Proceedings of DETC'04
ASME 2004 Design Engineering Technical Conferences and Computers and Information in Engineering Conference
Salt Lake City, Utah, September 28 – October 2, 2004

DETC2004/DTM-57487

AN INTEGRATED LATENT VARIABLE CHOICE MODELING APPROACH FOR ENHANCING PRODUCT DEMAND MODELING

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

Wei Chen[1]
Integrated Design Automation Laboratory
Department of Mechanical Engineering
Northwestern University
Evanston, Illinois 60208-3111
weichen@northwestern.edu

Jie Cheng
Power Information Network LLC.
Troy, Michigan 48098

Agus Sudjianto
Ford Motor Company
Dearborn, Michigan 48124

ABSTRACT

In today's highly competitive economy it is increasingly important to consider customer desires in engineering design on a systems level, that is, there is a need for integrating business decision-making and engineering decision-making. Building upon the earlier work on using the discrete choice analysis approach to demand modeling, in this work, the discrete choice analysis method is enhanced by introducing latent variables to include the customer's attitude and perception in a demand model. The latent variable approach better captures psychological factors that affect the purchase behavior of customers and facilitates the understanding of the relationship between customers' desires and product features. The approach is expected to enhance the predictive accuracy of demand models and help verify a designer's intent by assessing the contributions of various product attributes to the customer's perceived product performance. In this work, the existing binary latent variable discrete choice model is extended for multinomial choice and the mathematical formulation is developed to integrate a latent variable model and a multinomial logit choice model. The demand modeling of passenger vehicles with emphasis on studying the impact of engine design attributes is used to demonstrate the potential of the proposed approaches. It is important to keep in mind that the example's emphasis is on demonstrating the approaches rather than the results *per se*.

Key words: decision-based design, demand modeling, discrete choice analysis, vehicle design, latent variable analysis

1 INTRODUCTION

In enterprise-driven product design, a product demand model is crucial for the evaluation of both the profit made from a product and the cost associated with producing that product. The Decision-Based Design (Hazelrigg, 1998; Antonsson *et al.*, 2003; Olewnik *et al.*, 2003; Wassenaar and Chen, 2003) is such an approach that utilizes the expected utility of the profit as the single criterion for the design alternative selection. The need for a demand modeling approach that can accurately predict a product's demand as a function of its design options is shown in Figure 1. It shows that demand (and price) play a prominent role in assessing the profit (net revue) a product can bring.

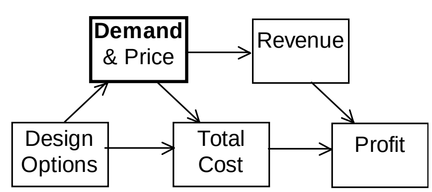

Figure 1. Critical role of demand analysis in Decision-Based Design

[1] Corresponding Author: 2145 Sheridan Road, Tech B224, Evanston, IL 60208-3111; Phone: (847) 491-7019, Fax: (847) 491-3915.

Though important, little research on demand modeling exists in the field of engineering design. Under the context of enterprise-driven design, a few approaches have been proposed for estimating market share or product demand. One is the S-Model Approach proposed by Cook (Donndelinger and Cook, 1997), of which the key concept is the Taylor expansion about a reference point where the value and price of all products are identical, resulting in equal demand for each product. Li and Azarm proposed the Comparing Multi-attribute Utility Values Approach (2000) and later a Customer Expected Utility Approach (Besharati et al., 2002). The former approach estimates the demand by comparing multi-attribute utility values obtained through conjoint analysis (Green and Wind, 1975), while the later approach assumes that the customer purchases a product when the product attributes satisfy certain limits (thresholds). In the work of Georgiopoulos, et al. (2002), a demand-based enterprise economic model is developed for product portfolio valuation. Michalek, et al., (2004) developed a method to coordinate marketing models of user and producer preferences with engineering design models of product performance.

In our earlier work (Wassenaar and Chen 2003; Wassenaar, *et al*., 2004), the discrete choice analysis (DCA), a disaggregate approach is proposed for demand modeling under the DBD framework. As pointed out in these works, it is not sufficient for a design engineer to use a (normative) multiattribute (utility) decision-making approach to represent the design preference of a whole market population. For similar reasons, the use of aggregated demand models for engineering design decision-making is limited as they insufficiently capture the variability found among customers.

Disaggregate demand models such as DCA, use data of individuals instead of group averages, thus enabling a more accurate capturing of the variation of characteristics of individuals and can avoid paradoxes associated with aggregating preference of multiple individuals.

In the DCA demand model of Figure 2 it is shown how the choice of *individual* customers y is modeled as a function of the key customer attributes $\boldsymbol{A}$, the customer's socioeconomic background $\boldsymbol{S}$ (e.g., gender, age, income, education), and the price P, using the customer utility U as an (unobserved) intermediate variable. The discrete choice model enables the prediction of a customer's choice y by assuming that the customer maximizes utility U. The key customer attributes capture the physical features (e.g., power, brand, price, and warranty) of the set of the products (choice set), describing the product (model inputs) as a list of facts with quantitative engineering descriptions. However, the way a customer views and values a product depends on the customer's experience and knowledge. It is likely that the customer's perception of product attributes differs from the attribute's literal meaning and level. Ratings customers give about the fuel economy of their vehicle illustrates as an example this perception issue. It is discovered from the J. D.Power survey data that a vehicle with a fuel economy of 24.1 mpg receives an average customer rating of 8.1 for its fuel economy. However, another vehicle (different brand but belonging to the same midsize car segment) with the same fuel economy receives an average rating of 7.7. This data indicates that *the actual attribute level and the level as perceived by the customer may not match*. Another issue is that customers may not compare products at the detailed attribute level but rather on a more abstract level. A vehicle can be described by engine power, number of seats, trunk space, etc., but what the customer really cares about is transportation, safety, status, etc., which are labeled top-level customer desires (Wassenaar *et al.*, 2003).

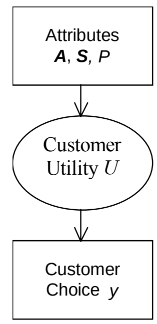

Figure 2. Standard discrete choice model

To overcome the aforementioned difficulty, latent variable analysis is proposed in this work to capture the customer's top-level desires and perceptions and integrate them with discrete choice analysis. Latent variable modeling (Everitt, 1984; Loehlin, 1998) is a technique that can capture the customer's perception through the use of psychometric data, obtained through conducting surveys. Psychometric survey questions ask the consumers to indicate how satisfied or dissatisfied they are with respect to aspects of latent variables. For instance, in Figure 3 the customer's *perceived performance* is modeled as a function of customer background and product attributes relating to the vehicle's performance. The level of the latent variable *performance* can be measured using the ratings customers give on aspects related to *performance*, e.g., *passing power at highway speeds*. In Figure 3 it is shown that the levels of latent variables are *caused* by the product features and related to the customer's background, while the levels of the indicators are *caused* by the levels of the latent variables.

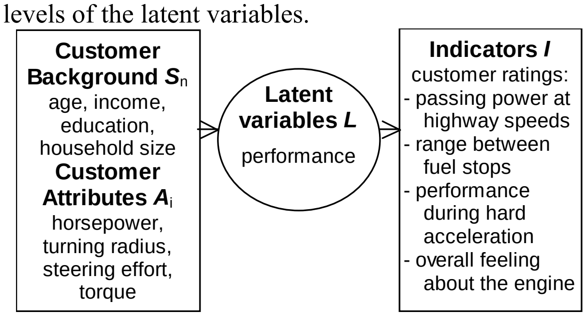

Figure 3. Latent variable example for performance

Latent variable analysis is related to factor analysis and structural equation modeling in that similar analysis techniques are used. Keesling (1972), Joreskog (1973), Wiley (1973), and Bentler (1980) have contributed to the latent variable theory by developing the structural and measurement equation framework and a methodology for specifying and estimating latent variable models (Ben-Akiva et al., 2002). Different approaches for including indicators of psychological factors into choice models have been developed. For example, Koppelman and Hauser (1979) included indicators directly into the customer utility function. Phrashker (1979) proposed to first fit the latent variables to indicators and subsequently to include the fitted latent variables in the customer utility function (Phrashker, 1979). However, these approaches have their shortcomings. Ben-Akiva extended the above approaches by formulating a general treatment of the inclusion of latent variables in discrete choice models (Ben-Akiva et al., 2002), shown in Figure 4. The integrated model takes into account the influence of the customer's perception on purchase behavior by employing the latent variables ***L*** as *additional predictors* to help predict the customer's utility *U* and customer choice *y*.

In this work, we address the need to consider the customer's top-level desires and perceptions with respect to the relationship between engineering design and customer purchase decisions. The discrete choice modeling approach is enhanced with latent variable choice modeling and a systematic approach for implementing the integrated latent variable demand modeling approach for engineering design is developed. The current theory and application of latent variable modeling seems to be limited to binomial choice, i.e., buy/not-buy decisions. However, many realistic choice settings involve more than binomial choice. Therefore, there is a need to extend the integrated latent variable discrete choice model to multinomial choice situations.

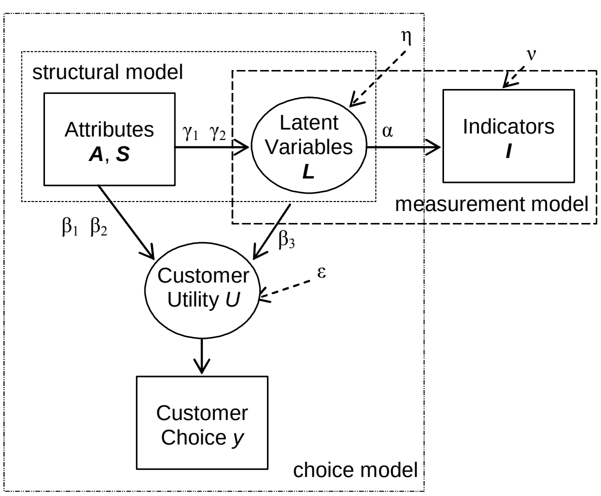

Figure 4. Integrated latent variable model with structural model and measurement model

The outline of this paper is as follows: in Section 2, the necessary background of choice modeling and latent variable analysis is provided. In Section 3 of this paper the discrete choice modeling approach is extended to the integrated latent variable choice modeling and an approach to implement the integrated approach in five steps is developed. Additionally, the maximum likelihood function of the integrated latent variable model for multinomial choice is derived. The proposed approaches are demonstrated in Section 4 using a (passenger) vehicle demand modeling as an example. The focus of the example is on demand modeling rather than decision-based design optimization. The use of demand models for product design decision-making has been illustrated (Wassenaar, et al. 2003), but is not the focus of this study. Concluding remarks and direction for future research are presented in Section 5.

2 CHOICE MODELING AND LATENT VARIABLE MODELING BACKGROUND

Discrete choice analysis (DCA) identifies patterns in choices customers make between competing products and it generates the probability that a particular 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 Equation 1). The deterministic part of the utility can be parameterized as a function of observable independent variables (key customer attributes ***A***, socioeconomic and demographic attributes ***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 Equation 2) The β-coefficients and utility functions are indicated with the subscript n, representing the nth 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. Assuming that the error terms (ε) are independently and identically distributed (IID property), the choice probability of the multinomial logit model can be formed as shown in Equation 4, where $Pr_n(i)$ is the probability that respondent n chooses alternative *i* from a set of *J* competing choice alternatives.

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).

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

The latent variable model (Figure 4) consists of two parts, a *structural model* and a *measurement model*. The structural model models the latent variables ***L*** as a function of the customer attributes ***A***, and the customer background ***S***, and consists of one equation per latent variable. The measurement model measures the relationship between the latent variables ***L*** and the indicators ***I***. A structural model equation is presented in Equation 5, where the indices for the latent variables, customer attributes, and customer background are omitted. The random disturbance η, which captures the variability of the customer's perception, is often assumed normally distributed.

$$L = \gamma_1 A + \gamma_2 S + \eta \tag{5}$$

Similarly, the structural equation for the customer utility *U* with the random disturbance ε can be written, which accounts for unobserved attributes, taste variation, and modeling deficiencies (compare to Equation 1).

$$U = \beta_1 A + \beta_2 S + \beta_3 L + \varepsilon \tag{6}$$

Note, that the functional form of the structural equations is not prescribed but is often assumed additive. The measurement model contains one equation per indicator as shown in Equation 7. One measurement equation per latent variable is constrained to one (1) for identification (Loehlin, 1998).

$$I = \alpha L + \upsilon \tag{7}$$

The random disturbance ν is often assumed to be normally distributed and is an indication for the model fit, i.e., ν is an indication of how well the measurement equation explains the variance found in the observed indicator values, where a value of 0 for ν indicates a perfect fit. The choice model is, strictly speaking, also a measurement equation in that choice *y* can be seen as an indicator of the customer's utility *U*. The relationships contained in the measurement and structural models should be causal. The Equations 4, 5, 6, and 7 combined form the integrated discrete choice latent variable model.

3 INTEGRATING A LATENT VARIABLE MODEL INTO DEMAND MODELS FOR PRODUCT DESIGN

A step-by-step approach to implementing the integrated latent variable demand modeling approach is presented in this section and the maximum likelihood function of the integrated latent variable model for multinomial choice is derived.

<u>Steps for Implementing an Integrated Latent Variable Demand Model</u> The descriptions in this section are kept succinct in favor of more detailed explanations in the example section (Section 4). Implementing the integrated latent variable modeling approach to product design involves five steps. Step 1, determination of what key customer attributes *A*, and customer background *S* to consider, is similar to when using the traditional multinomial demand model in previous work (Wassenaar and Chen, 2003) for discrete choice modeling and will not be repeated here. In Step 2 the goal is to assess how many and what latent variables *L* should be considered given *A* and *S*. The choice of latent variables is largely based on designers' insight into the problem and their understanding of the cause/effect relationships in the system. Unfortunately, like any model formation, there is no formal guideline for selecting the latent variables, even though factor analysis (Hair *et al*., 1998) (see details in Section 4) can be used to verify the number of latent factors chosen. Step 3 involves the identification of a sufficient number (i.e., four or more) of indicators *I* for each latent variable to measure their levels, again the choice is mainly based on designers' insight. Step 4 entails the determination of the structural equation and measurement equations for the latent variable model and for the choice model. The structural and measurement equations have been detailed in Section 2. Step 5, the development of the maximum likelihood function and estimation of the integrated latent variable model is explained in more detail next.

<u>Development of the Likelihood Function for the Integrated Multinomial Model</u> A likelihood function of the integrated latent variable model with a multinomial logit choice model for estimation of grouped choice data is currently not available in the literature. As a part of our research, we develop the generic likelihood function LF for the integrated model as presented in Ben-Akiva (Ben-Akiva et al., 2002), which is shown in Equation 8.

$$\text{LF} = f(y, I | Z; \alpha, \beta, \gamma, \Sigma_\varepsilon, \Sigma_\nu, \Sigma_\eta) = \int_L \underbrace{\Pr(y|Z, L; \beta, \Sigma_\varepsilon)}_{\substack{choice \\ model}} * \underbrace{g(L|Z; \gamma, \Sigma_\eta)}_{\substack{structural \\ model}} * \underbrace{h(I|Z, L; \alpha, \Sigma_\nu)}_{\substack{measurement \\ model}} dL \tag{8}$$

where $Z = (A, S, P)$, L - latent variables

In Equation 8 it is shown how the likelihood consists of the integration of the choice model over the distribution of the latent variables ***L***, including the indicators ***I***, which is the joint probability of the observed variables *y* and indicators ***I***, given the attributes ***Z***, which includes the customer attributes ***A***, customer background ***S***, and price *P* and unknown parameters α, β, γ, and the variances of the random disturbances Σ_ε, Σ_η, and Σ_η. Note, that the covariance of the random disturbance components η, ε, and ν in this model are assumed to be zero. Using Equation 8, Ben-Akiva derived the likelihood function for binary (probit) choice, shown in Equation 9 (Ben-Akiva et al., 2002)

$$\mathrm{LF} = \int_L \int \Phi\{y(Z\beta_1 + L\beta_2)\} * \prod_{k=1}^{nl} \frac{1}{\sigma_{\eta_k}} \phi\left[\frac{L_k - Z\gamma_k}{\sigma_{\eta_k}}\right] * \prod_{r=1}^{ni} \frac{1}{\sigma_{\nu_r}} \phi\left[\frac{I_r - L\alpha_r}{\sigma_{\nu_r}}\right] dL \quad (9)$$

where *nl* and *ni* indicate the number of latent variables and indicators respectively, ϕ represents the standard normal density function (pdf), Φ represents the cumulative density function (cdf), and *σ* represents the standard deviation of the random error of the latent variables and the indicators. The likelihood is obtained by integrating the density function over the latent variables *L*. However, the binary choice model contained in Equation 9 is too limited for demand models used for engineering design decision-making.

Our objective is to derive the likelihood function, LF for the latent variable model integrated with a multinomial logit choice model. The log likelihood function LLF for a multinomial choice model (e.g., logit, or probit), is shown in Equation 10 (Greene, 2003).

$$\mathrm{LLF}_{choice} = \sum_{i=1}^{n} \sum_{t=1}^{T_i} d_{it} \log \Pr(y_i = t) \quad (10)$$

where *n* represents the number of respondents, T_i the number of choice alternatives the respondent *i*, chooses from and $\Pr(y_i=t)$ represents the probability that choice alternative *t* is selected by respondent *i*, and $d_{ij} = 1$ if alternative *t* is chosen by respondent *i*, otherwise $d_{ij} = 0$. The probability of observing the observed choice set divided by the summed probability of observing all possible choice sets is shown in the detailed equation for grouped logit (i.e., multinomial logit), Equation 11.

$$\Pr(y_i \mid \textstyle\sum_{t=1}^{T_i} y_{it} = k_i) = \dfrac{\exp(\sum_{t=1}^{T_i} y_{it} Z_{it} \beta)}{\sum_{d_i \in R_i} \exp(\sum_{t=1}^{T_i} d_{it} Z_{it} \beta)} \quad (11)$$

The variable k_i indicates the number of alternatives chosen by respondent *i* and R_i represents the set of all possible choice sets d_i respondent *i* chooses from, i.e., all possible combinations with k_i ones and $(T_i$-$k_i)$ zeros. Thus, d_{it} is 1 or 0 conditional that $\sum_{t=1}^{T_i} d_{it} = k_i$ is satisfied, which ensures that the sum of choices equal to k_i. Further, ***Z***, which represents the key customer attributes ***A***, customer background ***S***, and price *P*, and the β-coefficients are used in the same fashion as in discrete choice modeling. Taking the log of Equation 11, results in the log likelihood function LLF of the multinomial logit choice model, which is presented in Equation 12.

$$\mathrm{LLF}_{choice} = \sum_{i=1}^{n} \left[\textstyle\sum_{t=1}^{T_i} y_{it} Z_{it} \beta - \log \sum_{d_i \in S_i} \exp(\sum_{t=1}^{T_i} d_{it} Z_{it} \beta) \right] \quad (12)$$

The modeling of multiple positive choices per choice set is facilitated using Equation 12. However, it is often unnecessary to consider multiple positive choices per choice set and the number of possible combinations S_i can be greatly reduced when assuming that exactly one alternative from the choice set is selected, i.e., setting k_i to 1, then Equation 12 can be simplified to Equation 13.

$$\mathrm{LLF}_{choice} = \sum_{i=1}^{n} \left[\textstyle\sum_{t=1}^{T_i} y_{it} Z_{it} \beta - \log \sum_{t=1}^{T_i} \exp(Z_{it} \beta) \right] \quad (13)$$

Assuming an additive function structure for the structural equations and the measurement equations, as shown in Equations 5, and 6, and also assuming normally and independently distributed random disturbances η and ν, orthogonal latent variables and further independence of the indicators, then the likelihood for the structural equations and the measurement equations can be formed as shown in Equations 14 and 15 (Ben-Akiva et al., 2002).

$$\mathrm{LF}_{structural} = \int_L \prod_{k=1}^{nl} \frac{1}{\sigma_{\eta_k}} \phi\left[\frac{L_k - Z\gamma_k}{\sigma_{\eta_k}}\right] dL \quad (14)$$

$$\mathrm{LF}_{measurement} = \int_L \prod_{r=1}^{ni} \frac{1}{\sigma_{\nu_r}} \phi\left[\frac{I_r - L\alpha_r}{\sigma_{\nu_r}}\right] dL \quad (15)$$

The likelihood function of the structural model, the measurement model of the latent variable model, and the log-likelihood function of the multinomial logit choice model can be combined to obtain the log-likelihood of the integrated model, presented in Equation 16.

$$\mathrm{LLF} = \sum_{i=1}^{n} \left[\textstyle\sum_{t=1}^{T_i} y_{it} Z_{it} \beta \prod_{r=1}^{ni} \frac{1}{\sigma_{\nu_r}} \phi\left[\frac{I_r - L\alpha_r}{\sigma_{\nu_r}}\right] - \log \sum_{t=1}^{T_i} e^{Z_{it}\beta} \right] \quad (16)$$

When fitting the integrated model, it is necessary to insure that the integrated latent variable model is identified, assuring that a latent variable model is not underdetermined (Loehlin, 1998). A good practice is to have at least four (4) indicators per latent variable. Further, one path for each latent variable should be fixed, i.e., constrained to a pre-selected value. Necessary, but not sufficient (for identification) is that the number of unknown parameters should be less than the number of unduplicated variances and covariances in the observed covariance matrix (i.e., $p(p+1)/2$ for *p* observed variables) of the data considered. Identification can be confirmed by using different starting points, which should yield the same estimated model parameters. Another method is to use the obtained latent variable model to generate synthetic data (indicator values and choice data). The latent variable model estimated using this synthetic data should be comparable within a reasonable margin of accuracy (Ben-Akiva et al., 2002).

An implementation of the integrated latent variable discrete choice demand modeling approach is demonstrated in the next section that employs the log likelihood function of Equation 16 as the objective function to determine the model coefficients. The Monte Carlo sampling method is used to assess the random disturbance η of the latent variables, which is assumed normally distributed and a simulated annealing algorithm is adopted for maximizing the log-likelihood function. This method is also tested and verified using mathematical problems (Wassenaar - Ph.D. thesis, 2003).

4 THE INTEGRATED LATENT VARIABLE MODEL FOR VEHICLE DEMAND ANALYSIS

An implementation of the latent variable modeling approach and the creation of the integrated latent variable multinomial logit model is demonstrated for demand modeling of passenger vehicles, with emphasis on studying the impact of engine design attributes (Wassenaar et al., 2003). First it is demonstrated how latent variable modeling can facilitate the understanding of the relationship between customer desires and product features by evaluating the engine design engineer's intent to increase customer satisfaction.

4.1 Latent Variable Model To Verify Designer's Intent

This illustrative example is limited to engine performance measurements and the customer's perception of engine performance for brevity. It should be noted that for making engine design decisions, on additional engine analysis model is needed to establish the relationship between the engine performance and the (detailed) design options, which is not the focus of this study. Engine performance measurements and the average customer ratings (both normalized) that measure the customer's perception of engine performance are presented in Table 1. The customer ratings are obtained from an existing customer survey of a vehicle model that is offered to the customer with a choice between a base engine and a more powerful engine, albeit for a different price. Thus both vehicle trims are the same except for the engine and price.

The data in Table 1 indicates that improving engine performance positively impacts the customer's perception of the engine's performance. However, the contribution and interaction of the different engine performance measurements to the customer's perception of the engine performance cannot be obtained from Table 1. For instance, an important design issue considered by an engine design engineer is the importance of horsepower versus torque. Evaluating customer ratings does not provide much insight here (a rating for torque is not available for this study). However, a latent variable model of engine performance can assess the contributions of the different engine performance measures to the customer's perceived engine performance and thus facilitate the engine design engineer in making important design decisions. A latent variable model of engine performance is presented in Figure 5. The indicators shown in Figure 5 are selected from the year 2000 J.D. Power VQS study (Vehicle Quality Survey) based on their relation to the engine performance measurements considered in this example. The measurement path between the latent variable and the customer rating for "performance during rapid acceleration" is constrained to one for identification. Other coefficients shown on the diagram are identified by fitting the latent variable model, which is based on the ratings of 373 survey respondents collected in the year 2000. Of these respondents, 300 opted for the base engine and 73 for the more powerful engine. The latent variable model shows that torque contributes more to the customer's perception of engine performance than horsepower (0.61 versus 1.38). This intuitively makes sense as car drivers predominately cruise at relatively low engine speed where torque is more prominent and seldom drive their vehicle at maximum horsepower. Additionally, it seems that both horsepower and torque are more important than fuel economy which concurs with the decreasing national fuel economy average in favor of more powerful vehicles like SUVs. Note that this study is based on year 2000 data, before the current steep increase in gas prices. The low residuals (ν) of the indicators show that the model predicts the indicators well given that the observed variance for the indicators is close to 1.3. This example illustrates how latent variables can be used to assess how product features relate to customer desires, helping the design engineer to design a product that customers want.

Table 1. Engine performance and customer ratings

Eng.id	Engine performance				Customer ratings for Engine				
	Torque	Power	Fuel Econ	Price	Power	Fuel Econ	Eng. Sound	Engine	Vehicle
Eng.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Eng.2	1.12	1.32	0.98	1.07	1.05	1.02	1.01	1.04	1.01

Once the latent variable model is fitted, the latent variable model can be used to predict the indicator levels, i.e., the customer ratings as a function of customer attributes A that describe the vehicle and the customer background S. Though predicting customer ratings is not considered for this case study, customer ratings are used by J.D. Power to determine the APEAL score (Automotive Performance, Execution And Layout), which is an important measure for vehicle manufacturers. In the next section it is shown how a latent variable model can be integrated into a discrete choice model in order to strengthen the relationship between the engineering design and market demand.

4.2 Integration of Latent Variable Model Into Discrete Choice Demand Model

The explanatory power of demand models and the causal relationship between (engine) design and market demand can be enhanced by integrating a latent variable model and a discrete choice model, demonstrated in the following example which is adapted from Wassenaar and Chen (2003). The choice alternatives are the same 12 trims of the 7 vehicle models of the mid-size car market segment. The integrated latent variable model is implemented using the five steps laid out in Section 3.

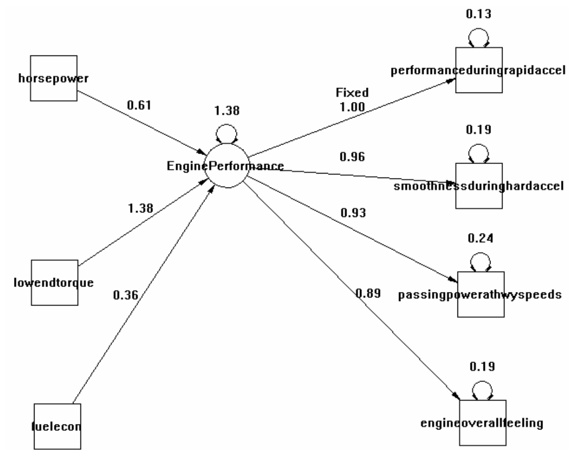

Figure 5. Latent variable example for performance

Step 1. Determine what Attributes to Consider

Details related to deciding what attributes to consider can be found in Wassenaar and Chen (2003). A set of ten customer attributes is used for this example (Table 2) that relate to the price, reliability, engine performance, fuel economy, and the vehicle's size such that a general description of the whole vehicle is obtained. It is necessary to capture a complete description of a vehicle to avoid biased coefficient estimates.

Step 2. Identify the Latent Variables

This step consists of determining the number of latent variables that will be considered in the integrated model, and subsequently identifying the latent variables themselves. In literature, there are no specific guidelines of what and how latent variables should be identified. Intuitively, the number of latent variables can be thought to depend on the number of top-level customer desires (Wassenaar and Chen, 2003). The key customer attributes $\boldsymbol{A}$ used for this case study concern mainly financial, performance, and dimensional properties that can be classified into three categories: cost, performance, and comfort. These can be seen as three possible latent variables.

Loehlin (1998) suggested an approach to determine how many and what latent variables to include in the latent variable model by analyzing the data using factor analysis. Factor analysis often facilitates a reduction of the number of variables of a data set by transferring the data set to a new set of variables, called factors, that are a linear combination of the existing variables. Factor analysis (Hair et al., 1998) aims to identify underlying linear constructs (i.e., components) that reproduce the correlation matrix of the observed data by capturing as much of the variability found in the observed data as possible. In Equation 17 it is shown how the new variables (F_i) are a combination of the variables A of the existing data set.

$$F_i = \delta_{i1}A_1 + \delta_{i2}A_2 + \delta_{i3}A_3 + \ldots + \delta_{in}A_n \qquad (17)$$

The factor analysis approach is iterative in that it determines the δ-coefficients (called factor loadings) by first finding the factor F_1 that explains the most variance of the data, then the factor F_2 that explains the second largest amount of variance, etc. The factor loadings represent the correlation between the factor and the variable. Hence, the square of a coefficient δ_{in} indicates the amount of variance of variable A_n explained by factor F_i. The eigenvalues of the factors can be thought of as a measure of how much variance of the data each component explains. The larger the factor's eigenvalue, the more variance is explained by that component. The complete set of factors explains 100% of the variance found in the observed data. However, the goal is to obtain a reduced set of factors that captures the data variance well. A rule of thumb, known as the Kaiser-Guttman rule (Loehlin, 1998), is to only consider the factors with an eigenvalue larger than one (1). Another approach is to graph the eigenvalues as shown in Figure 6. This approach, known as Scree plot analysis (Cattell, 1966), considers the data as a mountain with debris at its foot. Because the mountain contains the valuable components of the data (high eigenvalues), these components are retained, and the debris at the foot is considered rubble and therefore omitted. Given the Scree plot, the earlier assumption of using three latent variables is quite reasonable although the eigenvalue of the third component is just slightly larger than one. Having determined the number of latent variables, the next task is to identify suitable indicators for each latent variable, which is detailed in Step 3.

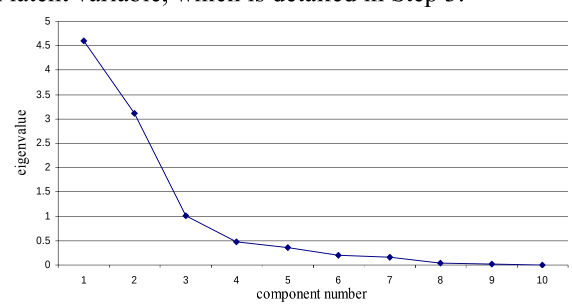

Figure 6. Customer attribute Scree plot

Step 3. Identify Indicators for Each Latent Variable

In this step the indicators are selected to measure each latent variable, i.e., cost, performance, and comfort. It would be ideal to hypothesize indicators that can measure each latent variable and then to collect data through a survey. However, in this example existing survey data is used for the indicators instead of collecting new data, which necessitates an assumption explained below.

Indicator Data for the Integrated Latent Variable Model

The items assessed in the VQS study are considered as possible indicators. Like choice data, the indicator values should be employed at the individual customer level. The survey used in the VQS study asks respondents (existing car buyers) questions about how they rate their vehicle's qualities. The VQS questionnaire considers eight aspects of the car at vehicle level, e.g., engine/transmission, cockpit, ride and handling, HVAC (heating, ventilating, air conditioning), comfort, sound system, seats, and vehicle styling. Each of these eight aspects at vehicle level is assessed using multiple questions concerning details of that aspect, which survey respondents rate on a linear scale ranging 1 – 10. For engine/transmission these are: sound while idling, sound at full throttle acceleration, performance during rapid acceleration, passing power at highway speeds, range between fuel stops, overall transmission performance, etc. These ratings provided by the customers are used as indicator values in the latent variable model. The linear rating scale used in the VQS questionnaire does not require a tradeoff among attributes, thus it seems reasonable to assume that the VQS consumer ratings are free of interaction effects. However, the brand image and halo-effect (customer is excited about the new purchase) may taint the ratings customers give to the vehicle they purchased. This effect is called justification bias, an approach to account for this bias is presented in Ben Akiva (2003).

Given the limited availability of indicator data related to cost, the latent variable is dropped for cost and two latent variables, performance and comfort, are retained. Two latent variables seems reasonable given that the eigenvalue of the third factor is close to one (see Step 2). The indicators and customer attributes identified for each latent variable are shown in Table 2. Each customer attribute and indicator is considered for its relevance to each latent variable.

Table 2. Latent variable attributes and indicators

Latent variable ***L***	Attribute ***A***	Indicator ***I***
L_1 performance	A_4 engine horsepower A_5 engine torque A_6 fuel economy A_{10} vehicle mass	I_1 engine performance during rapid acceleration I_2 engine fuel economy I_3 engine overall feeling I_9 vehicle overall feeling
L_2 comfort	A_7 front head room A_8 front leg room A_9 trunk space A_{10} vehicle mass	I_4 comfort-front leg room I_5 comfort-front head room I_6 seats overall feeling I_7 comfort-overall feeling I_8 comfort-cargo space I_9 vehicle overall feeling

The complete integrated latent variable model is shown in Figure 7, where A_1, A_2, and A_3 represent MSRP-price, VDI (dependability index), and the binary variable USA/imp.

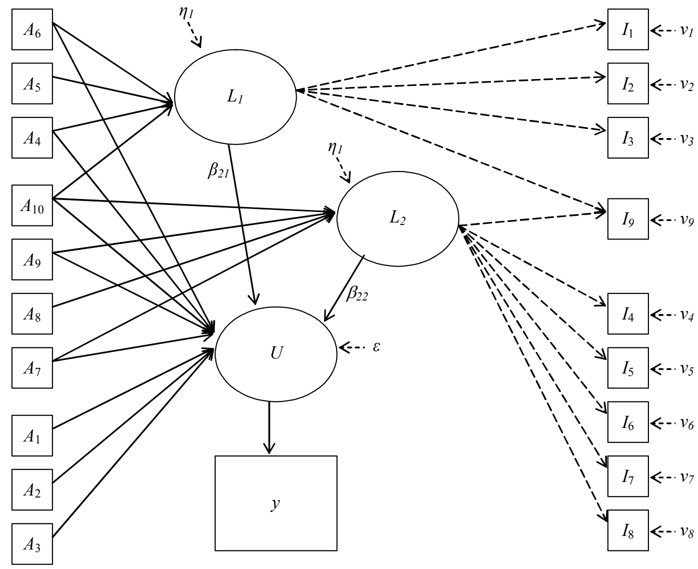

Figure 7. Complete integrated latent variable model for vehicle engine design example

An important issue of latent variable modeling is "identification," discussed in Section 3. In this example, each latent variable is measured by at least four indicators. Further, the number of data entries of the observed data covariance matrix exceeds the number of unknown parameters (i.e., γ, α, and β coefficients and the variances of the indicators, ν). A latent variable model is called over-identified when the number of unknown parameters is exceeded by the number of (observed) variances and covariances of the covariance matrix. Over-identification is a good thing, but does not guarantee identification of the latent variable model. That is, identification should be checked, using the procedure presented in Section 3. Additionally, one measurement equation per latent variables must be fixed to one for identification. Thus far, the rules for identification are satisfied. The next step is to form the structural equations and the measurement equations.

Step 4. Determine the Structural Equations and the Measurement Equations

The forms of the structural equations and the measurement equations are determined. In our study, the linear form of equation is used.

Step 5. Fitting the Integrated Latent Variable Model

In this step the structural equations and measurement equations are fitted using a MatLab program. The maximum likelihood function of Equation 16 serves as the objective function for optimization using simulated annealing. The observed and predicted market shares are presented in Table 3 and measurements that indicate the model fit such as the r-squared values of the predicted market share and observations at individual level and the Bayesian Information Criterion

(BIC) are presented in Table 4. While the results indicate a satisfactory fit, i.e., a good match between observed and estimated market shares, it should be noted that like any other curve (model) fittings, the accuracy of a model needs to be verified for its predictive capability using new data.

A related issue is that one may be interested in the actual improvement realized by incorporating the latent variables compared to the model without latent variables developed in our earlier study (Wassenaar and Chen, 2003). It should be pointed out that such comparison is not meaningful because when constructed properly, both models can provide perfect fits for the data used. However, whether a model is better or worse largely depends on its predictive capability, which needs to be verified through model validation. The model validation for a demand model also involves other aspects of demand analysis, such as considering the impact of economic conditions. Model validation is beyond the scope of this study.

Our focus is on illustrating the potential of the proposed method. Specifically, the integrated latent variable model can be used to not only assess the impact of design decisions on market demand but also to evaluate *how design changes impact the customer's perception of the product*, in particular performance and comfort. The integrated latent variable model enables the design engineer to evaluate the impact of engineering decisions on customers' perception of the product, i.e., from sporty (performance) to luxury (comfort). Such capability is valuable for product differentiation (ensuring that products are perceived as being different by customers) and the avoidance or minimization of market share cannibalism. This example clearly demonstrates how the integrated latent variable approach can be implemented, resulting in potential benefits related to design decision-making by explicitly considering the customer's perception.

Table 3. Example problem market share estimates

Vehicle id.	Observed market share	Estimated market share
1	0.100	0.1059
2	0.075	0.0723
3	0.125	0.1303
4	0.075	0.0927
5	0.100	0.1053
6	0.075	0.0811
7	0.075	0.0614
8	0.075	0.0567
9	0.025	0.0492
10	0.175	0.1520
11	0.075	0.0556
12	0.025	0.0376

Table 4. Integrated latent model result summary

	Integrated latent variable model
LLF value	-6908.60
r^2_MS	0.8538
r^2_observations	0.0167
BIC	252.91

5 CONCLUSION

In this work an extension related to the discrete choice analysis approach to demand modeling with latent variables to facilitate the consideration of the customer's perception is developed. The integrated latent variable discrete choice model for binomial choice is extended for multinomial choice and a method is developed to estimate the model coefficients. In a classical discrete choice model the perception is thought to be captured by the random disturbance ε (epsilon) of the customer utility function, but this is a simplistic approach. The irregular relationship between attribute value and customer perception as reflected in the fuel economy example implies that the ε may not be zero as assumed in logit choice models. The inclusion of alternative specific constants in the customer utility function may help but this approach may lead to over-specification of the choice model (Bierlaire *et al*., 1997). It is expected that explicitly considering the customer's perception in the demand model helps explain part of the random disturbance ε of the customer utility and so enhances the demand model's predictive capability. The proposed latent variable approach better captures psychological factors that affect the purchase behavior of customers than a discrete choice analysis approach by itself, which facilitates more accurate choice predictions.

An additional advantage of this approach is that the latent variable approach separates attributes into groups which can reduce collinearity issues. Thus, this approach can facilitate the consideration of more key customer attributes in the (integrated) demand model, enabling a stronger linkage with engineering design analysis models, which facilitates better guidance of engineering design decision-making. Another benefit of latent variable analysis is the improved understanding of how consumers' desires relate to product features as shown in the example in Section 4 (Figure 5). As such, a latent variable model can predict the impact of design decisions on the customer's perceived product image. Questions such as: does the customer notice and appreciate the improvement; and, is the improvement worth the additional cost, can be answered using latent variable models.

An efficient computational approach to estimate integrated latent variable demand models is planned for future development. Validating demand models is another challenging research topic as many other factors besides the customer attributes contribute to the product demand. Achieving a balance between the comprehensiveness of a model versus the loss of predictive accuracy when including many product attributes as input is another challenging topic for demand modeling of complex systems such as a vehicle.

ACKNOWLEDGMENTS

The specialists at J.D. Power and Associates and the Ford Motor Company are gratefully thanked for their thoughtful contributions and their efforts to gather data for the vehicle demand model. The support from the NSF grant DMI0335880 is also acknowledged.

REFERENCES

Antonsson, E.K., Zhang, Y., and Martinoli, "A. Evolving Engineering Design Tradeoffs", *ASME 15th International Conference on Design Theory and Methodology*, DETC2003/DTM-48676, Chicago IL., Sept. 2-6, 2003..

Bentler, P.M., Multivariate Analysis with Latent Variables, *Annual Review of Psychology 31*: pp. 419-456, 1980.

Besharati, B., Azarm, S., and Farhang-Mehr, A., "A Customer-Based Expected Utility for Product Design Selection", Proceedings of the ASME DesignEngineering Technical Conference, Montreal, Canada, 2002.

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

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., and Polydoropoulou A., "Integration of Choice and Latent Variable Models", Chapter 21 of *In Perpetual Motion: Travel Behaviour Research Opportunities and Application Challenges,* ed. by Mahmassani, E., Elsevier Science, 2002.

Bierlaire, M., Lotan, T., and Toint, P., "On the Overspecification of Multinomial and Nested Logit Models Due to Alternative Specific Constants", *Transportation Science*, Vol. 31, no. 4, November 1997.

Cattell, R.B., The scree test for the number of factors, *Multivariate Behavioral Research*, 1, p245-276, 1966.

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

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

Everitt, B.S., An Introduction to Latent Variable Models, *Monographs on Statistical and Applied Probability*, Chapman and Hall, 1984.

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

Greene, W.H., *Econometric Analysis*, second edition, Macmillan Publishing Company, New York, 2003.

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

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

Joreskog, K.G., A General Method for Estimating a Linear Structural Equation System, *Structural Models in the Social Sciences*, A.S. Goldberger and O.D. Duncan, Eds. Academic Press, New York, 1973.

Keesling, J.W., Maximum Likelihood Approaches to Causal Analysis, Ph.D. Thesis, *University of Chicago*, 1972.

Koppelman, F. and Hauser, J., Destination Choice for Non-Grocery-Shopping Trips, *Transportation Research Record 673*: pp. 157-165, 1979.

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

Loehlin, J.C., *Latent Variable Models, an introduction to factor, path, and structural analysis*, 3rd ed., Mahwah, NJ: L. Erlbaum Associates, 1998.

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

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.

Michalek, J. J., Feinberg F.M., and Papalambros, P. Y., "Linking marketing and engineering product design decisions via analytical target cascading," *To appear in Journal of Product Innovation Management.*

Olewnik, A.T., Lewis, K.E., On Validating Design Decision Methodologies, *Proceedings of the 2003 ASME Design Engineering Technical Conferences*, Chicago, IL., September 2-6, 2003.

Prashker, J.A., Mode Choice Models with Perceived Reliability Measures, *Transportation Engineering Journal* 105 TE3: pp. 251-262, 1979.

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

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.

Wassenaar, H.J. and Chen, W., Enhancing Discrete Choice Demand Modeling for Decision-Based Design, DETC-DTM48683, *Proceedings of the 2003 ASME Design Engineering Technical Conferences*, Chicago, IL., September 2-6, 2003.

Wassenaar, H.J. and Chen, W., "An Approach to Decision-Based Design", *ASME Journal of Mechanical Design, Vol. 125*, 490-497, 2003.

Wassenaar, H.J., Chen, W., Cheng, J., Sudjianto, A., "Demand Analysis for Decision-Based Design of Vehicle Engine", accepted by the *2004 SAE World Congress*, Paper 04M-133, Detroit, MI., March 2004.

Wiley, D.E., The Identification Problem for Structural Equation Models with Unmeasured Variables, *Structural Models in the Social Sciences*. A.S. Goldberger and O.D. Duncan, Eds. Academic Press, New York, 1973.