in our view this is incorrect. According to Arrow's General Possibility Theorem (Arrow, 1963) and demonstrated by Hazelrigg (1996), it is not possible to construct a valid social welfare function, therefore it is wrong to attempt to maximize the satisfaction of a group of consumers. Further, one should also consider the cost of improving an attribute when deciding an attribute's importance, which is not considered when rank-ordering the attributes using the Kano method. Therefore, in our view, *the use of Kano classification provides nothing more than an indication, that is, an econometric justification, of the function form.* In this work, we propose to enhance the predictive capability of DCA by using Kano method to assess the generalized shape of the customer utility functions in Eq. 2, used in the choice model of the DCA analysis. In the next section we detail how an approach for demand modeling that facilitates engineering decision-making can be implemented.

3 IMPLEMENTING DISCRETE CHOICE ANALYSIS FOR DEMAND MODELING

To facilitate engineering decision-making, a demand model is expected to relate the market demand (how much customers like the product) to engineering measures of product attributes that can be used to guide product design decision-making. In this paper, we provide guidelines for implementing DCA for product demand modeling and discuss the potential issues involved in each phase of demand modeling. Our discussion follows the sequence of the three major phases for implementing DCA:

Phase I — Identify key customer attributes $\boldsymbol{A}$, the range of price P and survey choice set; (Attributes and Choice Set Identification)

Phase II — Collect quantitative choice data of proposed designs versus alternative choice options and record customers' socioeconomic and demographic background $\boldsymbol{S}$; (Data Collection)

Phase III — Create a model for demand estimation based on the probability of choice. (Modeling)

In the later part of this section, we discuss a few other important issues related to demand modeling, including demand estimation and dynamic demand models.

3.1 Phase I - Attributes and Choice Set Identification

One can fit a mathematical model to some data using any set of explanatory variables (input of a demand model), however, the value of such a model is questionable. A useful demand model requires that the selection of customer attributes (explanatory variables) is based on econometric reasoning, that is, there exists a *causal relationship* between the customer attributes and the customer's choice which alternative to purchase. There are several methods available to assess what customers desire, what product attributes customers consider (Otto and Wood, 2000), and what competing alternatives should be considered in a discrete choice survey. Focus groups (Krueger, 1994) can be used for both existing products and products that are new (e.g., innovative design).

The outcome of surveys, focus groups, and interviews can be a long list of competing alternatives and customer desires. The customer desires identified should be clustered together into groups that share similar characteristics that are important to the customers such as, cost, performance, safety, operability, comfort, style, convenience, etc. These groups can be considered as "top-level *customer desires*". An example of specific customer desires identified for each top-level group of customer desires is provided in Fig. 3. The next step is to identify the *customer attributes* $\boldsymbol{A}$ that contribute to each customer desire. That is, to translate the customer language of the customer desires (e.g., good engine sound quality) into quantifiable engineering language (e.g., sound level, harmonics, frequency spectrum). Suitable units of measurement need to be identified for each customer desire. This transformation is very important in order to use the demand model for engineering decision making. This task consists of cooperation between market researcher and engineering specialist, and perhaps consultations with customers to verify the correct understanding of the customer desire.

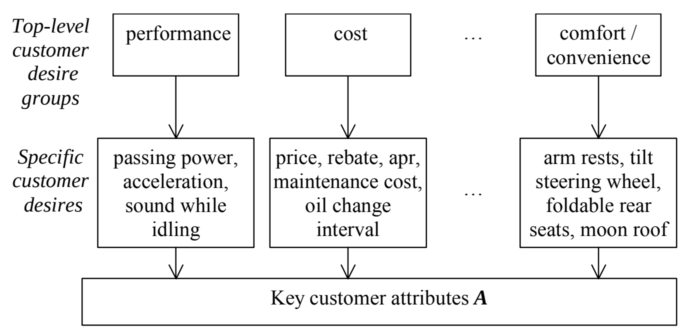

Figure 3. Cascading Customer Desires

Figure 4 further demonstrates an example on how the top-level customer desire group is cascaded to specific customer desires (customer language), and then to key customer attributes $\boldsymbol{A}$ in engineering language. The customer desire "low engine sound while idling" could be considered to belong to the group of (engine) performance. Radiated sound can be considered as the key customer attribute for measuring the engine sound while idling and engineering models can relate radiated sound to other engineering attributes such as the crankshaft stiffness, which is a function of crankshaft material (design option) among others.

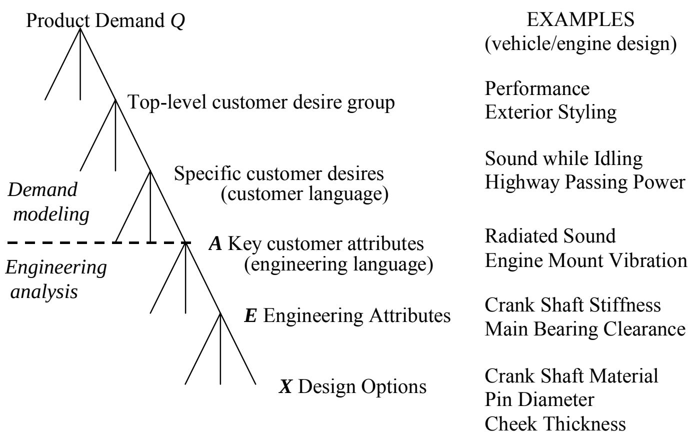

Figure 4. Attribute Hierarchy

3.2 Phase II – Data Collection

Stated Choice Vs. Revealed Choice

There are two ways of collecting choices: stated choice and revealed choice. Revealed choice concerns actual (purchase) behavior that can be observed in real choice situations. Stated choice concerns controled choice experiments that ask the respondents to state their *intend* without needing to commit to the consequence of their decision (e.g., pay the purchase price). Revealed choice can be used when similar products or services exist, e.g., when redesigning a powertool, while stated choice can also be used for innovative designs, product features, or services that do not yet exist, which have new attributes or whose levels of the key customer attributes lie outside the attribute level range of existing products. One should be aware that either stated choice or revealed choice has advantages and disadvantages (Louviere *et al,* 2000).

Survey Respondent Sampling

Several techniques can be used to sample a population (Ben-Akiva and Lerman, 1985), however any sampling design starts with defining the target market population, i.e., what consumers (customer background distribution) buy what (competing) products, the target market population size, etc. and definition of the sampling unit, i.e., a customer or a product purchase. Two common sampling strategies are random sampling and stratified random sampling. Random sampling entails randomly drawing an observation (customer or product purchase) such that the probability of that observation to be drawn equals 1/N, where N represents the market population. Random sampling cannot adequately capture the choice behavior of a very small population subgroup. This can be addressed using stratified random sampling, which divides the market population into mutually exclusive and exhaustive segments. Random samples are then drawn from each market segment. A demand model for each market segment can be constructed to predict each market segment's demand, which can then be properly weighted to arrive at an unbiased estimate for the total market demand.

3.3 Phase III - Modeling

Phase III is a quantitative process to generate the demand model. Modeling techniques such as logit or probit can be used to create a choice model that can predict the choices individual customers make and to forecast the market demand for a designed artifact based on the collected discrete choice data, whether it be revealed or stated choice. Details of logit have been introduced in Section 2.

Econometric motivation for determining the customer utility function shape

There is no functional form imposed on the form of the utility function *W*. When fitting a choice model it is common to initially assume a linear shape of the customer utility function (Eq. 2) and then to test different functional shapes (e.g., quadratic, exponential) for improvement of the model fit. Note, that we do not discuss possible interactions between attributes, e.g., by considering the ratio price/income as an attribute. It is obvious that such a try-and-improve approach lacks econometric reasoning, causality is lacking. A model obtained this way may fit the (survey) data well but its predictions may lack accuracy.

We propose to use the Kano method, introduced in Section 2, to facilitate the identification of the appropriate functional relationship between customer satisfaction and product performance, (i.e., the customer utility function) that provides the best fit of the choice model rather than try-and-improve. Quadratic or exponential function shapes can be used to fit the customer attributes that are classified by the Kano method as must-be and excitive. The predictive capability of the demand analysis should improve as the functional relationships identified using the Kano method, like random utility, better capture the underlying behavior of consumers as opposed to fitting a mathematical model to some data by randomly trying different functional shapes without proper econometric reasoning.

The maximum log-likelihood (MLL) method can be used to fit the choice model (i.e., to estimate the β coefficients). An advantage of using the MLL method is that the log-likelihood is less sensitive to a non-normal distribution of the random disturbance ε than the least square error method. Commercial software that can handle discrete choice data and offer logit or probit modeling capabilities include GENSTAT (www.vsn-intl.com), LIMDEP (www,limdep.com), SAS (www.sas.com), SPSS (www.spss.com), STATA (www.stata.com), or SYSTAT (www.systa.com).

Including a high number of key customer attributes in a choice model facilitates engineering decision-making. For instance, to capture the sound quality as experienced by the car occupants (may depend on the occupant's position in the car), the key customer attributes could include: noise level, harmonics, and frequency. These attributes included in the demand model can be used to guide engineering decision-making such as the air intake design, engine configuration, firing order, exhaust design, engine mount design, noise insulation, etc. However, while including more explanatory variables (attributes) may improve the model fit as with each additional variable more data variance can be explained, using too many explanatory variables may lead to the model fitting aspects of the data that are not due to underlying parametric features (e.g., sampling variability). Two criteria that can be used for comparing model fit and for determining whether including additional explanatory variables is useful are Akaike's Information Criterion (AIC), Eq. 5, and the Bayesian Information Criterion (BIC), Eq. 6. Both criterions penalize models for having too many explanatory variables.

$$AIC = -2L + 2p \tag{5}$$

$$BIC = -2L + p\ln(n), \tag{6}$$

where L is the log-likelihood, p the number of explanatory variables and n the number of observations (sample size). According to both criteria, the best-fitting model is the model with the lowest score. A difference of 6 points on the BIC scale indicates strong evidence that the model with the lower value is to be preferred (Raftery, 1995). In general, the BIC is more conservative than the AIC.

Another issue that may arise when using large numbers of explanatory variables is collinearity, that is, some explanatory variables may be explained by combinations of other explanatory variables. If collinearity occurs then these explanatory variables cannot be used in the choice model simultaneously. Factor analysis (Hair, 1998) could be used to combine key customer attributes (explanatory variables) that can be related to each other using econometric reasoning into a fewer number of factors. The factors could then be used to fit the choice model thus reducing the number of explanatory variables and the possibility of collinearity and over-fitting the data.

3.4 Demand Estimation

The choice model obtained using logit (Eq. 4), probit, or any other choice model structure can predict the choice probabilities for each alternative in the choice set (e.g., see Table 2, Section 4.3) given the customer's background (***S***) and descriptions of the choice alternatives. Sample enumeration is the most advanced and most accurate approach to determine a product's market demand. Sample enumeration uses a random sample of the market population *N*, to predict for each sampled individual *n* the choice probabilities to estimate the demand for the entire market population (Ben-Akiva and Lerman, 1985), see logit demand model presented in Eq. 7, where <u>*i*</u> denotes the choice alternative.

$$Q(i) = \sum_{n}^{N} \Pr_n(i) = \sum_{n}^{N} \frac{e^{W_{in}}}{\sum_{l=1}^{J} e^{W_{ln}}} \tag{7}$$

The accuracy of demand prediction can be improved by estimating a choice model per market segment to account for systematic variations of taste parameters (β coefficients) among population subgroups. Ultimately one can assume taste parameters log normal distributed across the market population (Ben-Akiva and Lerman, 1985), shown in Eq. 8. Including customer specific data in the customer background ***S*** that relates to the customer's (potential) use of the product can improve the accuracy of the demand predictions. E.g., in case of a car one can think of annual mileage driven, type of usage (commuting/recreational), etc. This product usage data can be recorded for each respondent when collecting the customer data and incorporated in the demand model.

$$Q(i) = \sum_{n}^{N} \Pr_n(i) = \sum_{n}^{N} \frac{e^{\beta_n W_{in}}}{\sum_{l=1}^{J} e^{\beta_n W_{ln}}} \tag{8}$$

A different approach for estimating the market demand is to use the choice model to predict the average choice probabilities (i.e., market shares) of the market population, (e.g., by using sample enumeration). A separate, specialized, model can be formed to estimate the total market sales volume. An advantage of this approach is that a separate model for predicting the market sales volume may be more accurate by accounting for economic growth, seasonal effects, market trends, etc., potentially leading to more accurate demand predictions.

3.5 Dynamic Demand Modeling

An issue of using demand models related to engineering design decision-making is that the market introduction of the designed product is at some point in future, while the demand model is formed at current market conditions. Since Decision-Based Design considers a product's entire lifecycle the demand model needs to predict the demand for a period of time in future, possibly spanning multiple years. Obviously market conditions, competition, and market population are subject to change, affecting the accuracy of the demand predictions. The static demand can be modeled as function of product attributes and customer background, possibly including data relating to the customer's product usage. Separate models can be constructed to predict changes over time in key customer attributes of the alternatives that are considered in the choice model, changes in demographics (e.g., birth rate, aging, ethnic shifts, etc), socio-economics (e.g., income, education, etc), and product usage (e.g., frequency, duration, type of use), while a specialized model for market volume can predict changes in market volume (Georgiopoulos *et al*, 2002) given expected economic developments, market trends, etc. To our knowledge, little research exists that combines the aforementioned models in demand analysis.

4 IMPLEMENTATION EXAMPLE

In this section we show an implementation of the three major phases of the discrete choice approach to constructing a vehicle demand model with emphasis on evaluating engine design changes in a DBD model. The demand model developed in this case study can be used to assess the impact of engine design changes on vehicle demand, facilitating the evaluation of engine design and making proper tradeoffs between performance and cost. This example is developed in collaboration with the Engine Engineering & Analytical Powertrain division of Ford Motor Company and the Power Information Network group (PIN) at J.D. Power & Associates. The statistical software package STATA is used to estimate the multinomial choice model coefficients based on the maximum likelihood criterion. Twelve vehicles (7 models, 12 trims) are considered in the demand model representing the midsize car segment, which includes vehicles like Ford Taurus, Toyota Camry, and Honda Accord. The data in the tables is normalized to protect proprietary rights of manufacturers.

Our implementation is subject to the following assumptions. A simplifying assumption is that customers *only* consider vehicles from the midsize car segment and in specific the 12 vehicle trims considered when purchasing a vehicle, in reality this may not be true. The demand model developed is a static model, as such demand changes over time are not considered. It takes approximately 48 months to redesign an engine, that is, a vehicle demand model used for engine design should be capable of predicting vehicle demand 4 years into the future. We assume however that engine design changes are immediately introduced in the market. We also assume that the designs of the other vehicles do not change, which can have a large impact on market demand. Next the three phases of implementing discrete choice analysis are presented.

4.1 Vehicle Demand Modeling - Attributes and Choice Set Identification

Based on J.D. Power's Vehicle Quality Survey we identify five top-level customer desires related to vehicle choice at the vehicle system level, these are: price/cost, engine/transmission performance, comfort & convenience, ride/handling performance, and roominess. For reasons of simplicity we do not consider customer desires related to sound system, seats, and style. Specific customer desires can be identified for each top-level vehicle system customer desire. Take engine/transmission performance as an example, the specific customer desires include performance during rapid acceleration, passing power at highway speeds, fuel economy, range between fuel stops (which highly impacts the *perceived* fuel economy), sound while idling, and sound at full throttle acceleration. To facilitate the demand evaluation of product design it is necessary to identify appropriate customer attributes that can capture the customer's intend reflected in the customer desires. Interaction between engineering experts at Ford Motor Company and market research specialists from J.D. Power helped identify the key customer attributes corresponding to the customer desires. Linking the key customer attributes with engineering attributes, and then design options like the cascading shown in Fig. 4 is also an important activity of designers, however, it is not covered in our case study as our study focuses on demand modeling and does not reach beyond the key customer attribute level.

In total we consider 21 key customer attributes related to the general vehicle design (e.g., price, vehicle length, legroom), 10 key customer attributes related to engine design at the system level (e.g., horsepower, fuel economy), 5 socio-economic attributes of customers (e.g., age, income), and 2 dummy variables. The dummy variables can be used to separate the slopes (i.e., separate β coefficients) of key customer attributes when there is a strong correlation between vehicle origin (domestic or import) or engine configuration (inline 4 cylinder or V6) with the recorded customer choice, which may lead to problems (e.g., illogical coefficient estimates) when fitting the logit choice model.

4.2 Vehicle Demand Modeling – Data Collection

The demand model will be used to evaluate design changes of an existing product. Therefore, the demand model is created using revealed choice data at the respondent level provided by J.D. Power. The data consists of 2552 observed individual vehicle purchases from the US year 2000 vehicle market of the seven vehicles considered in the case study including respondent background. The values of customer attributes related to the general vehicle descriptions such as weight, fuel economy, legroom, etc. of the 12 discrete choices are obtained from Ward's Automotive. Measures such as 0-60 mph time, 0-30 mph time, or 30-50 mph time can capture the engine/transmission performance as experienced by the customers however this data is not available for all vehicle trims. Instead, the engine performance is considered using (maximum) horsepower, torque and low-end torque. By taking the reciprocal of these measures with respect to the vehicle mass a measure of the vehicle's performance is obtained. The values of customer attributes like, ride, handling, noise, and vibration are provided by Ford Motor Company. Part of the data's correlation matrix is presented in Table 1.

Table 1. Partial Correlation Matrix

	gender	age	income	usa/import
gender	1			
age	-0.192	1		
income	-0.074	-0.176	1	
usa/import	0.150	-0.220	0.087	1
msrp_price	0.006	-0.041	0.141	0.183
rebate	-0.101	0.256	-0.141	-0.869
apr	-0.072	0.173	-0.017	-0.425
resale index	0.178	-0.215	0.031	0.869
hp/mass	-0.011	-0.104	0.180	0.212
torque/mass	-0.051	-0.005	0.148	0.013
le torq./mass	-0.087	0.036	0.120	-0.255
fuel econ.	0.127	-0.047	-0.102	0.444
fuel range	0.138	-0.063	-0.045	0.680
wheel base	-0.106	0.076	0.050	-0.667
veh. width	-0.119	0.157	-0.066	-0.918
veh. lengt	-0.149	0.154	-0.038	-0.907
overhang	-0.137	0.185	-0.097	-0.854
front-headrm	-0.013	-0.103	0.145	0.290
frong-legrm	0.072	-0.094	0.116	0.762
rear-headrm	-0.162	0.132	0.053	-0.695
rear-legrm	-0.140	0.157	0.013	-0.731
trunk space	-0.132	0.139	0.004	-0.844

The variables gender and usa/import of Table 1 are binary variables: female = 1, and import = 1, otherwise 0. Given the negative correlation between gender and vehicle size it seems that women buy smaller cars. The negative correlation between usa/import and rebates indicates that imports are generally sold with smaller rebates. The correlation between customer background and key customer attributes appears to be very weak. However, this is not necessarily a disadvantage. Highly correlated variables are prone to being collinear, which is undesirable. Further, high correlation between the dependent variable (in this case the vehicle choice) and independent explanatory variables implies that few variables are sufficient to predict vehicle choice, making it difficult to include many variables (customer attributes) required for engineering design decision-making. Obviously, high correlation does not necessarily indicate a causal (i.e., econometric) relationship required for accurate demand predictions. Note that import/usa is used as a dummy variable. The next phase is to create the demand model using the collected revealed choice data, and the vehicle description (key customer attributes).

4.3 Vehicle Demand Modeling – Multinomial Logit

Customer purchase data (revealed choice) by itself is not sufficient to construct the demand model. A database (input data) needs to be formed containing for each individual 12 rows of data, one for each choice alternative, each row containing the customer background, the key customer attributes that describe the vehicle, and the respondent's observed choice (real purchase). That is, the demand modeling specialist recreates the choice set considered by the customer when faced with the purchase decision *as perceived by the specialist*. An example of a choice set (which could be used to elicit stated choice in a discrete choice survey) containing the vehicles considered in this case study is presented in Table 2.

Table 2. Case Study Choice Set

example of choice set								totals	
vehicle id	engine type	displace ment	horse power	fuel econ.	msrp price	cust. choice	vehicle id	choice rate	% market share
1	I4	0.80	0.76	1.15	0.87	0	1	251	9.84
2	V6	1.08	1.02	0.98	1.15	0	2	190	7.45
3	I4	0.73	0.73	1.12	0.89	0	3	335	13.13
4	V6	1.04	1.02	0.98	1.05	0	4	220	8.62
5	V6	0.94	1.06	1.02	1.03	0	5	231	9.05
6	V6	0.94	1.06	1.00	0.89	0	6	192	7.52
7	V6	1.08	1.28	1.00	1.07	0	7	199	7.80
8	V6	1.08	0.93	1.00	0.96	0	8	167	6.54
9	V6	1.21	1.10	0.96	1.07	0	9	67	2.63
10	V6	1.19	1.01	0.98	1.11	0	10	435	17.05
11	V6	0.83	1.01	0.94	0.89	0	11	213	8.35
12	V6	1.08	1.01	0.90	1.02	1	12	52	2.04

The choice set presented in Table 2 shows the 12 vehicle choices with *partial* (normalized) description of the explanatory variables used in the customer utility function of the logit choice model. The customer choice is treated as a binary variable and in this particular case the customer selected vehicle 12. The second table shows the total number of vehicles purchased of each vehicle trim (choice rate) and the market shares as identified in the revealed data. In total the database contains 30624 observations (2552 respondents * 12 vehicles). STATA's conditional logit function is used to estimate the choice model, which is equivalent with multinomial logit (Greene, 1993). Conditional logit enables consideration of multinomial choice (i.e., one vehicle picked from the choice set) as a group of binary choices (pick/not-pick).

A linear customer utility function shape is initially considered for the utility function used in the logit choice model (Eq. 2). All customers share the same utility function coefficients, i.e., market segmentation is not considered. It is necessary to evaluate several different interactions of explanatory variables. We tested over 200 customer utility functions with different combinations of linear and interaction items. Eventually a model using 38 explanatory variable items is selected based on the Bayesian Information Criterion score (BIC) (see description in Section 3.3). Results of the top six most promising models are shown in Table 3. The BIC score indicates that model 6 should be preferred as its BIC score differs more than 6 points from the other models, which is considered strong evidence (Raftery, 1995). The MS_R2 represents the R2 error measure of the observed market shares vs. predicted market shares.

Though model 1 fits the data very well, model 1 demonstrates that fitting a mathematical model to some data may lead to poor predictions. All 11 explanatory variables of model 1 are key customer attributes related to the vehicle. This implies that model 1 cannot consider customer background. As such, model 1 may predict well for a market (e.g., state/region) whose population matches the population that produced the vehicle purchase data on which model 1 is formed, but not when the market population differs. Therefore it may be expected that model 1's predictions are poor given the lack of connectivity with the customer background. Table 3 shows that the observed market shares and the market shares as predicted by model 6 match quite well. The customer utility function of model 6 includes customer background, key customer attributes, and selected interactions between key customer attributes and customer background (e.g., horsepower/income). Except interactions, it should be noted that no higher than second order items are considered in the models tested so far.

Table 3. Model Comparison and Model 6 Market Share Prediction

model id	lik.hood.	Expl.Var.	MS_R2	BIC
model 1	-362457.6	11	1	725028.78
model 2	-350915.1	43	0.995346	702274.37
model 3	-355220.2	31	0.989278	710760.56
model 4	-350890.6	40	0.994903	702194.28
model 5	-350789	40	0.994941	701991.22
model 6	-350630.6	38	0.994941	701653.78

veh.id	observed	predict
1	0.0984	0.1002
2	0.0745	0.0771
3	0.1313	0.1304
4	0.0862	0.0851
5	0.0905	0.0846
6	0.0752	0.0718
7	0.0780	0.0771
8	0.0654	0.0669
9	0.0263	0.0266
10	0.1705	0.1760
11	0.0835	0.0840
12	0.0204	0.0202

As proposed earlier in Section 3.3, the Kano method is used to improve the predictive accuracy by providing the econometric justification for identifying appropriate shapes for the customer utility function of the choice model (Eq. 2). The Kano classification of the key customer attributes for this engine design case study is provided by Ford specialists. All key customer attributes should be considered *basic*, except *very high fuel economy* can be classified as *excitive*. It is decided to use a quadratic function shape for the key customer attributes *fuel economy* and *range between fuel stops* for model 6. The BIC score shown in Table 4 indicates that the demand model using the utility function shape as assessed by the Kano method should be preferred as its BIC score differs more than 6 points.

Table 4. Quadratic (Kano) vs. Linear Function Form

	Kano (quadratic)	Regular (linear)
MS_R2	0.998293	0.995984
MLL	-5820.69	-5831.48
BIC	11930.61	11941.85

4.4 Vehicle Demand Prediction and Integration with Decision-Based Design

A matching sample of the target market population (e.g., obtained using random sampling) is used to predict the market shares using choice model 6. Matching in the sense that the distribution of age, income, and gender, which are explanatory variables in model 6, of the sample and the target market are comparable. The 12 choice probabilities (one for each vehicle of the choice set) are determined for each sampled individual using that individual's background and the descriptions of the 12 vehicles of the choice set. The choice probabilities of the vehicles are aggregated across all individuals to obtain the predicted market shares. The market shares are then multiplied with the total vehicle demand of the target market to obtain the demand for each vehicle. Prediction of the impact of customer attribute changes (which reflect engineering design changes) on vehicle demand is possible by updating the vehicle descriptions and recalculating the predicted choice probabilities. In our study, our special interest is to study the impact on market shares when design changes are made to the engine (attributes) of a particular vehicle in the choice set. For example, the impact on market shares when improving the fuel economy of the Taurus with 5% yields a market share increase of 6.67%. The demand model that captures the relationship between key customer attributes and market demand can be used to explore "*what if*" scenarios or to set targets for key customer attributes that maximize market demand. Decision-Based Design (see Fig.1), i.e., the selection of the best design option $\boldsymbol{X}$ to maximize the profit is possible when an appropriate product cost model C can be formed and integrated with the vehicle demand model Q, along with the engineering analysis models that capture the relationship between design options $\boldsymbol{X}$ and key customer attributes $\boldsymbol{A}$ via engineering attributes $\boldsymbol{E}$. A detailed example of DBD is not provided due to the space limitation.

4.5 Cross Validation of Demand Model

Achieving a good model fit indicated by error measures such as the log-likelihood and R-square value indicates that the model will predict well on the data used to form the model, however, it does not guarantee accurate predictions of the demand model at new model inputs. The approach we take for validating the obtained vehicle demand model is through the technique of cross-validation (Breiman and Spector, 1992) which does not require the collection of additional data. Cross-validation entails dividing the data into k subsets of approximately equal size (k-fold cross-validation). The choice model is then fitted k times using k-1 subsets of data. Each time the fitted choice model is used to predict the choice of the remaining data set to calculate the measure of error (e.g., log likelihood score). The average of the obtained error measurements serves as the error measure of the model. A lower k value is chosen for large data sets, in general a 10-fold (k=10) or 5-fold cross-validation is recommended. Model 6 is validated using 5-fold cross-validation given the large data set. The dataset consisting of 2552 individuals is divided into 5 subsets of approximately equal size using random sampling. Model 6 is fitted to the combined data of 4 out of the 5 datasets. The fitted model is then used to predict the choice for the remaining choice set and the R2 value for the market shares, which is used as error measure, is calculated. This procedure is repeated 5-fold, every time using a different data set from the 5 data sets for prediction and error measure calculation. The R2 value of model 6 fitted on the full data set is 0.99. The R2 value decreased to an average of 0.92 for the 5 cross-validation tests, which is still an acceptable value. Cross-validation helps us build more confidence in using the proposed DCA approach to demand modeling and demand prediction. It also shows that the accuracy of model 6 is satisfactory.

5 CONCLUSION

Building upon our earlier work on using the discrete choice analysis approach to demand modeling we develop in this paper guidelines for implementing the discrete choice demand modeling approach in product design. The modeling of a hierarchy of product attributes is introduced to cascade customer desires to specific key customer attributes. *The advantages of the proposed demand analysis procedure can be summarized as*: (1) The method does not involve any ranking, weighting, or normalization, thus avoiding the paradox associated with many multicriteria approaches. (2) Probabilistic choice addresses the uncertainties associated with unobserved taste variations, unobserved attributes, and model deficiencies. (3) Competing products are considered, enabling analysis of market impact and competitive actions through "*what if*" scenarios. (4) Choices do not necessarily share the same set of attributes or attribute levels, expanding market testing possibilities and leaving more freedom to the marketing engineer. (5) The customer survey embedded in DCA resembles real purchasing behavior more closely, reducing respondent errors and enabling the analysis of more attributes. The Kano method is adapted to provide econometric justification for selecting the shape of the customer utility function, which should improve the capturing of underlying purchase behavior, enhancing the predictive capability of demand models. The obtained demand model for vehicle engine case study is shown to be satisfactory through cross-validation.

There are some misperceptions about the ability of the demand modeling approach to avoid Arrow's Impossibility (Arrow and Raynaud, 1986), which states that a group of decision-makers behaves intransitive, i.e., irrational. Problems with transitivity arise when a group of decision-makers (e.g., customers) is treated as if it were a single average individual. We do not attempt to draw conclusions based on aggregated data of a group of decision-makers but consider each decision-maker (customer) at the disaggregated individual level. To illustrate this, consider the following statements: John owns a car. John owns a bike. We may correctly conclude that John owns a bike and a car. We cannot project such conclusions on a group of decision-makers. For example, consider 100 customers; 35 own a bike, 20 own a bike and a car, and 45 own a car only. Then the following statements are true: the majority of the customers own a bike, the majority of the customers own a car. However, the conclusion that the majority of consumers own a bike and a car is false. The examples above show that *a group of decision-makers cannot be represented by an imaginary average individual*. As such, Arrow showed that it is impossible to construct a single social welfare function (Arrow, 1963) and it is therefore impossible to maximize some group utility, e.g., customer satisfaction as demonstrated by Hazelrigg (1996). *Optimizing demand (i.e., to maximize expected utility of profit) using DCA does not entail maximizing some customer utility function, but optimizing the summation of choice probabilities across individual decision-makers (customers), thereby avoiding Arrow's Impossibility*.

It should be noted that our proposed engineering design alternative selection approach, consisting of the integration of DBD and DCA, does not entail selecting the best alternative from the choice set used in the survey (e.g., choice set with alternatives A, B, or C) but determining what design alternative A from design alternative set A^* best competes against the alternatives considered in the choice set (e.g., competing products or services B and C). That is, the integrated DBD-DCA model is used to determine what design alternative performs best in the market, i.e., maximizes expected utility of profit. The addition or exclusion of alternatives from the design alternative set A^* does not affect the choice set and therefore cannot affect the valuation and thus the (absolute) value (i.e., expected utility) of the design alternatives that are included in A^*, thus avoiding Arrow's Impossibility.

Our proposal to employ the Kano method to select and econometrically justify the customer utility function shape requires further development. Another approach being developed is to enhance the capturing of the customer's perception of the key customer attributes through consideration of the unobservable top-level customer desires in the customer utility function using latent variables.

ACKNOWLEDGMENTS

We like to thank the specialists at J.D. Power & Associates and Ford Motor Company for their thoughtful contributions and their efforts to gather data for our vehicle demand model. We also thank J.D. Power & Associates for providing the opportunity to work with vehicle demand modeling experts during an internship in Summer 2002. The support from NSF grant DMII 0217702 are acknowledged.

REFERENCES

Arrow, K.J., "*Social Choice and Individual Values*", John Wiley & Sons, Inc., New York, 1963.

Arrow, K.J. Raynaud, H., *Social Choice and Multicriterion Decision-Making*, Massachusetts Institute of Technology, 1986.

Ben-Akiva, M and Lerman S.R., *Discrete Choice Analysis*, The MIT Press, Cambridge, Massachusetts, 1985.

Breiman, L., Spector, P., 'Submodel selection and evaluation in regression: The X-random case", *International Statistical Review*, 60, 291-319, 1992.

Cook, H. E., *Product Management: Value, Quality, Cost, Price, Profit, and Organization*, Chapman & Hall, London, 1997.

Daganzo, C., "Multinomial Probit, the theory and its application to demand forecasting", Academic Press Inc., New York, 1979.

Donndelinger, J., Cook, H.E. "Methods for Analyzing the Value of Automobiles," SAE Paper 970762, Warrendale, PA, *Society of Automotive Engineers, Inc.*, Feb. 1997.

Fowlkes, W.Y. and Creveling, C.M., *Engineering Methods for Robust Product Design*, Addison-Wesley Publishing Company, Reading, MA, 1995.

Georgiopoulos, P., Fellini, R., Sasena, M., Papalambros, P.Y., "Optimal Design Decisions in Product Portfolio Valuation", *Proceedings of the 2002 ASME Design Engineering Technical Conferences*, Montreal, Canada, September 29-October 2, 2002

Green, P.E. and Srinivasan, V., "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice", *Journal of Marketing,* 1990.

Green, P.E. and Srinivasan, V., "Conjoint Analysis in Consumer Research: Issues and Outlook", *Journal of Consumer Research*, Vol. 5, 1978.

Green, P.E. and Wind, Y., "New Ways to Measure Consumer Judgments", *Harvard Business Review,* 1975.

Greene, W.H., *Econometric Analysis*, second edition, Macmillan Publishing Company, New York, 1993.

Hair, J.F., (editor), Anderson, R.E., Tatham, R.L., Black, W.C., "*Multivariate Data Analysis*", 5th ed, Prentice Hall College Div,., 1998.

Hazelrigg, G.A., "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design", *ASME Journal of Mechanical Design*, 118, 161-164, 1996.

Hazelrigg, G.A., "A Framework for Decision based Engineering Design", *ASME Journal of Mechanical Design*, Vol. 120 , 653-658, 1998.

Hazelrigg, G.A., "An Axiomatic Framework for Engineering Design", *Transactions of the ASME*, Vol. 121, 342-347, 1999.

Hensher, D.A. and Johnson, L.W., "Applied Discrete Choice Modeling", Halsted Press, New York, 1981.

Krueger, R.A., *"Focus groups: a practical guide for applied research"*, 2nd ed, Thousand Oaks, Sage Publications, California, 1994.

Li H. and Azarm, S., "Product Design Selection under Uncertainty and with Competitive Advantage", *ASME Design Technical Conference,* DETC2000/DAC-14234, Baltimore MD, 2000.

Louviere, J.J, "Why Stated Preference Discrete Choice Modelling is NOT Conjoint Analysis (and what SPDCM IS)", Memetrics white paper, 2000.

Louviere, J.J., Hensher, D.A., Swait, J.D., "*Stated Choice Methods, Analysis and Application*", Cambridge University Press, 2000

Luce, R., *Individual Choice Behavior: A Theoretical Analysis*, *Wiley*, New York, 1959.

Phadke, M.S., 1989, *Quality Engineering using Robust Design*, Prentice Hall, Englewood Cliffs, New Jersey.

Marschak, J., "Binary Choice Constraints on Random Utility Indicators", *Stanford Symposium on Mathematical Methods in the Social Sciences*, K. Arrow, ed., Stanford University Press, Stanford, California, 1960.

Otto K.N. and Wood, K., *Product Design: Techniques in Reverse Engineering and New Product Development*, Prentice Hall, Upper Saddle River, NJ, 2001.

Raftery, A., "Bayesian Model Selection in Social Research", *Social Methodology*, 1995.

Shiba, S., Graham, A., Walden, D., ***N****ew American TQM: four practical revolutions in management,* Productivity Press, Cambridge, Mass. 1993.

Thurstone, L., "A Law of Comparative Judgment", *Psychological Review* 34, p273-286, 1927.

Train, K., "*Qualitative Choice Analysis, theory, econometrics, and an application to automobile demand*", MIT Press, Cambridge, MA, 1986.

Tversky, A., "Elimination by Aspects: A Theory of Choice", *Psychological Review* 79, p281-299, 1972.

Wassenaar, H.J. and Chen, W., "An Approach to Decision Based Design", *Proceedings of the 2001 ASME Design Engineering Technical Conferences*, Pittsburgh, PA, September 9-12, 2001. Accepted by ASME Journal of Mechanical Design.

