NORTHWESTERN UNIVERSITY

Configuring Engineering Systems Considering Consumer Heterogeneity

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Mechanical Engineering

By

Christopher Hoyle

EVANSTON, ILLINOIS

December 2009

Abstract

Configuring Engineering Systems Considering Consumer Heterogeneity

Christopher Hoyle

This dissertation is motivated by the need to develop methods which connect the engineering and marketing domains to enable identification of the preferred engineering system configuration, considering the real complexities in engineering system design and the heterogeneity of consumer preferences for such systems. The research includes a design process tool, an experimental design approach for human appraisal experiments, a multivariate statistical analysis methodology for human appraisal data, and finally an integrated Bayesian hierarchical choice modeling method which rigorously considers consumer heterogeneity and the nature of complex system design. This research primarily uses an automotive vehicle occupant package design as a motivating example, to both illustrate the issues in system design and demonstrate the features of the proposed design approach. The research can be divided into four primary contributions.

A new process tool called *Product Attribute Function Deployment* (PAFD) is introduced as a decision-theoretic, enterprise-level process tool to guide engineering design. The PAFD method is a model-based approach built upon established methods in engineering, marketing, and decision analysis to eliminate the need for user ratings and rankings of performance, priority, and attribute-coupling used in current process tools.

To collect data necessary to support preference modeling, an algorithmic *design of human appraisal experiments method* is developed to identify the optimal human appraisal experiment for a given set of requirements. The advantages of this approach over competing approaches for minimizing the number of appraisal experiments and model-building efficiency are clearly demonstrated.

An issue with human appraisal experiments is that the heterogeneity of the experimental respondents contributes to the response, and this heterogeneity must be understood to separate the influence of design factors from that of human factors. Multivariate statistical techniques are utilized to create *a human appraisal analysis methodology* to understand heterogeneity and preprocess the human appraisal data to enable preference modeling.

The *Integrated Bayesian Hierarchical Choice Model* (IBHCM) framework provides a unified choice modeling approach for complex system design. It utilizes multiple model levels to create a link between qualitative attributes considered by consumers when selecting a product and quantitative attributes used for engineering design.

Acknowledgements

I would like to express my sincere thanks to my PhD advisor, Professor Wei Chen, for providing me with her guidance, extensive support, inspiration, and encouragement throughout my Ph.D. studies. She has been an outstanding advisor and excellent teacher throughout my studies at Northwestern University. I am appreciative that she funded me given that I had been away from academia for 10 years. I am very thankful to my committee members: Professor Bruce Ankenman, Professor Frank Koppelman, and Dr. Nanxin Wang (Ford Motor Co.) for their suggestions and contributions to my research. Collaborating and co-authoring with them on my research projects were valuable experiences. I also owe special gratitude to Gianna Gomez-Levi and Richard Jakacki of Ford Motor Company for their assistance and use of the Programmable Vehicle Model at Ford, which made Chapters 4 and 5 of this dissertation possible, as well as their thorough review of the findings in this work. I would like to acknowledge Professor Irem Tumer of Oregon State University, formerly of NASA Ames Research Center for her guidance and collaboration on research projects outside of this dissertation. I would also like to acknowledge my master's thesis advisor, Professor Karthik Ramani of Purdue University for supporting my return to school and preparing me for research work during my master's degree. I would also like to acknowledge the faculty and staff of the Industrial Engineering Department at École Centrale Paris, particularly Professor Bernard Yannou, who hosted us for the summer of 2008 and took an interest in our research, as well as tolerating our lack of proficiency in French for the summer.

I am grateful to my great colleagues at the Integrated DEsign Automation Laboratory (IDEAL). Many thanks go to current members Xiaolei Yin, Shikui Chen, Lin He, Fenfen Xiong, Steven Greene, and Paul Arendt as well as previous members Huibin Liu, Deepak Dileep Kumar, Sanghoon Lee and Ying Xiong for their valuable discussions, advice, and help getting me started in school again. I really enjoyed the friendly and collaborative environment in IDEAL.

Finally, I would like to thank my friends and family for their continued support during my time in school again. Everyone has been quite interested in my progress at school, and available for discussion about various topics, research-related or otherwise.

This research was supported by the US National Science Foundation (NSF) (CMMI—0700585) and Ford University Research program. The support is gratefully acknowledged. I would also like to thank all the staff in the Department of Mechanical Engineering at Northwestern University for always being very friendly and helpful.

Glossary

Decision-Based Design (DBD): An approach to engineering design that recognizes the substantial role that decisions play in design, largely characterized by ambiguity, uncertainty, risk, and multiple trade-offs.

Discrete Choice Analysis (DCA): A statistical modeling technique that describes choices made by people among a set of mutually exclusive and collectively exhaustive alternatives. Aggregation of individual choice probabilities allows for demand estimation for a given alternative.

Ordered Logit (OL): A regression modeling technique specifically for modeling ordinal dependent variables, such as ratings.

Hierarchical Choice Model (HCM): A multilevel model used to describe choices made by people for a set of alternatives characterized as complex systems. The model is characterized by a DCA model at the top level, and a system of DCA or OL models at other levels to link consumer choices to engineering design attributes.

Customer-Desired Attributes (A): Attributes of a product or system which influence a consumer's choice or evaluation of the product or system, such as comfort, roominess, or exterior styling of an automobile.

Engineering Attributes (E): Attributes of a product or system used in engineering analysis and decision-making, such as horsepower, occupant package headroom, or fuel economy of an automobile.

Design Attributes (X): Specific attributes of a product or system which can be directly controlled by a designer to define an engineering attribute, such as a material type, dimension, or shape of an automotive component.

Demographic (or Human) Attributes (S): Attributes of the consumer including socio-economic (e.g. income), anthropomorphic (e.g. height), purchase history (e.g. Ford Focus) and product of system intended usage (e.g. commuting to work).

Model Attributes (Z): The set of all customer-desired **A** or engineering attributes **E,** and demographic attributes **S** included in the choice or rating model, including interactions among the model terms and high order terms.

Ratings (R): A method for a consumer to express his/her opinion of a product or system using an ordinal scale. Popular ordinal scales are 1-5, 1-7, or 1-10.

Programmable Vehicle Model (PVM): A computer-controlled, parametric vehicle hardware model capable of simulating a wide range of vehicles in a short amount of time for human appraisal experiments.

List of Abbreviations

AIC	Akaike Information Criterion
AIO	All in One
ASC	Alternative Specific Constants
ASV	Alternative Specific Variables
DBD	Decision-Based Design
DCA	Discrete Choice Analysis
det	Matrix determinant
GLS	Generalized Least Squares (Regression)
IBHCM	Integrated Bayesian Hierarchical Choice Model
inv	Matrix inverse
LR	Linear Regression
MNL	Multinomial logit
MXL	Mixed Logit
NL	Nested Logit
OLS	Ordinary Least Squares (Regression)
PAFD	Product Attribute Function Deployment
QFD	Quality Function Deployment
RE-OL	Random-Effects Ordered Logit
RP	Revealed preference
SP	Stated preference
SgRP	Seating Reference Point

Nomenclature

α	Pair-wise correlation coefficient for multinomial covariance matrix
β	Discrete Choice or Ordered Logit model coefficient for customer-desired attributes in a customer's utility function
B	Number of configurations given to a single respondent
bs	Rating Bias
C	Total Product Cost
$C.S.$	Choice share (aggregate product choice probability)
$D(t)$	Market Size (aggregate product type demand)
$\mathbf{D}_n$	Derivative of $\boldsymbol{\pi}_n=(\pi_{ip},\ldots,\pi_{BP})$
df	Degrees of Freedom
Ds	Index of dissimilarity
$E(U)$	Expected value of enterprise utility
ε_{ni}	Random disturbance of customer choice utility of alternative i by customer n
f	PDF of the logistic distribution
F	CDF of the logistic distribution
Fe	Design features
$\mathbf{f}(\mathbf{x})$	Extended Design Point, including intercept/cut points, interaction, and higher ordered terms than $\mathbf{x}$
$\mathbf{F}$	Extended Design Matrix, composed of $\mathbf{f}(\mathbf{x})$
G	Candidate set of design points
η_k	Dataset specific error, separate from ε_{ni}
i	A configuration or alternative
J	Set of competitive alternatives in the choice set

$\mathbf{k}$, k_p	Ordered Logit cut points
L (LL)	Likelihood (Log-Likelihood)
L^2	Likelihood Ratio Chi-Squared test statistic
M	Number of Configurations in a complete experimental design
$\mathbf{M}$	Fisher Information Matrix of an experimental design
$\frac{1}{\mu}$	Scale parameter that characterizes a Gumbel distribution. Directly related to the variance of the Gumbel distributed random variable
$\mathbf{Mf}$	Manufacturing process attributes
n	A block or respondent
Π	Profit
P	Product price
p	Ordered ratings categories
$\mathbf{P}$	Working Correlation Matrix
π_{nip}	Probability of Rating p for respondent n and configuration i
Q	Product demand
ρ	Pair-wise ratings correlation coefficient
ρ_0^2	Model fit statistic for ordered logit/probit model
s	A single case or observation
su	Rating scale usage
σ_u	Variance at the respondent level
σ_ε	Variance at the observation level
t	Time interval for which demand/choice share is to be predicted
T	Number of tries conducted in the algorithm
Tn	A training dataset

U	Enterprise utility, in units of *utils*
u_{ni}	True customer choice utility of alternative i by customer n
V	Selection criterion used by the enterprise (e.g. profit, choice share, revenues, etc.)
$\mathbf{V}$	Asymptotic variance-covariance matrix
W_{ni}	Observed part of the customer choice utility of alternative i by customer n
$\mathbf{x}$	Design Point product and human factors. A sub-set of $\mathbf{f(x)}$
$\mathbf{Y}$	Exogenous variables (represent sources of uncertainty in market)

Table of Contents

List of Tables

List of Figures

Chapter 1

PROBLEM DESCRIPTION AND RESEARCH OBJECTIVES

The research in this dissertation is motivated by the need to develop methods which connect the engineering and marketing domains to enable identification of the preferred configuration of an engineering system, considering the real complexities in engineering system design and the heterogeneity of consumer preferences for such systems. Configuring an engineering system requires multiple decisions to be made, specifically selection of a preferred system concept, deciding the features to offer in the system, and finally setting performance targets for the system. To make rigorous decisions for an engineering system, it will be shown in this research that it is necessary to consider factors outside the traditional engineering domain.

Traditional engineering design is conducted primarily with an engineering-centric viewpoint, in which the objective is to achieve the best performance given the budget available (monetary, human resources, etc.). In general, it has been noted in a variety of contexts (Clausing and Hauser, 1988; Krishnan and Ulrich, 2001; Ullman, 2002) that each of the major functional domains within a firm, or **enterprise**, such as engineering, marketing, production, and management, generally seeks to optimize a domain-specific objective, with limited input from the other functional domains. An example of the traditional product development decision making process in the marketing and engineering domains is shown in Figure 1.1. Such a disconnected decision process cannot be assured to make optimal decisions for an engineering system, most importantly because the engineering-centric approach does not consider consumer demand for the designs considered, whereas the marketing-centric approach does not consider the intricacies of engineering attribute coupling, and the resulting influence upon cost, for a product or system design.

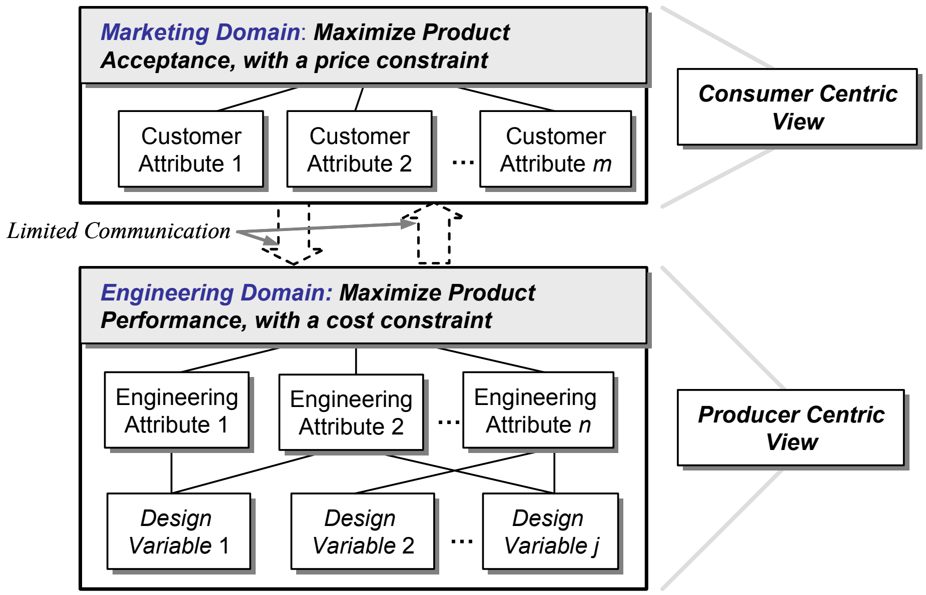


Figure 1.1: Disconnected Decision Processes

The need to consider potential consumer demand, together with cost and performance, when designing an engineering system is necessary to estimate the potential **profit** for the designed system, to determine the benefit of a given design to an enterprise. As will be shown in this research, an estimation of **demand** as a function of product attributes must explicitly consider the heterogeneity of the consumers and the market in which the product will compete, as illustrated in Figure 1.2, as well as sources of uncertainty to make product decisions.

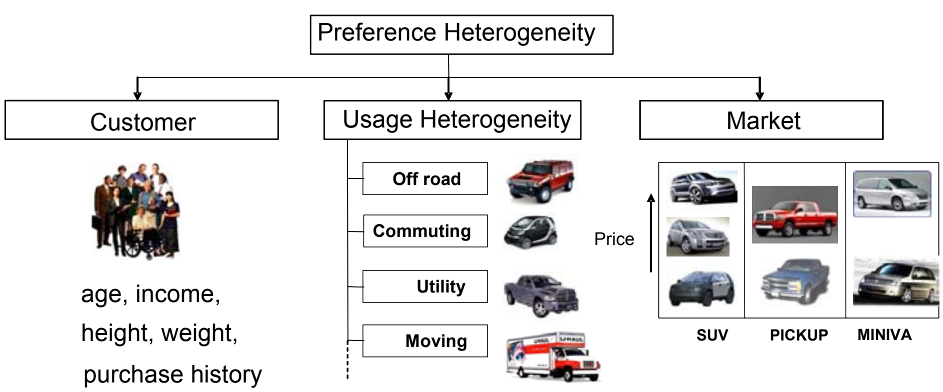


Figure 1.2: Heterogeneity of the Consumers and the Market

1.1 RESEARCH MOTIVATION AND CHALLENGES

Creating a connected decision process, primarily between marketing and engineering, has been a topic of research over the last few decades (Krishnan and Ulrich, 2001). Clausing and Hauser (1988) introduced the largely qualitative House of Quality (HoQ) methodology, based upon the **Quality Function Deployment** (QFD) methodology developed in Japan in the late 1960's. QFD is based upon the assumption that customer preferences can be aggregated and represented by group importance rankings and ratings, which is a potentially faulty assumption as will be demonstrated in this research. Recent efforts have used quantitative demand modeling approaches to estimate consumer demand for engineering system designs. Cook (1997) introduced a linear demand model derived from Taylor series expansion using product value and price to estimate demand, assuming aggregate consumer preferences. Li and Azarm (2000) utilized paired-comparison conjoint analysis (Green and Srinivasan, 1978) and estimated a deterministic linear part-worth utility model to estimate demand among the survey respondents for a given product. Alternatively, probabilistic choice modeling approaches using **Discrete Choice Analysis** (DCA) (Ben-Akiva and Lerman, 1985) to estimate demand have been utilized. DCA is a flexible approach which can model choice using a utility function composed of observed *product* and *consumer* level attributes, and can be estimated using survey or actual choice data, or a combination of both. Further, a "mixed" formulation of the model can be used to capture the distribution of unobserved, or random, preference heterogeneity. Using DCA to estimate demand entails estimating **choice probability** for a given design alternative over a sample population, and aggregating choice probability for a given design alternative to estimate its **choice share,** and ultimately its demand. Wassenaar et al. (2003; 2005; 2006) utilized DCA to model demand for an engineering system, demonstrating the method using a selection of quantitative product and consumer attributes for a consumer product and an automotive system. The approach was extended to include consumer perceptual preferences through the introduction of latent variable modeling (Wassenaar et al., 2004). Michalek et al. (2005) have considered random consumer heterogeneity only (i.e. the distribution of attribute preferences in a given data set) to enable the use of a DCA choice model.

While the previous work has laid the foundation for incorporating demand modeling in engineering design, there are several issues which must be resolved to enable the estimation of demand models for a general engineering system. The issues are summarized as follows:

- A systematic approach does not exist to model the effect of consumer heterogeneity upon demand. Previous approaches have accounted for heterogeneity of consumer preference in different manners: some have assumed that an "average" preference exists for the whole market, while others have assumed that the market could be segmented into groups in which preference is assumed to be homogeneous (i.e. latent class models (Train, 2003)). In the DCA methodology, both **systematic** and **random** heterogeneity have been modeled, but primarily as dictated by the form of model chosen and not based upon the nature of the problem.
- Except for the qualitative and potentially faulty QFD process, a **design process tool** for implementing the demand modeling approaches described above in a real design environment does not exist.
- Previous approaches have not adequately addressed the issue that customer-desired attributes used by a consumer may be qualitative in nature, and may be best expressed by an ordinal **rating** or **ranking** for the attribute, as opposed to a quantitative measure in a choice model.
- A comprehensive method for designing, conducting, and analyzing **human appraisal experiments** for use in guiding the engineering design process does not exist.

- The approaches do not adequately consider the nature of a complex system, in which a **hierarchy** of sub-systems exists between the top-level consumer choice attributes and the design-level engineering attributes.
- The approaches presented have used a single data source collected at a single time, with no framework to combine data from **different sources** and **different times** throughout the design process.
- Major sources of **uncertainty** in the demand model have not been adequately quantified and included in the subsequent decision process.

Given the issues in engineering design, the objective of this dissertation is to develop a general design methodology for complex engineering systems which considers the effect of consumer heterogeneity. Specifically, the research tasks are: 1) to create a general design process tool, similar in format to QFD, but rooted in more rigorous design decision-making principles to *quantitatively* bridge the gap between engineering and marketing; 2) to formulate a design of experiments method specifically for human appraisals; 3) to develop a multivariate statistical analysis methodology to understand human appraisal experiments; and 4) to create a general hierarchical choice modeling framework to support target-setting for complex system design, which accounts for the hierarchical nature of complex system design, incorporates heterogeneity at all model levels, and quantifies model uncertainty.

1.2 ENTERPRISE-DRIVEN DESIGN APPROACH TO CONFIGURING ENGINEERING SYSTEMS

The enterprise-driven approach to system configuration proposed in this dissertation and illustrated in Figure 1.3 is a comprehensive process to address the issues enumerated in Section 1.1. Specifically, methods are provided for *selecting* a preferred design concept, *assessing* consumer preferences for various system, sub-system, and component attributes, *utilizing* data collected in human appraisal experiments, and setting *target levels* of performance for a preferred system concept. The methods proposed in this research are based upon the principles of enterprise-driven Decision-based Design (DBD) to make engineering system configuration decisions. This proposed design methodology begins at the conceptual design phase in which a number of design concepts are brainstormed by a design team and a preferred design concept is selected for further development. Following selection of a preferred concept, performance targets must be set to define the preferred configuration for the selected system concept. The process of configuring the system starts with an understanding of heterogeneous consumer preferences for system, sub-system, and component level design features. These preferences are elicited using designed *human appraisal experiments*. These *data are analyzed and processed* to understand consumer heterogeneity and to structure the data in a format which supports efficient modeling of consumer preferences. With structured data, preference models are created which link consumer preferences to engineering attributes *using the Bayesian Hierarchical Choice Model*, which together with cost models and enterprise-level objectives, enable identification of target engineering performance levels to be selected which meet the needs of both the consumer and producer.

To realize the enterprise-driven approach to system configuration, research is required in four core areas: the Product Attribute Function Deployment Method, the Human Appraisal Experimental Design Method, Multivariate Statistical Data Analysis and Processing Techniques for Human Appraisals, and the Integrated Bayesian Hierarchical Choice Model. Research in these four core elements forms the focus of this dissertation; each research task is described in more detail in the following paragraphs.

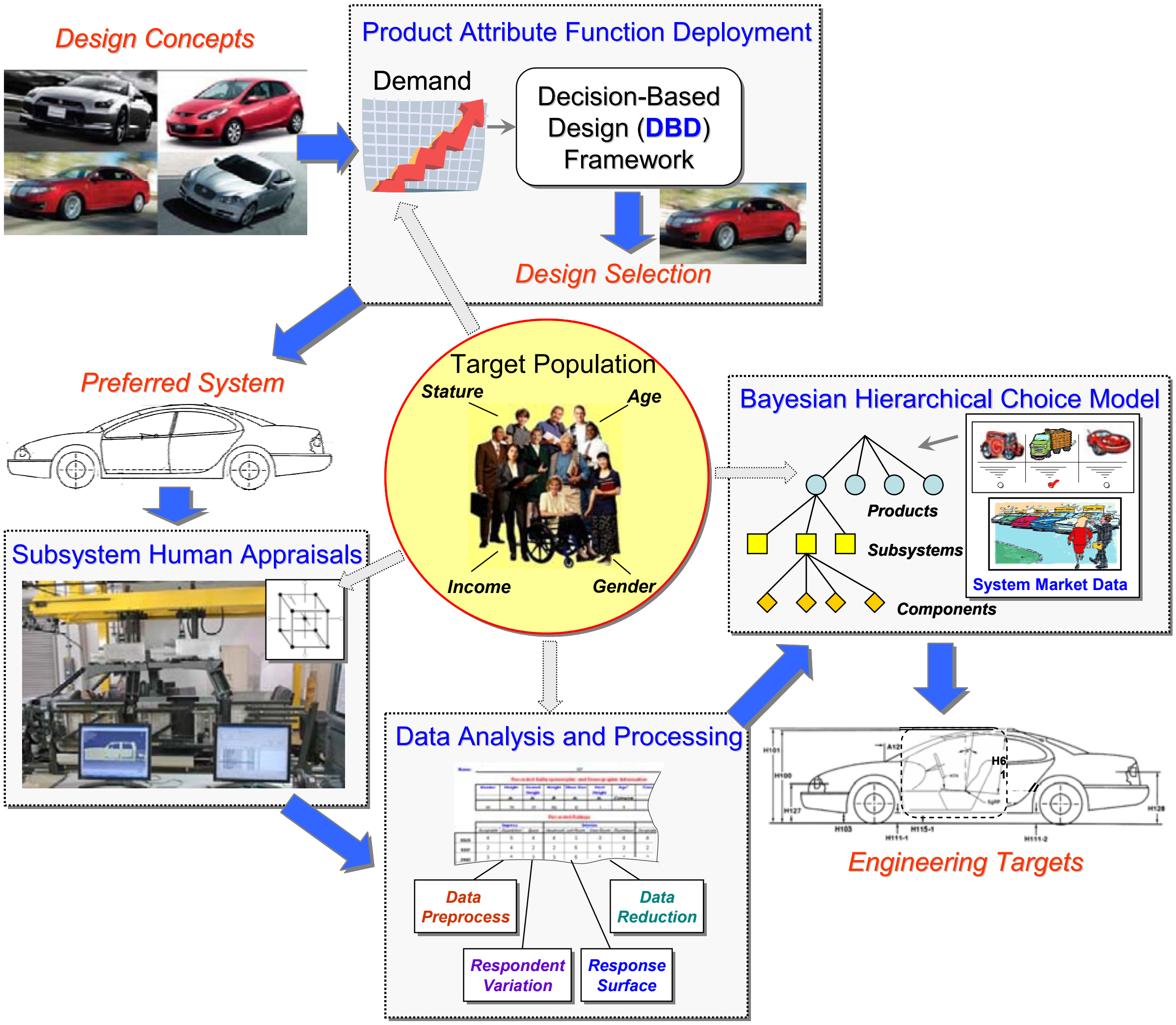


Figure 1.3: Enterprise-Driven Approach to System Configuration

Research Task 1—Product Attribute Function Deployment: As noted, design decisions for an engineering system require consideration of engineering performance and cost, as well as market acceptance to ensure the resulting design will be profitable and benefit the enterprise. The current methods outlined previously for bridging this gap are either the qualitative, and potentially flawed QFD method, or the quantitative optimization frameworks, using analytical engineering and market demand models, but lacking a methodology for implementation in a product development setting. In engineering product development, design process tools are needed to guide the development process in a systematic way, with a clear flow of activities. The PAFD method is provided as a *design process tool, rooted in rigorous decision-making principles, to bridge the gap between engineering and marketing to guide product development activities, select preferred design concepts, and set target engineering performance levels.*

Research Task 2—Design of Subsystem Human Appraisal Experiments: Surveys are required to elicit consumer preferences for system and sub-system design features and to provide the data needed to estimate preference models. *An optimal design of human appraisal experiments methodology is developed, which considers that the experiments are completed by heterogeneous human respondents, and supports modeling of human preferences explicitly including the impact of heterogeneity.* Features of the human appraisal experimental design method are that the experiment is optimized to estimate a response surface ordered logit model, a large number of product and demographic factors can be accommodated, fatigue of human respondents is mitigated, the unique rating style of individual respondents is accounted for, and specific factor combinations can be included or excluded from the design..

Research Task 3—Analysis of Human Appraisals for Modeling Consumer Heterogeneity: While the ultimate goal of this research is the use of the hierarchical choice model to set engineering attribute targets, the data used in the modeling process must be analyzed to determine the best modeling method, maximize the goodness-of-fit of the resulting models, and gain insights into the heterogeneity of human preferences that may not be obvious from a preference model alone. Specific methods are developed to analyze human appraisal experiments, which present a unique set of challenges compared to industrial or scientific experiments due to the effects of respondent heterogeneity and human behavior. *Multivariate* *statistical methods are utilized for this task to understand respondent heterogeneity and to support consumer preference modeling.*

Research Task 4—Integrated Bayesian Hierarchical Choice Modeling: A link between customer-desired attributes in the choice model and engineering design attributes used for product development is required for the design of complex systems. The complex system problem is characterized by a hierarchy of attributes in choice model estimation, a hierarchy of consumer demographic descriptors, and data from multiple sources of varying degrees of richness. Additionally, it cannot be assumed that the data needed to estimate a choice model for a complex system resides in a single survey, but rather is contained in several sub-system surveys. *The Integrated Bayesian Hierarchical Choice Modeling approach is formulated to mathematically map qualitative choice criteria to quantitative engineering attributes*. The approach considers the hierarchy of components and subsystems in a complex system, utilizes multiple sources of data, and affords a mechanism to quantify uncertainty and minimize model errors for a hierarchical system of preference models. The mixed logit (MXL) choice model is used to capture systematic and random heterogeneity, and Bayesian methodology is used for *integrated* estimation of the system of models in the hierarchy.

Development of techniques for the four listed research tasks enables implementation of the Enterprise-Driven *Decision-Based Design (DBD) framework*, which provides the basis for a rigorous decision making methodology for engineering design. The Decision-Based Design framework will be described in Chapter 2. The research developments will be illustrated using the **vehicle occupant packaging design** problem, which provides the necessary complexity and attribute trade-offs to demonstrate the proposed techniques.

1.3 Vehicle Occupant Packaging Design

Vehicle occupant package design is a multidisciplinary design activity that requires setting package design targets in terms of standard Society of Automotive Engineers (SAE) dimensions, in the presence of overall vehicle design considerations, such as structural, safety, and styling dimensions, illustrated in Figure 1.4. The design problem is characterized by often conflicting objectives among the design of the exterior styling, the vehicle structure, and the occupant package layout. In actual design practice, the conflicts are often resolved using extensive benchmarking of competitive vehicles and heuristic approaches. Several methodical approaches to capture the interaction of occupant packaging with other vehicle sub-systems have been investigated in the literature (Parkinson and Reed, 2006; Noui-Mehidi, 1997; Hamza et al., 2004; de Weck and Suh, 2006). While these approaches consider specific interactions between occupant packaging attributes and select vehicle attributes, they do not consider the trade-offs among multiple vehicle attributes, while simultaneously considering customer preferences. Further, models have been used for posture prediction (Reed et al., 2000; Reed et al., 2002) but have not addressed the relationship between anthropomorphic attributes and customer preferences for packaging.

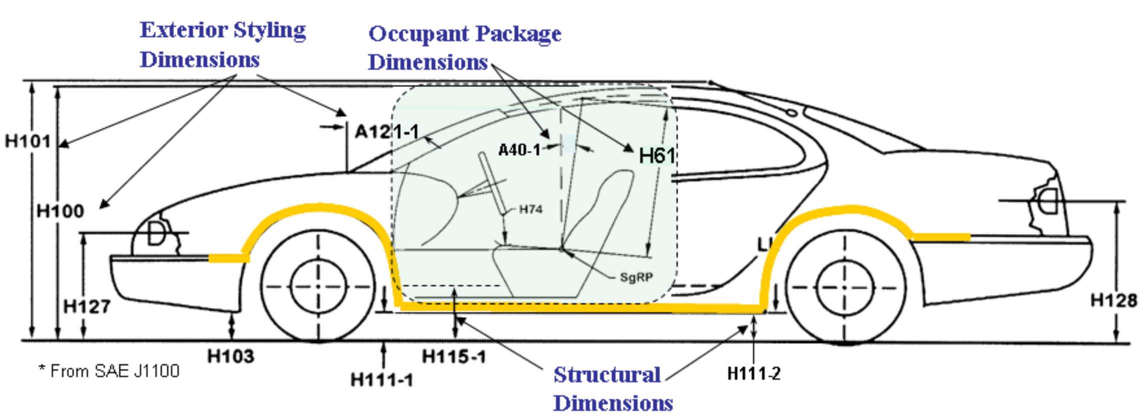


Figure 1.4: Vehicle Occupant Packaging Design Trade-offs (Society of Automotive Engineers, 2002)

The vehicle package must be designed to best meet the needs of a demographically diverse target population, characterized by diversity in socio-economic attributes (e.g. age, income), anthropomorphic attributes (e.g. height, weight), expectations based on previous purchase history (e.g. vehicle brand, size), and intended usages (e.g. commuting, moving), as illustrated in Figure 1.2. The market is also heterogeneous, composed of many market segments in which similar vehicles compete (e.g., small SUV, compact sedan). Unlike other vehicle specifications, setting package targets has been heavily influenced by qualitative considerations, such as overall roominess of the occupant package. In addition, customer-perceptions of the vehicle occupant package can be influenced by external factors, such as the market/product segment (SUV vs. midsize car) and the perceived status (luxury vs. economy) of the vehicle. Due to such complexity, targets for packaging attributes have traditionally been determined primarily through benchmarking of competitive vehicles and experience, limiting the potential for optimization of a vehicle design for a given market segment.

1.4 Organization of the Dissertation

The organization of the dissertation is illustrated in Figure 1.5. Chapter 2 presents both the technical background and the previous work underlying the four research tasks described in Section 1.2. Chapter 3 presents the Product Attribute Function Deployment (PAFD) method for design selection to address Research Task 1. While the method is general and can be used for both selecting a preferred design concept early in the process and setting target performances later in the design process, the method is demonstrated for design concept selection early in the design process. The design example used is an automotive pressure sensor rather than the vehicle packaging problem introduced in Section 1.3 because the choice model structure is straightforward and allows for a clear demonstration of the PAFD method and principles. The

PAFD method provides a process tool for design selection; however, the method does not address acquiring the data needed to build a choice model or how to create choice models for complex systems. Chapter 4 presents the optimal design of experiments for Human Appraisal method to address Research Task 2. This method provides the means to collect human preference data which is optimal for building preference models and understanding consumer heterogeneity. Chapter 5 provides a methodology to statistically analyze preference data to understand consumer heterogeneity as well as to preprocess the data to create efficient preference models. The methods presented address Research Task 3. Chapter 6 presents the Integrated Bayesian Hierarchical Choice Model (IBHCM) approach which provides a comprehensive choice modeling approach for complex systems in which both qualitative and quantitative choice attributes are considered (Research Task 4). The model is estimated using both data collected from experiments conducted using design of experiments for human appraisal method of Chapter 4 and processed using the methods of Chapter 5, as well as market survey data. Chapter 7 details the contribution of this research as well as areas for future research.

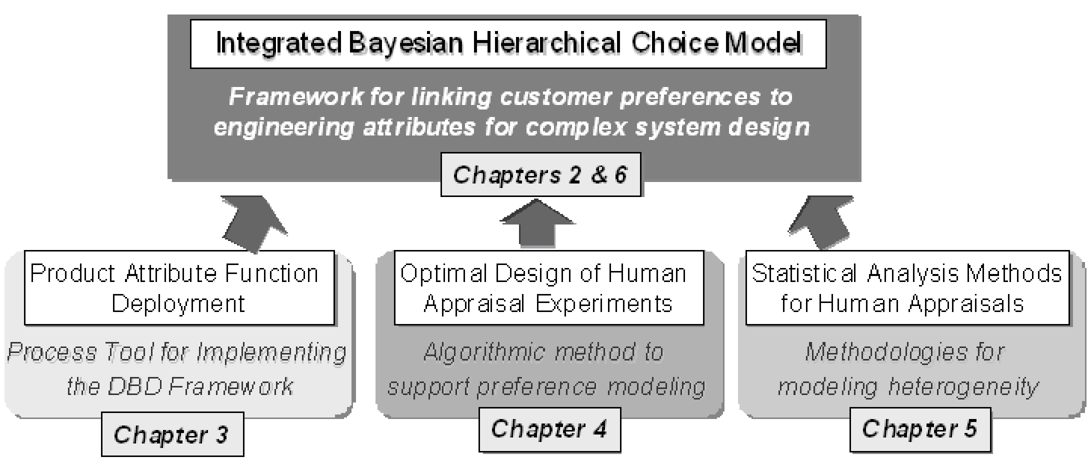


Figure 1.5: Organization of Research Presented in the Dissertation

Chapter 2

TECHNICAL BACKGROUND

The work in this dissertation is rooted in the discrete choice method for modeling product demand, as part of a larger effort to enable enterprise-driven Decision-Based Design (DBD). Demand modeling is necessary to estimate the potential profit of an engineering system or product, which is used as the selection criterion in the DBD framework. In this chapter, the Decision-based Design (DBD) framework is introduced, a brief tutorial on discrete choice analysis for demand modeling is provided, the hierarchical choice modeling approach is introduced, multilevel formulations of the DBD framework are described, design of experiments for human appraisals are introduced, and methods for statistical analysis and preprocessing of data are provided.

2.1 DECISION-BASED DESIGN (DBD) FRAMEWORK

2.1.1 DBD Motivation and Overview

Within the engineering research community, there is a growing recognition that decisions are the fundamental construct in engineering design (Lewis et al., 2006; Marston et al., 2000; Shah and Wright, 2000; Dong and Wood, 2004; Herrmann and Schmidt, 2002; Gu et al., 2002; Wassenaar and Chen, 2003). Based upon this premise, the ***Decision-Based Design*** framework has been developed, which merges the separate marketing and engineering domains into a single enterprise-level decision-making framework. The framework utilizes a decision-theoretic methodology to select the *preferred* product design alternative for the enterprise undertaking the design activity, as well as set *target levels of performance* for the product. This is accomplished as shown in Figure 2.1 through a hierarchical model linkage in which design concepts and variables (**X**) are linked to demand, *Q*, through engineering analysis and attribute mapping between **engineering attributes E** (e.g. fuel economy, horsepower) and **customer-desired attributes A** (e.g. comfort, performance). Also key is the inclusion of **demographic attributes S** (e.g. age, income, height), in addition to customer-desired attributes **A,** in the estimation of demand, to capture the heterogeneity of consumer preference.

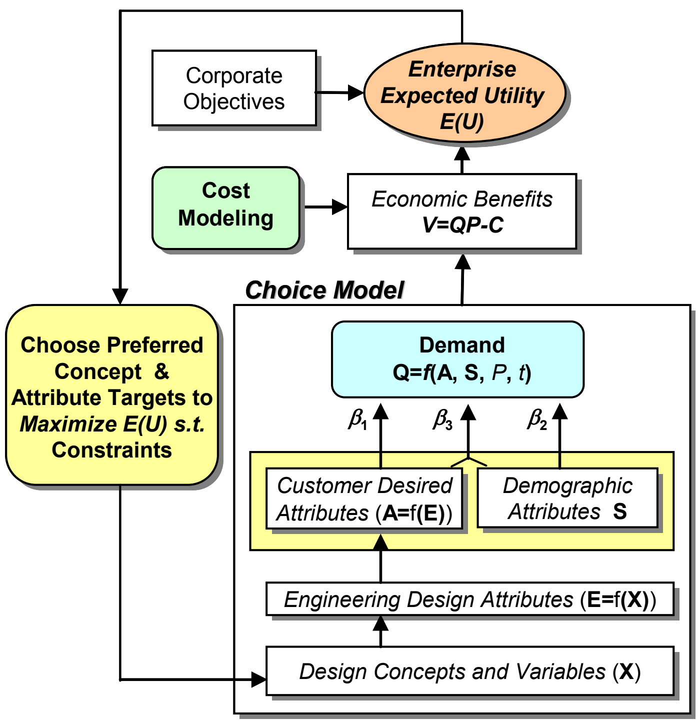


Figure 2.1: The Decision Based Design Framework for Conceptual Design

In the DBD implementation (Wassenaar and Chen, 2003), a single criterion, *V*, which represents economic benefit to the enterprise, typically profit, is employed as the selection criterion. This single-objective approach avoids the difficulties associated with weighting factors and multi-objective optimization which can be shown to violate Arrow's Impossibility Theorem (Hazelrigg, 1996). A **utility function**, *U*, which expresses the value of a designed artifact to the enterprise, considering the decision maker's risk attitude, is created as a function of the **selection criterion**, V. A preferred concept and attribute targets are selected through the maximization of enterprise utility.

2.1.2 Enterprise-Driven DBD Formulation

In the DBD formulation, utilizing **profit**, Π, as the selection criterion (V) captures the needs of both the consumer and the producer stakeholders, resulting in maximum benefit to the enterprise when utility is maximized. The profit function is intended to represent the profit contribution attributed to *design attributes* and not enterprise-level profitability for a product as a whole, similar to the traditional use of cost functions in engineering optimization (Siddall, 1982) used to represent design cost and not total enterprise costs. Profit is expressed as a function of product demand Q, price P, and total cost C, where demand Q, is expressed as a function of customer-desired attributes $\mathbf{A}$, customers' demographic attributes $\mathbf{S}$, price P, and time t:

$$V = \Pi = Q(\mathbf{A}, \mathbf{S}, P, t) \cdot P - C \quad . \tag{2.1}$$

Similar to "customer attributes" in QFD, $\mathbf{A}$ are product characteristics that a customer typically considers when purchasing the product. To enable engineering decision-making, qualitative customer-desired attributes $\mathbf{A}$ must be expressed as a function of quantitative engineering attributes $\mathbf{E}$ in the demand modeling phase. This functional relationship can consist of a *hierarchy* of models mapping $\mathbf{A}$ to $\mathbf{E}$ to establish the relationships necessary for decision-making. Cost, C, is a function of the design attributes, $\mathbf{E}$, exogenous variables $\mathbf{Y}$ (the sources of uncertainty in the market), demand, Q, and time t. Price, P, is an attribute whose value is determined explicitly in the utility optimization process, or obtained from a separate price optimization process. Based upon these functional relationships, the selection criterion can be expressed as:

$$V = \Pi = Q(\mathbf{A}(\mathbf{E}), \mathbf{S}, P, t) \cdot P - C(\mathbf{E}, \mathbf{Y}, Q, t). \tag{2.2}$$

It should be noted that uncertainty is considered explicitly and the objective is expressed as the maximization of the expected enterprise utility $E(U)$, considering the enterprise risk attitude:

$$\mathbf{max}: \; E(U) = \int_V U(V)\, pdf(V)\, dV \tag{2.3}$$

where V is the single selection criterion in Eq. (2.2).

As seen in Figure 2.1, decision-making regarding a preferred design concept, as well as optimal levels (targets) of engineering design attributes **E** is performed using optimization to maximize the expected enterprise utility $E(U)$, subject to appropriate constraints.

2.2 Discrete Choice Analysis (DCA) for Demand Modeling

Discrete Choice Analysis (DCA) (Ben-Akiva and Lerman, 1985; Koppelman et al., 2005) is used to model product demand by capturing *individual* customers' choice behavior, in which performance of a given product is considered versus that of competitive products. It should be noted that in this formulation, the customers could be either individual consumers or industrial customers. DCA is based upon the assumption that individuals seek to maximize their personal **customer choice utility**, u, (not to be confused with enterprise utility, U) when selecting a product from a choice set.

2.2.1 Formulation of the Discrete Choice Analysis Model

The concept of choice utility is derived by assuming that the individual's (n) true choice utility, u, for a design alternative, i, consists of an observed part W, and an unobserved random disturbance ε (unobserved utility):

$$u_{in} = W_{in} + \varepsilon_{in} \ . \tag{2.4}$$

While there are a number of DCA techniques popular in literature (e.g., Multinomial Logit, Nested Logit, Mixed Logit), they are distinguished from each other by the degree of sophistication with which they model the unobserved customer choice utility error ε and heterogeneity in customer preferences. In the **Multinomial Logit** (MNL) model, the coefficients (β) of the observed customer choice utility function (W) for the product attributes are identical across all customers. However, heterogeneity is modeled by considering demographic attributes **S** (e.g., customer's age, income, etc.) in the customer choice utility function. Assuming this utility function can be expressed as a linear combination of attributes, W follows the form:

$$W_{in} = \boldsymbol{\beta} \cdot \mathbf{Z} = \beta_{0i} + \boldsymbol{\beta}_{1i}\mathbf{S}_n + \boldsymbol{\beta}_2\mathbf{A}_i + \boldsymbol{\beta}_3\left(\mathbf{S}_n \cdot \mathbf{A}_i\right). \tag{2.5}$$

where β_{0i} is an Alternative Specific Constant (ASC), $\boldsymbol{\beta}_{1i}$ is an Alternative Specific Variable (ASV), and **Z** is the set containing **A** and **S**. The MNL model exhibits the Independence of Irrelevant Alternatives (I.I.A.) property, which leads to proportional substitutions patterns among the alternatives considered. In cases in which this property is undesirable, the **nested** or **mixed logit** formulations can be used to relax this assumption.

The mixed logit model (MXL) is distinguished from the MNL model in that it allows for random taste variation, i.e. the parameters β vary over respondents. Therefore, the mixed logit probabilities are integrals of the multinomial logit probabilities over a density of parameters, as expressed in the form:

$$\Pr_n(i) = \int \left(\frac{\exp(W_{in}(\beta))}{\sum_j \exp(W_{jn}(\beta))} \right) pdf(\beta) d\beta \ . \tag{2.6}$$

where $pdf(\beta)$ is the probability density function of model parameter β s. The mixed logit model has been demonstrated to be capable of approximating any random utility discrete choice model (Train, 2003). One of the most important advantages of the mixed logit model is that heterogeneity in customer preferences is decomposed into a systematic part, expressed by **S**, and a random part expressed by random coefficients β; in MNL, only the systematic part is estimated, with the random heterogeneity lumped into the error term ε_{in}. No closed form solution exists for Eq. (2.7). Therefore in practical applications, the mixed logit choice probability is approximated (i.e. by $\hat{\Pr}_n(i)$) using numerical simulation by taking a finite number of draws $r = 1,2,3,...,R$ from the distribution:

$$\hat{\Pr}_n(i) = \frac{1}{R}\sum_{r=1}^{R}\Pr_{n,r}(i) = \frac{1}{R}\sum_{r=1}^{R}\frac{\exp\left(W_{ni}(\beta_r)\right)}{\sum_j \exp\left(W_{nj}(\beta_r)\right)}. \tag{2.7}$$

where R is the number of random draws, $\Pr_{n,r}(i)$ is the probability of respondent n choosing product i in the r^{th} draw, and β_r is the corresponding simulated random coefficients.

2.2.2 Estimation of the Discrete Choice Analysis Model

The choice model is estimated using Maximum Likelihood Estimation (MLE) or Hierarchical Bayes Estimation (HBE). In the MLE method, model parameters (i.e. β) are found through maximization of the likelihood function L for the MNL or MXL model:

$$L\left(y_n \mid \beta\right) = \prod_{n=1}^{N}\prod_{i=1}^{J}\left(\Pr_n(i)\right)^{y_{ni}} , \tag{2.8}$$

where y_n is the response, i.e. the individual choices in the MXL model. To aid the solution process, the log-likelihood function (LL) is typically maximized because the LL function is additive as opposed to multiplicative.

In order to reduce the computational burden associated with multivariable sampling for MLE of the mixed logit model, Hierarchical Bayes Estimation methods were developed utilizing Markov Chain Monte Carlo methods with a Gibbs sampler to estimate the mixed logit model (Rossi et al., 2005). In the Hierarchical Bayes choice modeling paradigm (Gelman et al., 2004), the choice probability is modeled using a method in which the posterior distribution of the β_n parameters, characterized by a mean b and covariance matrix Σ, is found as a function of the prior distribution of b^0 and Σ^0, and an information source of observations, Y. In the hierarchical prior distribution, the distribution of β_n is conditional upon the distribution of the population-level hyper-parameters b and Σ. The population-level hyperparameters characterize the distribution of β_n in the population as a whole. Thus, model parameters β, b, and Σ are given by the parameter posterior distribution, pdf^*:

$$pdf^*(\beta, b, \Sigma \mid Y) \propto \prod_{n=1}^{N} L(y_n \mid \beta_n)\, pdf(\beta_n \mid b^0, \Sigma^0)\, pdf(b^0, \Sigma^0), \tag{2.9}$$

where pdf is the prior distribution (the denominator is excluded for simplicity), L is the likelihood function of the MXL model, and b is the mean vector and Σ is the full variance-covariance matrix of β.

The expression in Eq. (2.9) demonstrates a fundamental difference between the HBE and MLE approaches: the Bayesian method estimates the *mean* of a distribution, whereas the MLE solution estimates the maximum, or *mode*, of a distribution. The HBE method has several advantages over MLE for model estimation. If the prior distribution of β_n are assumed to be multivariate-normally distributed, i.e. $\boldsymbol{\beta} \sim MVN(b, \Sigma)$, estimation of random parameters is more computationally efficient than classical MLE methods. The Bayesian method allows for estimation of the true posterior distribution and recovery of the individual level β_n, unlike the MLE method which only provides point estimates of the mean b and variance Σ of the assumed distribution of β_n. Through the specification of hierarchical prior distributions, this solution technique estimates the posterior distribution of $\boldsymbol{\beta}$, and provides a mechanism for model updating through the definition of the prior distribution as information evolves.

2.2.3 Demand Forecasting using Discrete Choice Analysis

Estimation of the customer choice utility function (W) allows the **choice share**, $C.S.$, for a choice alternative i to be determined by summing over the market population, N, all probabilities, $\Pr_n(i)$, of a sampled individual, n, choosing alternative i from a set of J competitive choice alternatives:

$$C.S.(i) = \sum_{n}^{N} \Pr_n(i) = \sum_{n}^{N} \frac{e^{W_{in}}}{\sum_{k=1}^{J} e^{W_{kn}}}. \tag{2.10}$$

The set of choice alternatives J may include both the new designed product and the existing competitive alternatives available. The choice alternative set is composed of either actual consumer purchase choices from a set of product alternatives, i.e. **Revealed Preference** or simulated product choices, such as those resulting from a market survey, i.e. **Stated Preference**. Demand for a given alternative, i, at time t, $Q(i)_t$, is the product of choice share, $C.S.(i)$, by the total **market size** (or aggregate market segment demand), $D(t)$, for a given market segment (e.g. automobile mid-size sedan):

$$Q(i)_t = C.S.(i) \cdot D(t). \tag{2.11}$$

2.3 HIERARCHICAL CHOICE MODELING

A large-scale design problem is characterized by attribute-hierarchies in demand model estimation, a hierarchy of consumer demographic descriptors (**S**), and data from multiple sources with varying degrees of richness (e.g., in-house marketing surveys, purchase data, exit interviews). Existing demand modeling approaches in the design literature require that product attributes considered in the choice model be quantitative. However, many criteria used by customers to choose between complex engineering systems tend to be qualitative, especially those at the system level. Also, as noted in Section 1.1, existing demand modeling approaches used in engineering design do not adequately account for consumer heterogeneity, nor do they consider multiple data sources.

2.3.1 Challenges in Choice Modeling for Complex System

A challenge in choice modeling of complex engineering systems is modeling the ***heterogeneity of customer preferences***. For the design of a complex design artifact like an automobile, it is important to model the diversity in customer-preferences in a more complete way. In general, capturing customer heterogeneity is a necessary component in understanding the perception of a design for a given population segment. Most existing approaches in the design literature do not consider heterogeneity of preference in modeling (i.e. systematic and random heterogeneity do not appear in the demand model). As discussed earlier, Li and Azarm (2000), and Michalek et al. (2005) used conjoint analysis, in which individual choice preferences were aggregated. Michalek et al. (2005) have considered random heterogeneity only in using a mixed logit choice model. Cook (1997) used a linear model derived from Taylor Series expansion which used product value and price to estimate demand. Wassenaar et al. (2003; 2005) considered the systematic heterogeneity only by including a limited number of demographic attributes (e.g., age, gender) in a DCA model. Wassenaar et al. (2004) also considered the use of an integrated latent variable modeling approach; however, the implementation of the approach was not completely successful due to the high computational expense, and the large number of explanatory variables involved in a complex system.

While the current demand modeling approaches described previously consider customer preferences when choice (purchase) decisions are involved, they do not fully consider the impact of ***customer preferences for individual product features***. In the automobile market, for instance, customers have distinct preferences for individual product features like engine characteristics (e.g., acceleration, noise, fuel economy), interior characteristics (e.g., roominess, instrument panel, material, seating), etc. as well as component-level attributes like suspension, tires, steering, etc. Attributes considered by customers in a choice situation may be qualitative, and require mapping to physical, measureable design attributes at the subsystem and component levels. While it may be possible to include all mapped quantitative component-level attributes in the product-level choice model, it is not realistic due to the sheer volume of such attributes and more importantly, such an approach does not consider that decisions on design-level attributes may only be required for a subset of all possible design attributes. Integrating preference models at different levels allows us to examine the effect of design changes at the component level on customer ratings for different product features as well as on customer choice.

Design of large artifacts is usually distributed over several teams, often spread across different geographical locations requiring design teams to work autonomously. In such a scenario, designers usually conduct surveys and human appraisals specific to subsystems and components (e.g., evaluation of an engine upgrade, vehicle-interior surveys and exterior/styling surveys) independently to preserve autonomy as well as to make the survey size manageable. In order to examine how customers trade-off between the different subsystem attributes when they make the purchase decision, it is necessary to ***combine data sources*** to simultaneously consider multiple feature-specific surveys. Estimating such pooled models is known as **model fusion** (Allenby et al., 2005) or **data enrichment** (Louviere et al., 2000) in the transportation literature. Existing approaches in the design literature have only used data from a single source—either stated preference (SP) data (Michalek et al., 2005) or revealed preference (RP) data (Wassenaar and Chen, 2003). RP data refers to actual choice (i.e., purchase) behavior that is observed in real choice situations. SP surveys are used to learn about how people are likely to respond to new products or new product features through a market survey. Preliminary work on combining RP and SP sources of data specifically for product design has been examined in (Kumar et al., 2007; Kumar et al., 2009), but more comprehensive methods are required.

2.3.2 Previous Work in Hierarchical Choice Modeling

To deal with the challenges presented in Section 2.3.1, a **Hierarchical Choice Modeling** strategy has been proposed (Kumar et al., 2009) as shown in Figure 2.2, in which the top system level choice model only contains a reasonable set of system-level customer-desired attributes **A** (including price *P*), while the lower level models establish the relationships between qualitative customer perceptual attributes **A** as functions of quantitative engineering design attributes **E** and demographic attributes **S,** i.e., **A**=f(**E, S).** This ensures a more manageable model at each level, and mitigates the model estimation issues that accompany an all-in-one approach. The proposed approach uses customer **ratings** for qualitative attributes in the choice model, which are expressed in terms of quantitative engineering attributes through a hierarchy of *linking* models. For example, qualitative attributes in the top-level DCA analysis model, labeled $\mathbf{M_1}$ in Figure 2.2, may be linked to engineering attributes through a series of **ordered logit** ratings prediction models for the subsystems, labeled $\mathbf{M_2}$ and $\mathbf{M_3}$ in the figure.

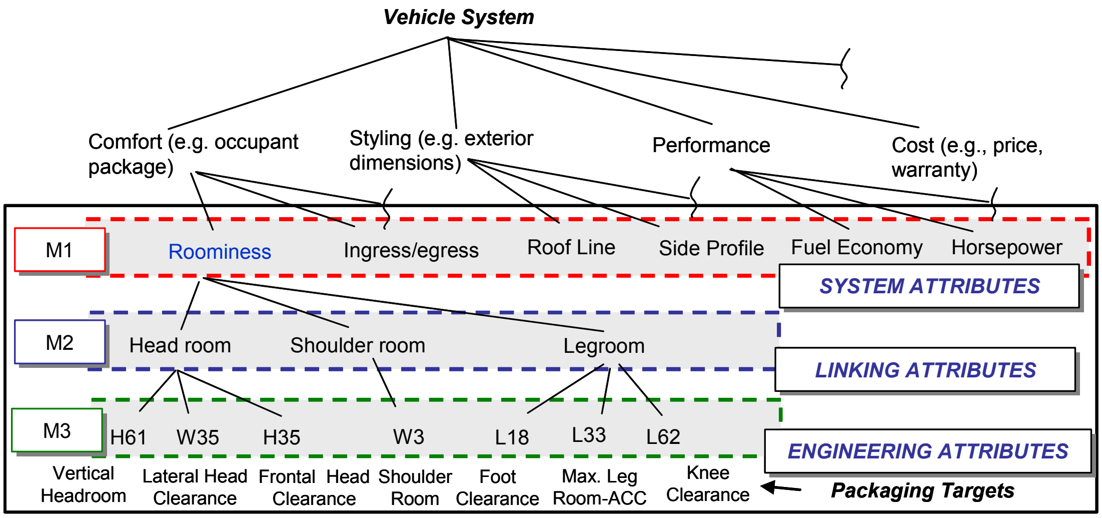


Figure 2.2: Hierarchical Choice Model Approach

A key issue is determining the set of **S** to include at each model level. In order to ensure comprehensive modeling of systematic heterogeneity and thus minimize unexplained heterogeneity, a taxonomy of **S** has been developed for model estimation (Kumar et al., 2009). The proposed taxonomy of Figure 2.3 is expressed as the following:

- $\mathbf{S}_1$*: Socio-Economic attributes* (e.g., age, income)
- $\mathbf{S}_2$: *Anthropometric variables* (e.g., stature, weight)
- $\mathbf{S}_3$: *Purchase History* (e.g., vehicle type last purchased)
- $\mathbf{S}_4$: *Usage Context attributes* (e.g., construction, moving)

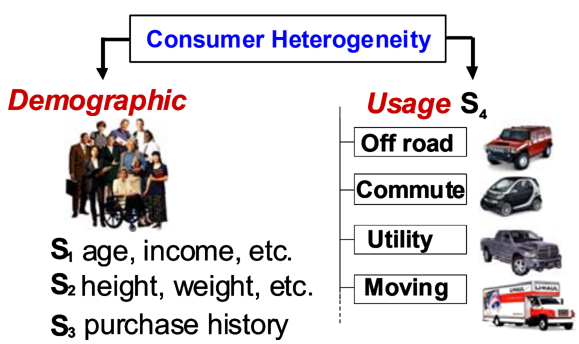


Figure 2.3: Taxonomy of S

2.3.3 Ordered Logit for Modeling Rating Responses

As discussed in the previous subsection, methods are required to model consumer preferences expressed as ratings as a function of quantitative engineering attributes to enable the hierarchical choice model. To fit a predictive model to survey ratings, or *ordinal data* (*e.g.*, 1=poor, 2=fair, 3=good; rating from 1 to 10), alternative methods to standard linear regression are required. A key assumption of linear regression is violated when used to fit ordinal data because the expected model error cannot be assumed to be of zero mean with constant variance: the true value of the dependent variable is not a linear function of the explanatory variables **Z,** as shown Figure 2.4 (McKelvey and Zavoina, 1975). Further, an ordinal dependent variable is not unbounded as required by linear regression (Lu, 1999), but rather takes on a fixed number, p, of discrete values as defined by the survey design (e.g., rating scales of 1-10, 1-7).

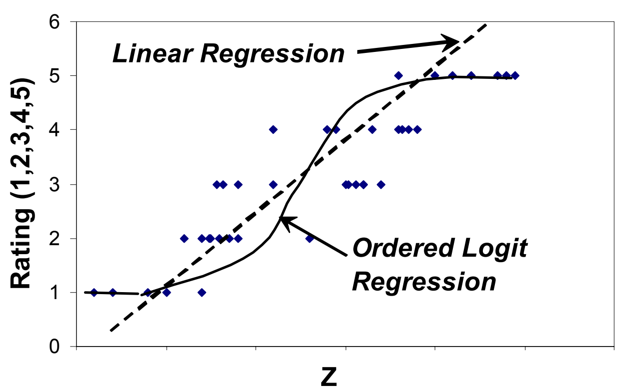


Figure 2.4: Illustration of the Variation of Ratings vs. Explanatory Variables Z (McKelvey and Zavoina, 1975)

For this reason, the **ordered logit model** is used in this work to estimate models for ordinal customer ratings. McKelvey and Zavoina (1975) introduced ordered probit regression for ordinal data, in which the ordinal ratings were assumed to be discrete representations of a continuous underlying, normally distributed opinion or utility. McCullagh (1980) introduced ordered logit in which the underlying distribution is logistically distributed, leading to the proportional odds model. In this model, the cumulative odds ratios are identical across ratings categories. Hedeker and Gibbons (1994) developed a random effects ordered probit formulation, which considered the β to be random and potentially functions of respondent level attributes (e.g., age, income), or *covariates*. Tamhane et al. (2002) modeled the underlying utility response using the beta distribution to allow greater flexibility (i.e. not symmetric) and to enable a bounded response.

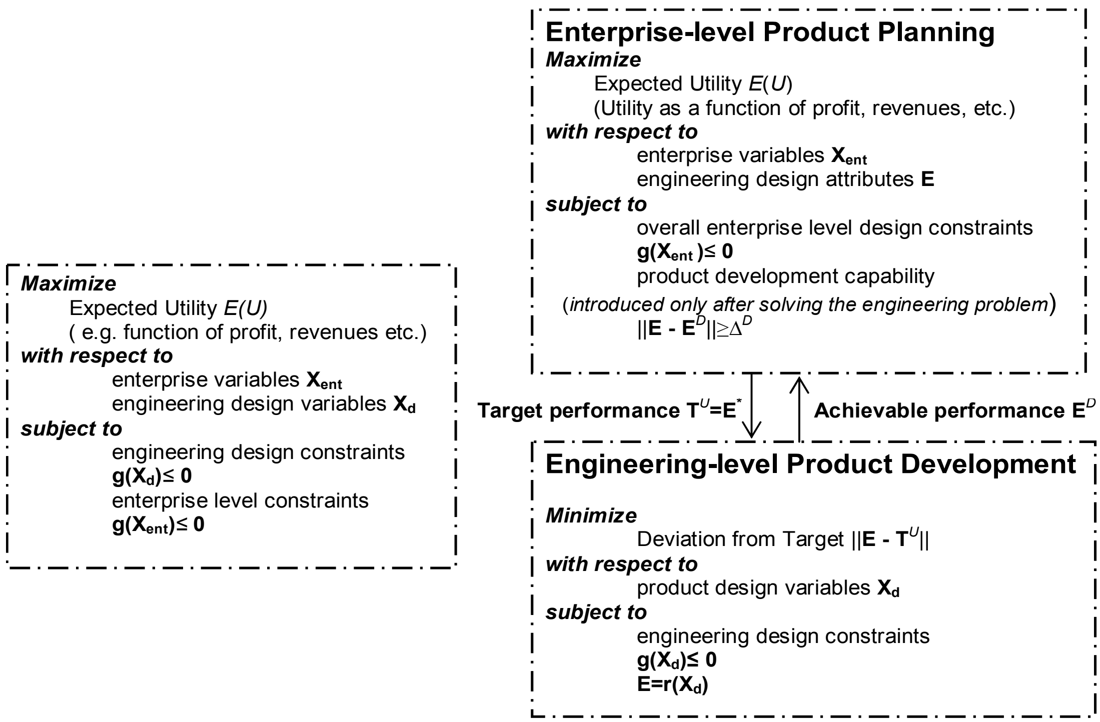
Ordered logit assumes that the p ordered ratings, **R**, are discrete representations of a continuous, underlying ***utility***, u_{in}, associated with each alternative, i, which is rated by each survey respondent, n. In the ordered logit formulation, the underlying utility measure, u_{in}, is based upon the same concept as the discrete choice model utility in that it is assumed to be the sum of a parameterized observable component, $W_{in} = \boldsymbol{\beta}\cdot\mathbf{Z}$, and an unobserved error component ε_{in}, as given previously by Eqs. (2.4) and (2.5). Also in the OL approach, it is assumed that the error variance is smallest at maximum or minimum values of **Z** and largest for moderate values of **Z** (*i.e.* responses at the ratings extremes are more certain than those in the middle regions). This appears to be a more realistic assumption compared to that used in linear regression. OL seeks to model the underlying utility, u_{in}, while the predicted discrete ratings, **R**, are estimated through the use of (p-1) ***cut points***, **k**, imposed on the distribution of the u_{in}, estimated to match the proportions of **R** present in the actual survey data. The ordered logit model is derived under the assumption that the probability, Pr, for any rating R_p is a function of observed utility and cut points, and that the unobserved errors ε_{in} are distributed logistically:

$$\begin{aligned}\Pr\left[R = R_p\right] &= \Pr\left[k_{p-1} < u_{in} < k_p\right] && \left(p = 1,2,\ldots,P\right)\\ &= \frac{e^{k_p - \boldsymbol{\beta}'\mathbf{Z}}}{1 + e^{k_p - \boldsymbol{\beta}'\mathbf{Z}}} - \frac{e^{k_{p-1} - \boldsymbol{\beta}'\mathbf{Z}}}{1 + e^{k_{p-1} - \boldsymbol{\beta}'\mathbf{Z}}} && \left(k_0 = -\infty, k_p = +\infty\right)\end{aligned} \quad (2.12)$$

The model parameters, **β**, and cut points, **k**, are determined using Maximum Likelihood Estimation (MLE) or Bayesian estimation. A **random-effects** version of the model is used in this work in which a random intercept term is used to capture random heterogeneity (Hedeker and Gibbons, 1994). When used for prediction purposes, the utility for an alternative, i, for a particular person, n, is first calculated, and then transformed to a rating using the (p-1) series of estimated utility cut points. As an alternative to the latent variable approach, ratings are used in this research to capture qualitative customer preferences. Ratings represent relative, or *ordinal*, preferences for an attribute, as opposed to absolute, or *cardinal*, preferences and thus require special consideration in modeling.

2.4 MULTILEVEL OPTIMIZATION FORMULATION TO DBD

The DBD formulation described in Section 2.1 was described as ultimately a process of maximizing the expected utility, $E(U)$, of a designed artifact to the enterprise. For complex systems, an All-in-One (AIO) method of solution to the maximum expected utility problem may not be feasible or desirable, and a multilevel method of solution may be implemented. Figure 2.5 illustrates the difference between the AIO approach and a multilevel optimization formulation to DBD. The AIO approach in Figure 2.5 (a) treats the problem of maximizing the expected value of enterprise-level utility $E(U)$ as a single optimization problem, where the decisions on product planning and product development are made simultaneously. Figure 2.5 (b) illustrates a decomposed multilevel framework (Kumar et al., 2006), reflecting a decoupling between the enterprise-level product planning and engineering product development. Following the “target cascading” paradigm (Kim, 2001; Kim et al., 2002; Kim, Michelena et al., 2003; Kim, Rideout et al., 2003; Michalek et al., 2005; Michelena et al., 1999) for multi-level decision making in industrial settings, engineering product development is viewed as a process for meeting the targets set from the enterprise level.



(a) All-In-One approach (b) Hierarchical approach

Figure 2.5: Comparison Between All-In-One and Hierarchical Multi-level Approach to DBD

Using a multilevel optimization formulation, at the upper level, the enterprise-level product planning problem maximizes the expected utility $E(U)$ with respect to the engineering design attributes $\mathbf{E}$ and enterprise-level variables, $\mathbf{X}_{ent}$. Decisions made on the optimal levels of engineering design attributes $\mathbf{E}$, represented as $\mathbf{E}^*$, are then used as targets or $\mathbf{T}^U$, passed to the lower level engineering product development process. The objective of the lower-level engineering product development is to minimize the deviation between the performance target $\mathbf{T}^U$ and the achievable product performance response $\mathbf{E}$ while satisfying the engineering feasibility constraints $\mathbf{g}$, for the design artifact. The achievable product performance $\mathbf{E}^D$ is then transferred to the enterprise level problem, and the optimizer creates new targets based upon the achievable performance.

The optimization problem at the engineering development level can be further decomposed and solved using multilevel optimization. Based on the nature of decomposition, either non-hierarchical or hierarchical, different multilevel optimization (MDO) formulations can be used. The motivation for decomposing the problem is the desire to leverage discipline specific knowledge in the formulation of the optimization problem for each sub-system, or to incorporate an existing sub-system optimization formulation in the overall system optimization problem. The Analytical Target Cascading (ATC) approach decomposes the original engineering problem hierarchically at multiple levels, and operates by formulating and solving a minimum deviation optimization problem (to meet targets) for each element in the hierarchy. Collaborative Optimization (CO) (Braun, 1996) can be utilized to decompose the problem non-hierarchically, reflecting a collaborative design environment. Under a multilevel design framework, the ideal product development scenario occurs when the utopia targets corresponding to the maximum enterprise utility are achievable by the engineering design. In most engineering design cases, however, it is uncommon to achieve the utopia target due to the trade-off nature of multiple attribute target values and physical feasibility (i.e., no feasible design is available to meet the targets perfectly). If the engineering feasible domain is disconnected in the space of performance attributes (i.e., multiple, discrete feasible designs are available), the task becomes more challenging.

In this work, the DBD problem will be solved using the AIO approach to demonstrate the features of the proposed method; however, the problems can be formulated to use a multilevel approach, such as ATC or CO, if desired.

2.5 DESIGN OF EXPERIMENTS FOR HUMAN APPRAISALS

Human appraisal experiments can be differentiated from other types of experiments in the literature. Industrial and scientific design of experiments have been well documented (Box et al., 2005; Montgomery, 2005; Myers and Montgomery, 2002) and utilized in practice. The response in such experiments is the output of a physical process, such as from industrial machinery or a new product test bed, and therefore fatigue is not generally an issue. This class of experiments is characterized by random error, $\varepsilon_{\mathrm{rand}}$, in the response due to uncontrolled nuisance factors. While advanced methods for reducing the random error of designed experiments, such as blocked and split-plot designs, are used in this class of experiments, the reasons are typically due to nuisances or compromises in the experimental design which introduce additional error or prevent full randomization, as opposed to being an integral feature of the design. Computer experiments have been studied extensively (Simpson et al., 2001; Jin et al., 2001) for the purpose of metamodeling, and are characterized by a lack of random error, and thus methods of blocked or split-plot designs are not used: the goal of computer experiments is a uniform coverage of the design space to minimize bias error. Conjoint experiments have been used for product or service evaluations in the marketing field (Green and Srinivasan, 1978; Green and Srinivasan, 1990; Louviere et al., 2000), and are characterized by random error, $\varepsilon_{\mathrm{rand},}$ in the response and blocks corresponding to each respondent; however, they have not considered human attributes **S** in the design of the experiments but rather have treated the **S** as covariates (i.e. quantities recorded during the experiment but not used in the design of the experiment). A comparison of optimal distribution of design points to minimize bias error in computer experiments, versus that of a conjoint experiment to minimize random error is shown in Figure 2.6. Garneau and Parkinson (2007) have demonstrated that both systematic and random anthropomorphic heterogeneity are significant predictors of preferences for product designs in which the design interacts with the human (e.g., an exercise bicycle seat); however, a general approach for designing experiments for such human appraisals and methods to separate respondent level variation from random variation was not presented.

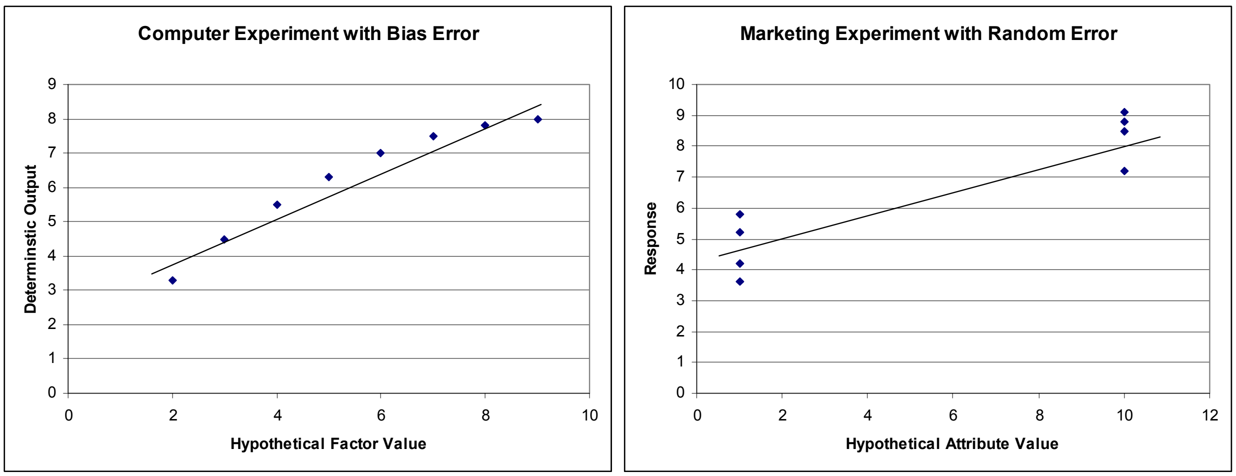


Figure 2.6: Error Comparison of Computer vs. Physical/Appraisal Experiments

The human appraisal experiment is presented as a separate class of experiment in this work, specific to product evaluations in which the human attributes of the respondent have an *observable, systematic* influence upon the response, in addition to the random effect captured by the random block effect as in a general conjoint analysis. Standard experimental designs and other experimental design approaches for human appraisals are generally not suitable for these experiments, which are conducted with the goal of creating a response surface model to understand respondent preferences as a function of product and human attributes. Standard split-plot designs based upon standard full factorial or fractional factor designs for response surface creation, considering significant respondent blocking, do not exist (Box et al., 2005; Myers and Montgomery, 2002). Orthogonal array designs (Phadke, 1995), such as the L_{18} design, are small enough such that each person can complete the entire experiment and blocking is not required;

however, while such designs allow estimation of linear and quadratic terms, interactions can not generally be estimated. Experiments specifically for human appraisals, with the goal of minimizing the number of configurations for each respondent to evaluate, have been developed for certain situations. Adaptive Conjoint Analysis (Green et al., 1991) uses a prescreening of preferences for factor levels to optimize the configurations presented; however, this approach requires gaining access to resources for the prescreening tests and ignores the importance of factor interactions. One-factor-at-a-time experiments (Frey et al., 2003) have been developed to reduce the number of configurations needed when the goal of the experiment is to identify an optimal configuration. While this approach is effective for optimization, the goal of the human appraisal experiment in this work is to create a response surface model over a design space to understand response behavior. Based on the limitations of existing approaches, an approach using the *D*-optimality criterion is implemented as the method for selecting a human appraisal experiment.

To select experimental designs for human appraisals, given a constraint on the number of configurations rated by a single respondent (due to fatigue) and multiple product and human attributes, optimal design of experiment methods are adapted to the specific needs of this class of experiments. Optimal design of experiments (DOE) have been studied for a variety of applications, such as industrial, agricultural, or scientific experiments, e.g. Atkinson and Donev (1992), and conjoint experiments, e.g. Kuhfeld et al. (1994) and Kessels et al. (2008). The methodology has been extensively developed for Ordinary Least Squares (OLS) modeling (Atkinson and Donev, 1992) and has been extended recently to Generalized Least Squares (GLS) to account for the error variance structure in blocked or split-plot experiments (Goos, 2002). Optimal DOE methodology has also been applied to multinomial logit (MNL) discrete choice analysis models (Kuhfeld et al., 1994; Sandor and Wedel, 2001; Kessels et al., 2006), as well as general logistic regression, including ordered logit and ordered probit (Zocchi and Atkinson, 1999; Chipman and Welch, 1996; Heise and Myers, 1996; Perevozskaya et al., 2003); however, a general approach to account for the combined split-plot and block structure of the human appraisal experiments has not been presented and is therefore a focus of this work.

2.6 STATISTICAL DATA ANALYSIS AND PROCESSING OF HUMAN APPRAISAL EXPERIMENTS

Results from human appraisal experiments require analysis prior to creating ordered logit models due to the multiple responses provided by respondents, the correlation of ratings elicited by a single respondent, and the many potential product and demographic factor forms that can be utilized in the modeling process. Multivariate statistical analysis methods have been developed for the purpose of data exploration, reduction, classification, and relationship identification (Johnson and Wichern, 2002). For the purpose of data reduction, Factor Analysis or Latent Class Analysis (McCutcheon, 1987) is used. The purpose of these methods is to describe the covariance relationship among many observed random quantities in terms of a few underlying, unobserved factors, or latent variables. Factor Analysis is used for continuous observed variables, whereas Latent Class Analysis is used for discrete (categorical or ordinal) observed variables. In the area of data exploration, cluster analysis is commonly employed, particularly in the area of market segmentation analysis (Green and Krieger, 1995). The goal of cluster analysis is to find natural groupings of items or variables based upon similarity of the items, or variables. For data classification, methods broadly classified as data mining techniques (Witten and Frank, 2005) are used to classify a set of objects or observations into groups, with different methods providing different insights into the classification process. For data relationship identification, regression methods broadly classified as generalized linear models, such as ordered logit modeling, are used to predict the value of a response variable based on a set of predictor variables. To understand the relationship before the modeling process, analysis of variation (ANOVA) methods (Box et al., 2005) are utilized to understand the portion of variation explained by each factor.

While the standard statistical techniques exist, the use of the techniques to support preference modeling in general, and application to the hierarchical choice modeling approach, must be examined.

Chapter 3

PRODUCT ATTRIBUTE FUNCTION DEPLOYMENT FOR DESIGN SELECTION

As noted in Chapter 1, product planning requires a design process tool to establish engineering priorities, select the preferred design concept, and set target levels of engineering performance while considering the needs of both the consumer and producer. In this chapter, a new design tool called Product Attribute Function Deployment (PAFD), based on the principles of Decision-Based Design (DBD), is introduced as a decision-theoretic, enterprise-level process tool to guide the conceptual design phase. Other process tools, such as Quality Function Deployment (QFD), have been developed as design process tools to translate customer needs into engineering characteristics; however, significant limitations have been identified with such process tools. While existing tools such as QFD provide a useful visual format and encourage an interdisciplinary design process, they rely upon subjective performance assessments and potentially faulty rating and ranking methods. The PAFD method extends the qualitative matrix principles of QFD while utilizing the quantitative decision-making processes of DBD. The PAFD method is built upon established methods in engineering, marketing, and decision analysis to eliminate the need for the user ratings and rankings of performance, priority, and attribute-coupling in the QFD method. The differences between the QFD and the PAFD processes are compared and contrasted, and the conceptual design of an automotive Manifold Absolute Pressure sensor is used as a case study to demonstrate the features and benefits of the PAFD method. The general framework presented in this chapter can be utilized with the Bayesian Hierarchical Choice Model developed in Chapter 6 to make design decisions for a complex system.

The chapter is organized as follows: Section 3.1 introduces the challenges in design selection, Section 3.2 describes the limitations of current methods used for design selection, Section 3.3 describes the use of DBD for design selection, Section 3.4 develops the new PAFD method, and Section 3.5 provides a demonstration of the method for automotive sensor design selection.

3.1 INTRODUCTION

In the early stages of product design there is a need to set engineering priorities, primarily through the selection of a preferred design concept, identification of key product attributes, and establishment of performance targets for the artifact or *product* under design. Because product decisions made in the early, or *conceptual*, design phase can account for up to 75% of the committed manufacturing cost (Ullman, 2002), it is essential that these decisions be rigorous and consistent with the objectives of the firm or *enterprise*. A design process tool utilized to guide these critical product planning activities must consider the needs of both the consumer and the producer in order to select concepts and set targets which will maximize the benefit to the enterprise as a whole. While design freedom is at a maximum in this phase, design knowledge is at a minimum, requiring that decisions made in this phase also explicitly consider uncertainty.

Within the engineering research community, there is a growing recognition that decisions are the fundamental construct in engineering design (Marston et al., 2000; Shah and Wright, 2000; Dong and Wood, 2004; Herrmann and Schmidt, 2002; Gu et al., 2002). Traditionally, discipline specific decision-making methodologies, utilizing mathematical behavioral models such as those used in marketing (e.g., conjoint analysis) and engineering (e.g., differential equations), have been adopted based upon the specific needs of the individual discipline. These methods have used domain specific objectives as the decision criteria, biased towards either consumer product acceptance or producer performance metrics. These methods in isolation cannot achieve the necessary enterprise-level decision process required during the product planning phase, a fact which has been acknowledged by the development of various process tools which bridge different enterprise domains to support product design activities (Krishnan and Ulrich, 2001).

Quality Function Deployment (QFD) was developed to bridge the marketing and engineering domains using a much simplified, consensus-driven qualitative analyses. This process was developed as a means to link product planning directly to the “Voice of the Customer”. It remains the leading tool for setting engineering priorities, determining target levels of product performance through benchmarking and, when supplemented with Pugh’s Method, selecting a design concept. As shown in Figure 3.1, the primary feature of the QFD process is the House of Quality (HoQ) (Clausing and Hauser, 1988), which provides an inter-functional product planning map to link engineering attributes to customer desires that are ranked in importance. The HoQ utilizes a weighted-sum multi-objective decision criterion, entailing technical test measures (benchmarking) analysis, technical importance rankings, and estimates of technical difficulty to enable a decision maker to set performance targets for a designed artifact. The QFD process has been supplemented by some practitioners with the Pugh Matrix for design concept selection (Terninko, 1997). The Pugh Matrix provides a method to compare alternative design concepts against customer requirements, with evaluations made relative to a base or favored concept, in a process independent from the HoQ analysis.

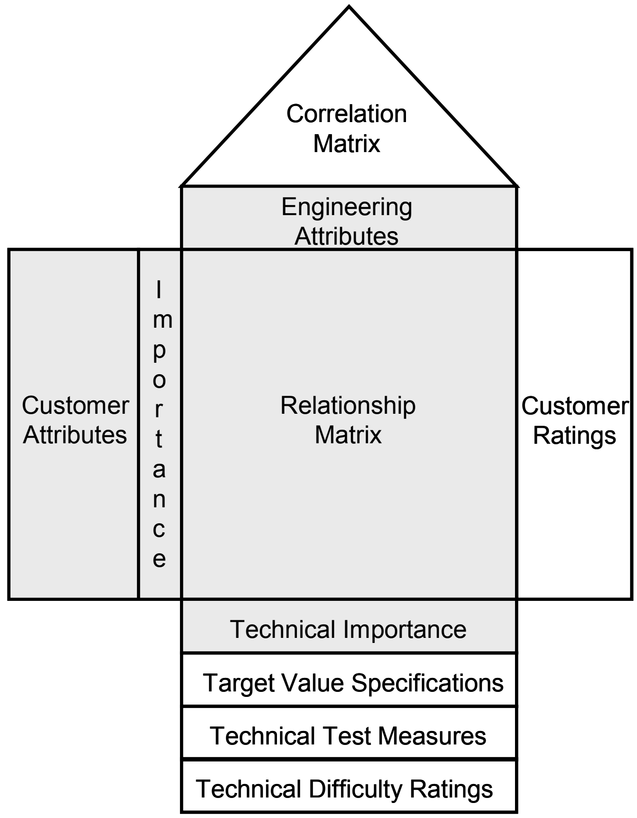


Figure 3.1: House of Quality, 1st House (Olewnik and Lewis, 2005)

3.2 ANALYSIS OF LIMITATIONS IN CURRENT DESIGN PROCESS TOOLS

3.2.1 Issues with QFD Analysis

Much literature has demonstrated both the successes and issues with the QFD methodology (Chan and Wu, 2002). Based on a survey of the literature and analysis conducted for this work, the QFD method suffers from several limitations which can lead to sub-optimal or irrational early product decisions. Firstly, according to Aungst et al. (2003), using only customer and competitor information to set targets without consideration of the physics of engineering attribute interactions or other product objectives, such as choice share and potential profit, can result in targets that can never be achieved in practice. Several proposed improvements to the QFD have been presented in the literature. Aungst et al. (2003) have presented the Virtual Integrated Design Method. Their method uses a quantitative, rather than qualitative, link between the conventional four HoQ matrices, and adds a new 5th house to account for customer perceptual attributes determined using factor analysis. Brackin and Colton (1999) have proposed a method in which analytical relations between the engineering attributes and customer attributes are created, and real values of engineering attributes are searched from an appropriate database to ensure targets are achievable. Locascio and Thurston (1998) have combined the QFD ratings and rankings into a design utility function to determine performance targets using multi-objective optimization. Although these methods improve upon the target setting methodology of QFD, they utilize customer group importance rankings and engineering rankings which have been shown to be potentially problematic (Hazelrigg, 1996).

In the QFD approach, the importance ranking assumes that all customers' preferences are the same and can be represented by a group utility; however, based on Arrow's Impossibility Theorem (AIT), Hazelrigg (1996) has shown that utility exists only at the individual, or disaggregate level. Each customer has a specific preference, and the demand for a product can only be determined by aggregating individual product choices. More recently, van de Poel (2007) has illustrated the methodological problems in the QFD process caused by the implications of AIT. Although the Analytical Hierarchy Process (AHP) was introduced (Armacost et al., 1994) to aid in the determination of importance rankings, Hazelrigg (1996; 2003) has shown through the use of AIT that the importance weightings for ranking the importance of engineering attributes can be irrational when more than two attributes are ordered. Further, Olewnik and Lewis (2005) have demonstrated through the use of designed experiments that the HoQ rating scale used in the relationship matrix yields results comparable to inserting random variables, or completely different scales in its place.

Additionally, due to its philosophy, the QFD method is overly biased towards meeting customers' requirements. Prasad (2000) presented an expanded QFD methodology called Concurrent Function Deployment (CFD) that expands upon the customer attributes to consider other corporate objectives, such as cost and manufacturing. Similarly, Gershenson and Stauffer (1999) developed a taxonomy for design requirements for corporate stakeholders, in the form of a hierarchically organized requirements database. They consider not only end-user requirements as in conventional QFD analysis, but also corporate, regulatory and technical requirements. These methods still employ conventional weighting and ratings techniques.

Although the consideration of uncertainty is imperative in engineering design, particularly in the conceptual design phase, conventional QFD analysis offers only a deterministic approach to ranking importance and setting target performance. It lacks a mathematical framework to incorporate uncertainty into decision-making. Recent work in the QFD methodology has focused on the use of fuzzy set theory to account for uncertainty in consumer importance assessments (Chan et al., 1999; Kim et al., 2000; Kahraman et al., 2006). While these approaches address the uncertainty in the human element of importance assessment, they do not address uncertainty in other elements of the decision process, such as in the technical requirements, or address the limitations of preference aggregation in the QFD method. Other significant limitations of QFD are the over-simplification of attribute-coupling in the "roof" of the HoQ, an inadequate reflection of the real design trade-offs due to the subjective nature of attribute ranking, and a lack of methodology for considering manufacturing/production constraints. Regarding the Pugh Matrix for concept selection, its major limitation is that it is not a comprehensive enterprise-level decision tool, but rather was formulated to make decisions in the engineering domain while considering product requirements, without consideration of uncertainty, customer demand, or enterprise profitability.

3.2.2 Issues with Other Common Design Process Tools

A comprehensive review of other target setting and design selection methods, such as *Taguchi's Loss Function*, *Design for Six Sigma*, and *Suh's Axiomatic Design*, has been conducted by Hazelrigg (2003). To summarize, none of these processes attempt to set targets or select a design concept utilizing an enterprise-level decision criterion. They each assume that product design decisions made using a domain-specific selection criterion, such as minimizing product defects or producing "uncoupled" designs, will result in a preferred design. Another relevant product planning tool is the *Requirements Traceability Matrix* (RTM), which is used to organize and track product requirements to ensure all requirements are met by the design artifact. The *Design Structure Matrix* (DSM) is utilized in systems engineering to decompose a system into components and determine the relationships among components which must be considered in the design process. Neither RTM nor DSM is intended to be used as an enterprise-level product planning tool, but rather each is used to help designers organize product or system requirements.

3.3 Use of the Decision-based Design Framework to Address QFD Limitations

The limitations in the previous section point to the need for a design planning tool which is supported by a rigorous decision-making framework to ensure that consumer preferences are accurately represented and targets set by the tool are achievable in engineering design. The Decision-Based Design (DBD) method, an emerging design paradigm (Lewis et al., 2006), provides such a rigorous framework by modeling design as a decision-making process that seeks to maximize the value of a designed artifact through the use of utility maximization. Recent efforts in DBD research resolve trade-offs among technical objectives by utilizing models of the producer's financial objective, such as net revenue or profit (Hazelrigg, 1998; Li and Azarm, 2000; Wassenaar and Chen, 2003; Michalek et al., 2005). At the core of the enterprise-driven DBD approach is the use of Discrete Choice Analysis (DCA) (Ben-Akiva and Lerman, 1985) for demand modeling to estimate the effect of design changes on a product's choice share, and consequently on the firm's revenues. In Chapter 2, the probabilistic choice modeling approach (DCA) was shown to have the ability to capture the effect of heterogeneous consumer choice behavior upon product design. The disaggregate DCA method takes account of every observed choice situation and the correlation between individual behavior and individual-level conditions and attributes to model choice behavior. A disaggregate DCA approach is more in line with microeconomic theory than an aggregate demand modeling approach and is necessary to understand the heterogeneity of consumer choice behavior (Small, 2006). Although the DBD approach provides a rigorous mathematical framework for decision-making, most formulations of DBD exist as optimization frameworks, not intended to be used as comprehensive process tools to guide real product development activities. The approach has not been applied widely due to the complexity of integrating *product planning* and *engineering product development* into an optimization formulation that incorporates various categories of product design attributes at different levels of abstraction. To manage the complexity of implementing the DBD approach, there is a need to develop a design process tool which effectively guides the execution of the method at an operational level. Such a tool can be tailored for use in the conceptual design phase to fulfill the identified need for a comprehensive product planning tool.

While the flaws associated with the QFD approach limit its use as a quantitative tool for decision-making, the HoQ analysis used in QFD does provide an effective visual tool and promote a rigorous thought process for qualitatively linking product attributes to ensure that product planning activities are conducted while considering the Voice of the Customer. Also, it facilitates a multidisciplinary design process among marketing, engineering, and production. Combining the strengths of DBD and QFD, the Product Attribute Function Deployment (PAFD) method is developed in this work as a comprehensive product planning process tool for the conceptual design phase. It extends the QFD mapping matrix concept to qualitatively identify relationships and interactions while employing the principles of Decision-Based Design (DBD) to provide quantitative assessments for concept selection, attribute target setting, and establishing engineering priority for the detailed product design phase. Our research development leads to a product design tool that overcomes the limitations of the QFD method and facilitates the implementation of the DBD approach for a real design problem.

3.4 The Product Attribute Function Deployment (PAFD) Method

Combining the strengths of the QFD and DBD methods, the PAFD method is developed in this work as a multi–stage process that utilizes two "houses" to establish the qualitative attribute mapping to set engineering priorities, select the preferred design concept, and determine target values, $\mathbf{E}^T$, for the engineering attributes. Because PAFD is intended as a replacement for QFD, a comparison of analogous QFD and PAFD process steps, categorized into three primary stages, is shown in Figure 3.2. In the first stage of both methods, customer preferences are quantified. PAFD uses a DCA model to express consumer *demand* for an entire product relative to the existing competing products, whereas QFD uses a ranking of consumer *preferences* for specific product attributes to assess consumer acceptance of a product as a whole. In the second stage, the engineering design is characterized. PAFD utilizes preliminary analysis models to capture the costs and technical trade-offs among $\mathbf{E}$ (details provided later), versus the technical difficulty rating and correlation matrix mapping used in QFD. PAFD explicitly considers engineering attributes resulting from customer, corporate, and regulatory sources, whereas QFD is primarily focused upon those engineering attributes resulting from customer-desired attributes, **A**.

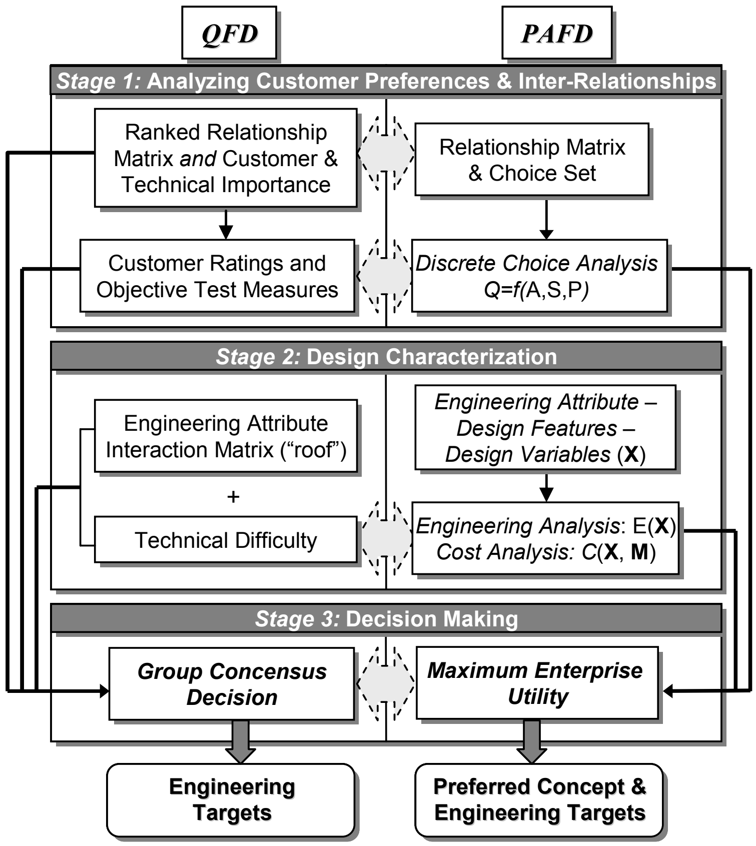


Figure 3.2: Comparison of QFD to PAFD Processes

In the third stage, PAFD provides design decisions in a single-step maximization of enterprise utility formulation, whereas QFD sets priorities using several ratings and rankings which must be synthesized by a human decision maker(s). The following subsections describe the three stages of PAFD in detail, with comparison to equivalent QFD processes.

Stage 1: Analyzing Customer Preferences and Attribute Interrelationships

A "house" structure is used to accomplish the Stage 1 processes of the PAFD method. Similar to conventional QFD analysis is the deployment of mapping between **E** and **A**, as well as the collection of engineering attribute levels from competitors' products (competitive analysis). The engineering attributes determined in this matrix are the **E** related to customer-desired attributes **A**, identified as $\mathbf{E_A}$. Also unique to PAFD, customer demographic attributes **S** and the **A·S** interactions (later transformed to $\mathbf{E_A \cdot S}$ for demand modeling) are identified to account for the heterogeneity of individual customers. This part of the expansion facilitates the construction of the DCA demand model to capture the impact of engineering design (engineering attributes) on customers' purchase behavior through estimation of product demand. As shown in Figure 3.3, House 1 contains two relationship matrices:

- **Matrix 1**: Mapping customer-desired attributes, **A**, to engineering design attributes, $\mathbf{E}_{\mathrm{A}}$.
- **Matrix 2**: Identifying interactions between demographic attributes, **S**, and **A**.

Additionally, a table is provided for tabulating competitive alternatives in the choice set:

- **Table A**: Table of competitive alternatives *J* with corresponding levels of $\mathbf{E}_{\mathrm{A}}$ and price, *P*.

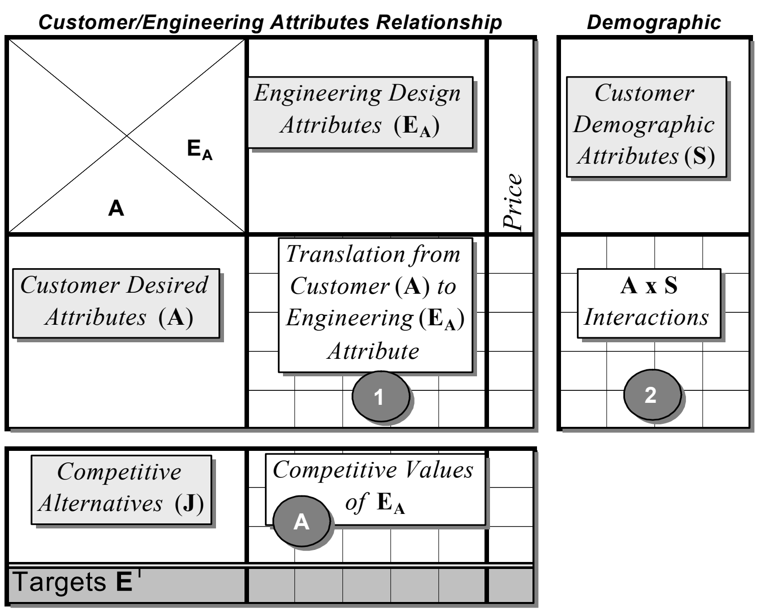


Figure 3.3: House 1 of the PAFD Method

The house can also be extended for use with the Latent Variable modeling approach as shown in Hoyle et al. (2006) by introducing an additional mapping between perceptual customer-desired attributes, **A**, to indicators, **I**, which is not shown in Figure 3.3 for simplicity.

The relationship matrices in PAFD show only the qualitative linking between attributes. Unlike QFD, a rating scale (i.e. 1, 3, 9) is not utilized to characterize the strength of the relationship; however, an "×" is used to indicate the presence of a relationship. The purpose of completing these relationship matrices is to ensure that each of the **A** has a corresponding $\mathbf{E}_\mathrm{A}$ (vector) and that inter-relationships among **A**, $\mathbf{E_A}$, **S,** are clearly identified to enable choice modeling. The "roof", which identifies the coupling of engineering attributes in QFD, has been eliminated in PAFD because engineering attribute interactions will be modeled explicitly using preliminary engineering analyses in Stage 2, to better associate the coupling with a specific design concept. As noted in Section 3.2.1 and illustrated in the case study in Section 3.5, the coupling of multiple engineering attributes, $\mathbf{E_A}$, can largely depend on the chosen design concept, with $\mathbf{E_A}$ coupling in different ways for different design concepts.

The DCA choice model is estimated using the $\mathbf{E}_\mathrm{A}$, P**,** and **S** identified in House 1 as explanatory variables, with J comprising the set of choice alternatives based on competitors' products. Unlike the competitive analysis in QFD (customer ratings), the competitive alternative set used in PAFD is for the purpose of estimating the DCA model as described in Section 2.2. The values of $\mathbf{E_A}$ and P for each alternative together with the consumer product *choice* form the basis for model estimation. The choice set is composed of either actual consumer purchase choices or simulated product choices, such as those resulting from a market survey, as described in Section 2.2. For a market survey, the list of **A** and $\mathbf{E_A}$ can help guide survey construction by providing an indication of the attributes that should be varied among the products presented in the survey (Louviere et al., 2000).

The form of the parameters in the choice model requires insight into customer choice behavior, with potentially several model iterations needed to maximize the model goodness of fit. Linear (e.g. E_i) and transformed (e.g. E_i^2) forms of the variables are explored during the modeling process based upon expected choice behavior. The relationship matrices are used to guide the modeling of $A_i \times S_i$ interactions in terms of the $\mathbf{E}_A$ and $\mathbf{S}$ necessary to make decisions at the engineering design level. *Alternative Specific Constants* (ASC) are utilized to represent preferences that are inherent and independent of specific attribute values. Conversely, *Alternative Specific Variables* (ASV) are utilized to capture the heterogeneity of consumer preference for each alternative due to the differing demographic attributes, **S**, of each consumer (Ben-Akiva and Lerman, 1985).

Stage 2: Design Characterization

This stage of PAFD results in *preliminary* engineering and cost analysis models which are intended to capture the high-level relationship between design concepts and both engineering performance and cost, as opposed to use in creating detailed product designs. The PAFD analyses explicitly consider specific design concepts, whereas the QFD analyses require the design characterization to be carried out at the engineering attribute level, with rankings of technical difficulty and attribute interactions used in place of established engineering and cost analysis methods.

To begin Stage 2, the $\mathbf{E}_A$ established in House 1 become one set of engineering attributes tabulated in House 2 (price, P, is not included in Stage 2 because its value will be determined directly in Stage 3) as shown in Figure 3.4. Unlike QFD analysis which is primarily focused upon the Voice of the Customer, the $\mathbf{E}_A$ form just one subset of the entire set of engineering attributes **E** in PAFD**.** In addition, those attributes which a customer does not consider explicitly in product selection but are essential to producer's interests (Gershenson and Stauffer, 1999), specifically those resulting from corporate $\mathbf{E}_C$, regulatory $\mathbf{E}_R$, and physical requirements $\mathbf{E}_P$, are also identified. This expanded set is essential to ensure all requirements of the design are considered in the decision-making phase to make certain achievable targets are set.

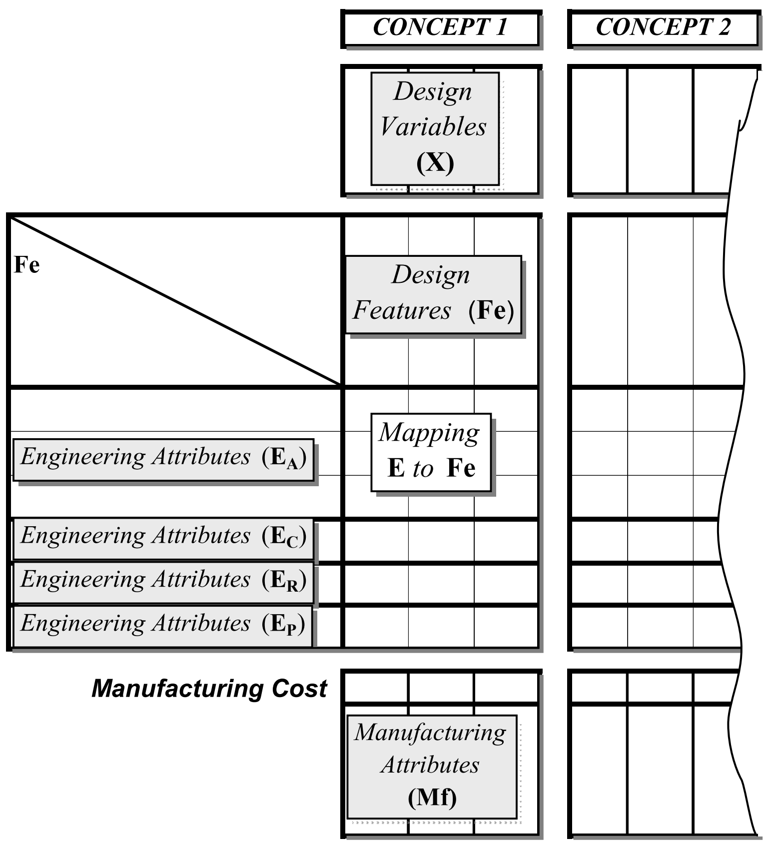


Figure 3.4: House 2 of the PAFD Method

With a comprehensive set of **E** determined and tabulated, designs can now be generated to fulfill these requirements. A *design concept* is defined as a high–level system configuration, composed of multiple subsystems and corresponding key *design features*, **Fe**. To facilitate preliminary cost and engineering analysis, each design feature, Fe_i, is represented by integer, discrete, or continuous *design variables*, **X**, such as material types, dimensions, etc. For each design concept, the attribute mapping in House 2 provides the qualitative relationship between the **E** and **X** through a mapping of **E** to **Fe** as shown in Figure 3.4. From the qualitative relationship, the quantitative functional relationship, $E_i=f(\mathbf{X})_i$, is established using preliminary engineering analysis. In cases where design options are highly conceptual, and an analytical relationship cannot be established, the range of achievable levels of **E** can be estimated. The design variables (**X**) selected *are the minimum, high–level* set necessary to estimate the cost, C_i, of each feature and to represent the coupling of the design features in the decision-making process (Stage 3). The specific form and complete set of the X_{i} will be established in the detailed design process.

After establishing the set of design concepts and specific high-level design features, preliminary *manufacturing process attributes*, **Mf**, are identified for each concept, and mapped to **Fe** (Figure 3.4). The **Mf** are used to estimate processing costs and to identify constraints on **X** resulting from manufacturing process limitations to be considered in the decision-making stage of PAFD, as well as to ensure appropriate manufacturing processes are identified for each design feature. Using the identified **X** and **Mf**, estimation of the total cost, C^{k}, for each design concept, k, is calculated using:

$$C^k(\mathbf{X}^k, \mathbf{Y}, Q, t) = \sum_R C_D^k\left(\mathbf{X}^k, \mathbf{Y}, Q, t\right) + C_C^k(t) + C_F^k(t) \tag{3.1}$$

where $C_D^k\left(\mathbf{X}^k, \mathbf{Y}, Q, t\right)$ is the material and processing cost for each design feature, R is the number of design features, $C_C^k(t)$ is the cost of capital, and $C_F^k(t)$ is fixed corporate overhead cost for each design concept. The reason for establishing both preliminary engineering and cost analysis in PAFD is to capture the real trade-off behavior of engineering attributes, to ensure design selections resulting from the tool are optimal, and target performances are actually achievable.

The number of design concepts considered in PAFD is not fixed, with the house structure repeated for each additional concept to be evaluated. The design concepts and key design features can be generated using several methods available in the literature. Brainstorming and functional decomposition techniques (McAdams et al., 1999; Stone and Wood, 2000) can be utilized to generate the design concepts and corresponding design features, while TRIZ (Theory of Inventive Problem Solving) principles (Altshuller and Williams, 1984) can be employed to aid in the creative process. Optionally, Suh's axiomatic design method (Suh, 1990) can also be employed with PAFD, enforcing an un-coupled or de-coupled relationship between the **E** and **X** and the **Mf** and **X**. While the features of the design concepts can vary significantly, it is assumed that the concepts share a common set of engineering performance (attributes) $\mathbf{E_A}$ that matter to customers.

Stage 3: DBD: Design Concept Selection & Target Setting (Decision-Making)

As shown schematically in Figure 3.2, PAFD evaluates designs through the maximization of expected enterprise utility $E(U)$, using the single selection criterion, V, constructed from the DCA demand (stage 1), engineering, and cost models (stage 2). In addition to selecting a preferred design concept and setting performance targets, PAFD, like QFD, can also aid in setting engineering priority through evaluation of parameter (β) importance in the DCA model and sensitivity analysis of the $E(U)$ function to determine which product attributes should receive the greatest resource allocation during the detailed design phase. In contrast, the evaluation process used by QFD is a (human) group consensus decision, in which the multi-attribute decision criterion requires synthesis of technical importance, technical test measures, technical difficulty, and attribute correlations by the decision maker(s). Engineering targets are set individually for each engineering attribute, based upon the best measured performances from the competing products. This methodology has been shown to be potentially faulty in Section 3.2.1.

Because the preliminary engineering and cost models are used for the purpose of capturing attribute trade-off behavior and are typically analytical expressions, the computational expense of evaluating such models, and hence the expense of the PAFD design selection method, is minimal. Additionally, the $\mathbf{X}$ can often be represented by discrete values in the conceptual design phase, for example representing catalog component options (Bradley and Agogino, 1994). The maximization of utility can be evaluated using a genetic algorithm which is commonly used in conceptual design selection when a combination of discrete and continuous design variables are present (Goldberg, 2002). Constraints are of the form $g(\mathbf{X}, \mathbf{E}) \leq 0$, and are estimated for each design concept based upon corporate, regulatory, physical, and manufacturing constraints upon the $\mathbf{X}$ and $\mathbf{E}(\mathbf{X})$ identified in House 2.

3.5 Automotive Sensor Case Study

The conceptual design of an automotive pressure sensor is used as a case study to demonstrate the PAFD methodology. The specific example considered is to design a standard next-generation Manifold Absolute Pressure (MAP) sensor for the automotive industry. The MAP sensor measures the air pressure in the intake manifold for fuel and timing calculations performed by the engine computer. The customers are *industrial customers*, composed of both automobile manufacturers and engine system sub–suppliers. The targeted market is the mid-size sedan segment. A high level function diagram of a MAP sensor is shown in Figure 3.5.

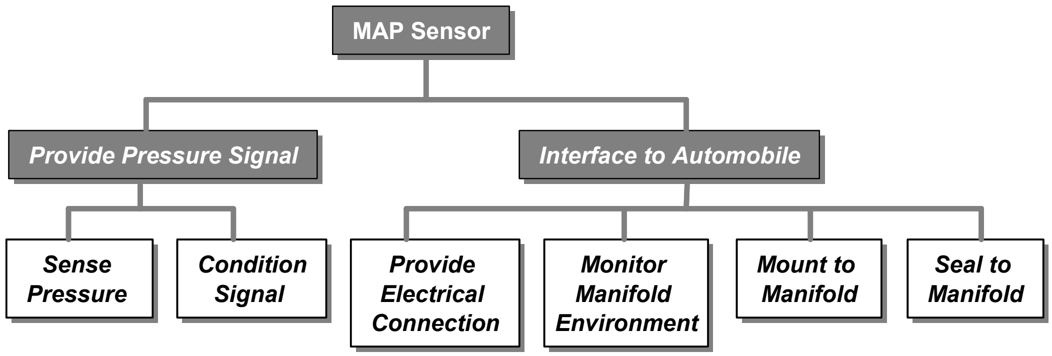


Figure 3.5: MAP Sensor Functions

Multiple sensing technologies exist for pressure measurement, and each technology drives specific corresponding high-level design features, resulting in differing levels of performance and cost structure for each design concept. Therefore, before detailed design of the sensor begins, the preferred design concept must be selected and target levels of product performance must be established. A *risk-averse* attitude is assumed for the enterprise, and the market size is assumed to grow by 10%/yr. over the time interval, t, of 4 years considered in the forecast. Both a QFD and PAFD analysis are conducted to better illustrate the parallels between the two methodologies, with the differences in the resulting design decisions demonstrated.

3.5.1 QFD Analysis of MAP Sensor

To begin the QFD analysis, the **A** (e.g. *High Accuracy* and *Withstand Temperature Extremes*) and the key $\mathbf{E_A}$, such as *Housing Footprint* (mm^2) and *Temperature Range* (°C), are placed in the appropriate rows and columns of the HoQ as shown in Figure 3.6.

Engineering Attribute (E) / Customer Attribute (A)	Importance	– Sense Element Accuracy	+ Pressure Span	+ Temperature Range	– Housing Footprint	+ Natural frequency housing	– Connector Mating Force	Customer Ratings 1 (L) ⟷ 5 (H)			
High Accuracy	8	9	3					*A*	*C*	*Ours*	*B*
Wide Pressure Range	6		9					*A*	*B*	*Ours*	*C*
Withstand Temperature Extremes	6			9				*C*	*Ours*	*A*	*B*
Small dimensional footprint	4		1		9		1	*C*	*A*	*B*	*Ours*
Withstand High Vibration	4			1	3	9	1	*A*	*Ours*	*C*	*B*
Low Connector Mating Force	2						9	*Ours*	*C*	*A*	*B*
Technical Test Measures: Sensor A		2.0	80	145	17.7	1400	35				
Technical Test Measures: Sensor B		1.0	185	180	16.4	2100	25				
Technical Test Measures: Sensor C		1.5	250	140	26.0	1700	35				
Technical Test Measures: *Our*		1.2	240	150	14.6	1600	40				
Technical Difficulty		8	4	3	5	7	2				
Technical Importance		*72*	*82*	*58*	*48*	*36*	*26*				
Targets		**1.0**	**250**	**180**	**14.6**	**1600**	**35**				

Figure 3.6: Comparison QFD Analysis of MAP Sensor

The engineering team must rank-order the importance of each **A**, fundamentally establishing a "group utility" for each attribute as described previously, and determine a "direction for improvement" for each of the $\mathbf{E_A}$ based on engineering judgment, as shown by the "+" and "–" signs preceding each $\mathbf{E_A}$. The relationship matrix is then completed, with the engineering team determining the strength of relationship between the $\mathbf{E_A}$ and **A**, using a largely subjective evaluation based on the experience level of the team members. With the relationship matrix complete, the *Technical Importance* is calculated for each E_A to determine engineering priority for each attribute, with a higher importance rating indicating higher engineering priority. The "roof" *Correlation Matrix* is completed, with ✓ indicating positive correlation and ✖ negative correlation between attributes (e.g. negative correlation indicates a performance improvement to one attribute degrades the performance of another attribute), and the *Technical Difficulty* rating is estimated (higher number indicates greater difficulty). These analyses can be viewed as highly simplified, empirical forms of the engineering and cost analyses explicitly formulated in the PAFD method.

To complete the *Customer Ratings*, a market study (Stated Preference) is conducted in which several customers are surveyed to determine consumer perceptions of current competitive MAP sensors on the market. The respondents are asked to rank-order the performance of three competitive sensors (labeled A, B, C), plus the current generation sensor (*Our*), with respect to each **A** they have identified, with the ranking results shown in Figure 3.6. For example, the customer group evaluation for *High Accuracy* indicates that Sensor B is perceived as having the highest accuracy and Sensor A the lowest accuracy. Note that with QFD, the customer ranking must be aggregated in order to achieve a single rank-order for each **A**, a process shown to be potentially problematic (Hazelrigg, 1996). To complete the QFD analysis, the actual measured performance level of each engineering attribute is determined for each of the four sensors and documented in the *Technical Test Measures* portion of the HoQ.

With the HoQ completed, performance targets for the sensor are determined through a multi-attribute consideration of the *Technical Test Measures*, *Customer Ratings*, *Technical Difficulty*, and *Correlation Matrix*. The performance target decision is made relative to the current levels of performance of *Our* sensor, with the values identified in the *Technical Test Measures* representing the best known levels of performance for each **E** which should be targeted by the new sensor**.** The *Technical Difficulty* and *Correlation Matrix* provide subjective constraints upon performance. Using the QFD methodology, the targets are shown at the bottom of the HoQ in

Figure 3.6. It was decided that the new sensor should have improved target performances for *Accuracy*, *Pressure Span*, and *Temperature Range*, since these have high technical importance, and *Our* current sensor is not perceived as the market leader in these areas. Also, it was decided to improve the target for *Connector Mating Force* since it has a very low technical difficulty. It was decided not to improve the target for *Housing Footprint*, since *Our* sensor is the market leader, or *Natural Frequency* due to high technical difficulty and low technical importance. With targets set, product design concepts may be further brainstormed by an engineering team, and the preferred concept selected with a tool external to the HoQ, such as Pugh's Method.

3.5.2 PAFD Analysis of MAP Sensor

Stage 1: Understanding MAP Sensor Requirements and Interrelationships

As the first step in PAFD, key customer-desired attributes $\mathbf{A}$ and engineering attributes $\mathbf{E}_A$ are placed in the appropriate rows and columns in the same manner as the QFD analysis (Figure 3.7). In contrast to QFD analysis, demographic attributes $\mathbf{S}$ (e.g. *Vehicle Market Segment*) are also identified and tabulated. Note that the $\mathbf{S}$ for the industrial customers are company-specific attributes, such as the corporate location or the specific market niche in which the company competes. As described in Section 2.2, the $\mathbf{S}$ account for the heterogeneity of customer choice, *i.e.* they explain why different customers choose different MAP sensors for similar applications. With $\mathbf{A}$, $\mathbf{E}_A$, and $\mathbf{S}$ identified, hypothesized relationships are marked by an "×" in matrix 1 identifying the linking of the $\mathbf{E}_A$ to $\mathbf{A}$, and in matrix 2 identifying the potential interactions among the $\mathbf{S}$ and $\mathbf{A}$ which influence choice behavior, such as the interaction of *High Accuracy* and *Vehicle Market Segment*.

Customer/Engineering Attributes Relationship								Demographic			
DCA Variables	AC	PS	TR	FT	NF	MF		AR	ER	MS	
Engineering Attributes (E_A) / Customer Desired Attributes (A)	Element Accuracy (% Error)	Pressure Span (kPa)	Temperature Range (°C)	Housing Footprint (cm^2)	Natural frequency housing (Hz)	Connector Mating Force (N)	***Sensor Price ($)***	Vehicle Origin: Asia	Vehicle Origin: Europe	Vehicle Market Segment	*Interaction Parameters* (β_3)
High Accuracy	×	×						×	×	×	∗
Wide Pressure Range		×								×	**0.3**
Withstand Temperature Extremes			×		×						
Small dimensional footprint				×							
Withstand High Vibration				×	×						
Low Connector Mating Force						×					
Sensor A	2.0	80	145	17.7	1400	35	10.0	Base	Base	Base	Base
Sensor B	1.0	185	180	16.4	2100	25	10.2	**7.4**	**7.0**	**7.3**	**-6.4**
Sensor C	1.5	250	140	26.0	1700	35	12.4	**8.0**	**9.5**	**12.8**	**-6.7**
Our	*1.2*	*240*	*150*	*14.6*	*1600*	*40*	*11.5*	**7.1**	**7.9**	**-2.4**	**-5.6**
DCA Parameters (β_1)	**-2.8**	∗	**0.02**	**-0.4**	**-0.001**	∗	**-6.9**	*ASV* (β_2)	*ASV* (β_2)	*ASV* (β_2)	*ASC*
Normalized DCA Parameters (β_1)	-5.6	∗	2.9	-9.6	-3.1	∗	-85.6				
Targets E^T											

Figure 3.7: PAFD House 1 for MAP Sensor

To acquire the choice data necessary to estimate the DCA model, simulated sensor purchase data (Revealed Preference) was utilized, unlike the QFD analysis in which respondents were asked to rank-order the performance of each sensor for each **A**. The purchase data for *Our* sensor represents the current generation sensor on the market; alternatively, a *Stated Preference* survey (Louviere et al., 2000) could be conducted using prototypes of the new sensor design, if desired. The demographic data **S** for each customer in the data set is recorded in the PAFD method. A sample of the purchase and recorded demographic data **S** is shown in Table A.1, Appendix A. A MNL DCA model is formulated as a function of the values of $\mathbf{E}_A$, P, and **S** using the choice data collected for the four sensors. The model parameters (β) estimated to create a choice model with good fit statistics are composed of linear (e.g. *Accuracy*, *Temperature Range*), interaction (e.g. *Accuracy* × *Vehicle Market Segment*) and alternative specific variables (e.g. $Alternative_j$ × *Vehicle Market Segment*), with alternative specific constants included to capture inherent preferences for each alternative. The results are shown in House 1 (Figure 3.7), which has been extended from the template shown in Figure 3.3 to include a summary of the β parameters in the grey region (note that not all E_A enter W as indicated by a ✳, as some parameters are not statistically significant or are highly correlated with other E_A). Referring to Eq.(2.5), the β parameters establish the customer choice utility function, W, of each alternative. In particular, each alternative shares a common set of product selection attribute parameters, which form the *common* customer choice utility function:

$$W_{Common} = -2.8(AC_i) + 0.02(TR_i) - 0.4(FT_i) - 0.001(NF_i) - 6.9(PRICE_i) + 0.3(PS_i \times MS). \quad (3.2)$$

The specific customer choice utility functions for each of the competitive alternatives is then determined for use in Stage 3, using the common utility formulation with the addition of the appropriate alternative specific constants (ASC) and variables (ASV):

$$\begin{aligned} W_{\mathbf{A}n} &= (W_{common})|_{i=1} \\ W_{\mathbf{B}n} &= -6.4 + (W_{common})|_{i=2} + 7.4(AR) + 7.0(ER) + 7.3(MS) \\ W_{\mathbf{C}n} &= -6.7 + (W_{common})|_{i=3} + 8.0(AR) + 9.5(ER) + 12.8(MS) \end{aligned} \quad . \quad (3.3)$$

A customer choice utility function is also developed for *Our* sensor design:

$$W_{\mathbf{OUR}n} = -5.6 + (W_{common})|_{i=4} + 7.1(AR) + 7.9(ER) - 2.4(MS). \quad (3.4)$$

Examination of the utility function provides insight into customer choice behavior. The sign of the parameter indicates the effect of an attribute upon W, for example increasing the *Price* (β = -6.9) of a sensor decreases W, and hence the probability of choice of that sensor, *ceteris paribus*. Additionally, the effect of **S** upon utility can also be examined. For example, W and hence the probability of choice of Sensors B, C and *Our* increases relative to the reference (Sensor A) if the customer is located in Asia (*AR*) or Europe (*ER*); the greatest increase in *W* is for Sensor C as indicated by the magnitude of the β parameters for *AR* (β = 8.0) and *ER* (β = 9.5) in the $W_{\mathbf{C}n}$ expression. To understand the engineering priority of each $\mathbf{E_A}$ and $\mathbf{E_A}\times\mathbf{S}$ in terms of their impact on demand, the β coefficients can be normalized as shown in Figure 3.7 to allow the importance of each attribute to be estimated based upon its magnitude. For example, *Price* is the most important attribute (β_{NORM} = -85.6) while *Temperature Range* is the least important (β_{NORM} = 2.9).

With a customer choice utility function available for each alternative, Eq. (2.5) can be utilized to determine the demand for the new design concepts based upon the values of $\mathbf{E_A}$ and *P* substituted into Eq. (3.4) during the decision-making phase in Stage 3.

Stage 2: MAP Sensor Design Concepts Identification and Characterization

Stage 2 begins by transferring the $\mathbf{E}_\mathrm{A}$ identified in House 1 to the **E** Column in House 2, Figure 3.8, and establishing the additional engineering design attributes derived from corporate, regulatory, and physical requirements, such as *Common Platform* as $\mathbf{E}_\mathrm{C}$, *UL Flammability Resistance* as $\mathbf{E}_\mathrm{R}$, and *Housing Stress* as $\mathbf{E}_\mathrm{P}$, to form the complete set of **E**. With **E** identified, design concepts and their corresponding design features **Fe** can be formulated. For this problem, two design concepts were identified: *Concept 1* utilizes a piezoresistive (PRT) sensing element with a micro–machined sensing diaphragm, which senses pressure due to bending of the diaphragm, and *Concept 2* utilizes a two–plate capacitive sense element, which senses pressure due to a change in the capacitor plate separation distance.

		DESIGN CONCEPT 1					DESIGN CONCEPT 2					
	Design Variables (X)	Length, Thickness	Resolution, Error (T)	Width, Length, Polymer Mating Force	Thickness, Material	Thickness, Material	Area, Separation	Resolution, Error (T)	Width, Length, Polymer	Thickness, Material	Mating Force	Thickness, Material
	System Engineering Attributes (E) \\ Component Design Features (Fe)	Piezoresistive Sense Element	Digital Signal conditioning IC	Integral housing/USCAR connector	Steel Housing Reinforcement	Rubber Pressure Seals	Capacitive Sense Element	Digital Signal conditioning IC	Polymer Housing	Steel Housing Reinforcement	USCAR Connector	Rubber Pressure Seals
E_A	Sense Element Accuracy (%)	×	×				×	×				
E_A	Pressure Span (kPa)	×	×			×	×	×				×
E_A	Temperature Range (°C)	×		×			×		×			
E_A	Housing Footprint (cm^2)			×					×			
E_A	Natural frequency housing (kHz)				×					×		
E_A	Connector Mating Force (N)			×							×	
E_C	Common Platform	×	×				×	×			×	
E_R	UL Flammability Resistance			×					×		×	
E_T	Housing Stress (kPa)				×					×		
	Estimated Manufacturing Cost	**$0.25**	*	**$1.00**	**$0.10**	**$0.20**	**$0.40**	*	**$0.40**	**$0.10**	**$0.55**	**$0.20**
	Manufacturing Attributes (Mf)	*Micromachining*	** Catalog*	*Insert Molding*	*Stamping*	*Transfer Molding*	*Micromachining*	** Catalog*	*Injection Molding*	*Stamping*	*Insert Molding*	*Transfer Molding*

Figure 3.8: PAFD Engineering Design House 2 for the MAP Sensor

Both design concepts are shown in Figure 3.9. Due to differences in the designs of the sensing elements, the piezoresistive concept is inherently less expensive and results in a smaller package, whereas the capacitive concept is more robust to temperature and pressure extremes.

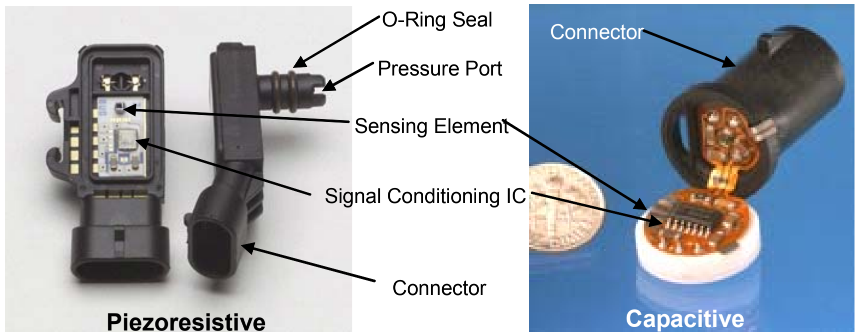


Figure 3.9: Comparisons of Concepts 1 and 2

The key design features for each concept are established and the corresponding high-level design variables, **X**, to model the technical trade-offs and cost for the decision-making problem are determined and tabulated (Figure 3.8). For example, piezoresistive sense element *thickness* is a continuous variable to be determined based upon the trade-off among element length, manufacturing limitations, and cost; integrated circuit A/D discretization *resolution* is a discrete variable to be determined based upon the trade-off between sensor accuracy and cost. Key conceptual manufacturing processes, **Mf**, (e.g., micro-machining, injection molding, etc.) are identified for each design concept, and placed in the columns corresponding to the associated design feature, **Fe**, shown in Figure 3.8. Manufacturing process costs are also estimated for each design feature for use in the cost model (Eq. (3.1)).

As demonstrated by this case study, the technology selection drives specific design features and the corresponding set of design variables for a given design concept. For example, the packaging of each sensor is fundamentally different as shown in Figure 3.9: Concept 1 uses an injection–molded housing with integral pressure port and connector, whereas Concept 2 requires a separate port and connector component because of the large size and electrical interconnect of the capacitive element.

Also noted, each set of high-level design variables **X** for a given concept has a different functional relationship with **E** (**E**=*f*(**X**)). Concept 1 utilizes the piezoresistive sensing element with a resistance output given by the relation (Hauptmann, 1993):

$$Pressure\ Span = k(\Delta L_E / L_E) \tag{3.5}$$

where the engineering attribute is *Pressure Span*, the design variable is diaphragm length L_E, and the piezoresistive k-factor, k, is a constant. Concept 2 utilizes a capacitive output given by:

$$Pressure\ Span = \varepsilon_0 \varepsilon_r \left(A_E / \Delta D_E \right) \tag{3.6}$$

where the engineering attribute is *Pressure Span*, the design variables are the plate area, A_E, and the plate separation distance, D_E, with absolute and relative dielectric constants, ε_0, and ε_r. A list of all engineering relations used in this analysis is provided in Appendix A, Table A.2. These analyses are intended to be preliminary analyses to capture the fundamental trade-offs among the critical design variables, and will be refined during the detailed design phase to enable final design of the sensor.

Each concept requires a specific manufacturing process, and the different sets of **Mf** result in a differing cost structure and place different constraints upon the **X**. For example, the micro-machining process used to manufacture the diaphragm of the piezoresistive sense element results in a minimum diaphragm thickness limitation, and hence, places a constraint on the minimum size of the sense element, independent of engineering analysis. Also confirmed by this study is that engineering design attributes, **E**, resulting purely from customer-desired attributes, **A**, are not sufficient to create an engineering specification (target setting). For example, consideration of the stresses induced by the manufacturing process on the sensor housing leads to a key constraint upon the sensor housing design which would not have resulted from customer-desired attributes.

Stage 3: Design Concept Selection and Target Setting

Stage 3 of PAFD is conducted by formulating the decision-making problem as shown in Table 3.1.

Table 3.1: Pressure Sensor Decision-Making Formulation

Given
Mid-Size Sedan Market Size: 1,000,000 [sensors/year] for 4 years
Demographic data of targeted industrial customers **S**

Engineering Attributes $\mathbf{E_A}$ (PAFD: House 1)
$\mathbf{E_A}$ determined as a function of the high-level design options **X** (**E**(**X**))

Design Concept (PAFD: House 2)
Two (2) Design Concepts considered (piezoresistive & capacitive sensing)

Sources of Uncertainty Y

DCA Model Parameters	S.E. of β
Cost Estimates	C^1=±10%, C^2=±30%
Normal Distribution of T_E and D_E	$\sigma = (0.1)\,\mu$

Cost Model (PAFD: House 2)
Cost of each alternative given by Eq.(3.1).

Demand Model *Q* (PAFD: House 1)
Obtained from the MNL model of the competitive alternative attribute data.

Single criterion $V = QP$-C (Eq.(2.2))

FIND:
Design Variables **X**, Target Engineering Levels $\mathbf{E^T}$ (PAFD: House 1) and Price *P*

MAXIMIZE:
$E(U)$, assuming an enterprise risk-averse attitude (Eq. (2.3))

SUBJECT TO (PAFD: House 2):
$g(\mathbf{X}, \mathbf{E}) \le 0$ $T_E - 14.0 \le 0$; $D_E - 12.0 \le 0$: **Constraints from Mf**
$g(\mathbf{X}, \mathbf{E}) \le 0$ $PS - 80.0 \le 0$; $NF - 1400.0 \le 0$: **Constraints from $\mathbf{E_C}$ and $\mathbf{E_P}$**

Three types of uncertainty are considered in the selection process:

- *Demand Model Uncertainty*: Uncertainty in all DCA model parameters (i.e. βs), as quantified by the standard error (S.E.) estimates, is considered.

- *Cost Estimation Uncertainty*: Because the costs are estimated, uncertainty in the estimates must be considered. It is assumed that the cost estimates for Concept 1 have ±10%

error, while estimates for Concept 2 have ±30% error, since it is assumed that the designers are more familiar with the design and costs of Concept 1.

- *Design Variable Uncertainty*: The Piezoresistive Sense Element Thickness, T_E, and Capacitive Sense Element Plate Separation Distance, D_E, are normally distributed random variables due to known variation in the element manufacturing processes.

These uncertainties create risk in the decision process. The preferred concept, considering uncertainty, depends upon the decision-maker's (i.e. the enterprise) risk attitude. The risk attitude assumed by the enterprise in this case study is moderately risk-averse. However, the decision-making process will be demonstrated for risk-averse, risk-neutral, and risk-seeking attitudes.

3.5.3 Comparison of PAFD and QFD Results

The results of the PAFD decision process are shown graphically in Figure 3.10 a) in which the full distributions of profit for both concepts are shown. Figure 3.10 b) illustrates the expected utility of each concept considering a variety of enterprise risk attitudes. The risk attitude is modeled using an exponential utility function in which higher relative risk tolerance indicates increasing risk-seeking.

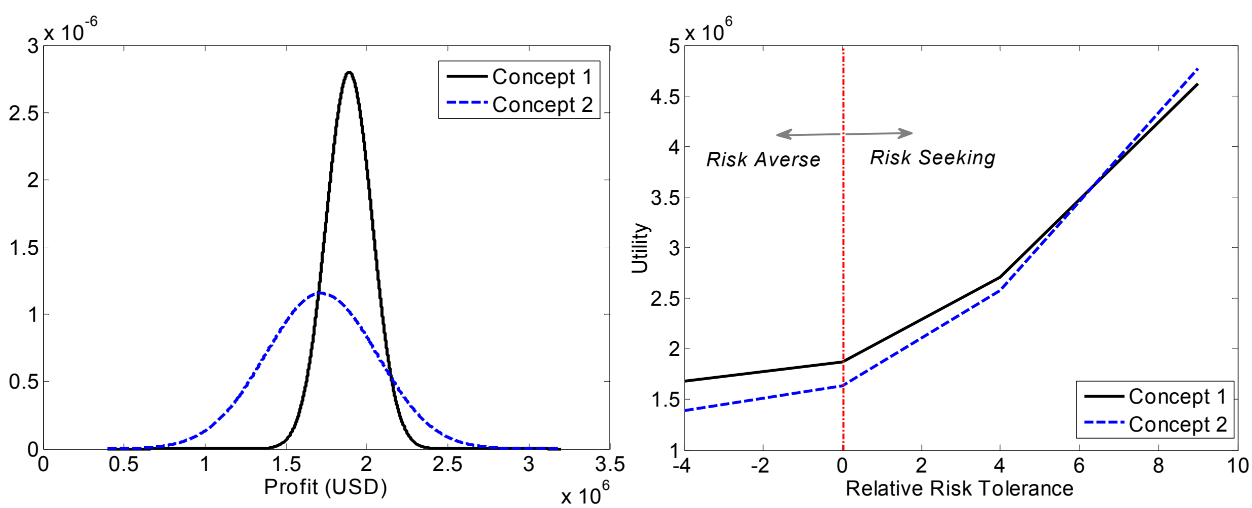


Figure 3.10 a) and b): Comparison of Profit and Utility for Concepts 1 and 2

As demonstrated, Concept 1 is preferred for risk-averse, risk-neutral, and moderate risk-seeking attitudes. However, Concept 2 is preferred for a high risk-seeking attitude, since the greater uncertainty in Concept 2 results in a higher upside potential than Concept 1.

The results of both the PAFD and QFD analyses corresponding to a moderate risk-averse attitude are shown in Table 3.2. The PAFD decision results in performance targets $\mathbf{E}^T$, and values of demand, price, and cost for both Concepts 1 and 2. The preferred design concept for this problem is *Concept 1*, which results in the highest utility for the enterprise considering uncertainty ($E(U)$ = 1,683,000 *utils*). The QFD analysis results in performance targets only, which are not associated with a design concept, and additionally QFD has no mechanism for determining price P. For the purpose of comparison, the unit price of the QFD design is set at the same price ($10.40) as Concept 1 (the preferred design from the PAFD method) and profit and utility estimated using this price.

Table 3.2: Comparison of Decision Results–Preferred Concept (shaded)

	PAFD **($\mathbf{E}^T$)**		***QFD***
Engineering Attribute E	**Concept 1**	**Concept 2**	**($\mathbf{E}^T$)**
Sense Element Accuracy (%)	1.23	1.19	1.0
Full Scale Span (kPa)	176.0	185.0	250.0
Temperature Range (°C)	150.0	150.0	180.0
Housing Footprint (cm^2)	16.9	17.2	14.6
Natural frequency (Hz)	1400.0	1425.0	1600.0
Connector Mating Force (N)	40.0	40.0	35.0
Q: Demand/year (*# sensors*)	**416,000**	**433,000**	541,000
P: Unit Price (*USD*)	**$10.40**	**$10.58**	$10.40
C: *Unit Cost* (*USD*)	**$9.25**	**$9.66**	$10.32
Expected Profit (USD)	**$1,905,000**	**$1,706,000**	$173,000
Expected (*U*) (*utils*)	***1,683,000***	***1,393,000***	*170,000*

Compared to the PAFD results, the QFD identifies targets based upon the best values of $\mathbf{E}_A$ identified in the competitive analysis, which subsequently leads to a lower value of $E(U)$ of 170,000 *utils*. The reason the QFD resulted in such low enterprise utility is that although the estimated demand, *Q*, for a sensor meeting the targets set by QFD is higher than those identified by PAFD, the cost to make such a sensor is significantly higher ($10.32). As described in Section 3.2.1, QFD is biased toward meeting customer product desires and does not explicitly consider cost, leading to a sensor design with good customer acceptance potential but low expected enterprise utility. Additionally, because parameter relationships identified through engineering analysis and constraints determined in the PAFD Stage 2 process are not utilized, it is not known with confidence if these QFD targets can actually be achieved in the subsequent sensor design. For the PAFD analysis, the target levels identified for the preferred concept reflect the actual achievable levels of $\mathbf{E}_A$ which maximize enterprise utility for this design concept, based upon the constraints imposed in the decision-making problem. This is further illustrated by noting that Concept 2 has different values of $\mathbf{E}^T$ corresponding to the maximum enterprise utility for that particular concept.

To set engineering priority using the PAFD analysis, a global sensitivity analysis (Chen et al., 2005) is conducted as recommended previously to study the total effect of individual engineering attributes on the *E(U)*. The results of this analysis indicate that the greatest resource allocation should be made to achieving the targets for *Housing Footprint* and *Pressure Span*, due to the sensitivity of enterprise utility to these parameters. For QFD, the *Technical Importance* measure is used to establish engineering priority, resulting in selection of *High Accuracy* and *Pressure Span* as the highest priority. The difference in priority results from the different focuses of the two tools, with PAFD focused upon maximizing enterprise utility and QFD focused primarily upon customer product acceptance. In summary, the PAFD method has provided a clear conceptual direction and engineering targets necessary to begin the detailed design of the MAP sensor; detailed engineering analysis can be utilized to create the specific feature designs which meet these targets.

3.5.4 Validation and Discussion of the PAFD Method

A primary feature of the PAFD is the use of a DCA model to predict customer demand for a design option. The choice model was validated using a cross-validation method, in which the data is partitioned into training and test sets (Tamhane and Dunlop, 2000). The results show approximately a 5-10% error in predicting the choice share on the test sets; however, such errors equally effect predictions for both Concepts 1 and 2 and do not change design selection result. To explore the effect of demand model specification on the selection process, a model was used in the process which did not include country (i.e. Asia, Europe) attributes. For such a model, cross-validation indicated greater errors on the test sets (10-18% errors) and did result in higher predicted choice share and approximately 8% higher profit for Concept 1 and 10% higher profit for Concept 2; however, the selection process was not effected as Concept 1 is still preferred for risk-averse, risk-neutral, and moderate risk-seeking attitudes using this model.

The case study presented demonstrates the advantages of the proposed PAFD method. It has been shown to preserve the primary strengths of QFD by offering a visual tool, maintaining ease of use, and promoting team work. The DCA model presented is becoming commonplace in the marketing discipline (Rossi et al., 2005) with commercial software solutions available (Sawtooth Software, 1999), while the engineering and cost modeling are standard practice among design and manufacturing engineers. The method can be expanded beyond the three stages shown in Figure 3.2, for example to include a specific stage for the design of choice experiments. The method can also find application in the service industry, to design the service to best meet the needs of both the customers and the enterprise providing the service.

The PAFD method can identify a preferred solution in situations in which cost and performance models can be formulated, and the risk attitude of the enterprise can be formalized. A potential limitation of the PAFD method is that it is based on the assumption that a DCA choice model can be estimated to represent consumer preferences. This assumption would not be valid in cases where the customer is a single or small group of customers (such as for a component sub-supplier of a major system), for highly specified designs, or in industries which do not have the infrastructure for consumer preference data collection.

3.6 DISCUSSION AND SUMMARY

In this work, the Product Attribute Function Deployment (PAFD) method is developed to offer a mathematically rigorous, decision-theoretic process tool for use during the product planning phase of a product development program. The need for developing such a method results from a close examination of the needs during the conceptual design phase, and the limitations of current methods, such as QFD, currently used for this purpose. The PAFD method extends the QFD mapping matrix concept to qualitatively identify relationships and interactions among product design attributes while employing the DBD principles to provide rigorous quantitative assessments for design decisions. In conceptual design, the PAFD method is used to select the preferred design concept, set target levels of engineering performance, and set engineering priorities. The PAFD method can be implemented, with minor modification, to work with alternative enterprise-driven design approaches to provide the necessary quantitative assessments.

In addition to presenting the PAFD method, a comprehensive comparison of QFD and PAFD was conducted in this work, demonstrating the parallels between the two methods and the improvements achieved by utilizing DBD principles in the new tool. The use of single-objective utility maximization provides a rigorous mathematical framework for decision-making under uncertainty, alleviating the difficulties associated with weighting factors and multi-objective decision-making in QFD. The use of profit as a single criterion better captures the real design trade-offs, incorporating the needs from both the producer and consumer to set engineering targets consistent with enterprise objectives. The heterogeneity of consumers is captured through the inclusion of demographic attributes, **S**, in the DCA model, addressing the aggregation issues present in QFD. The subjective ratings and rankings present in QFD are replaced with established methodologies in engineering, cost, and decision analysis to set targets for performance which can be achieved in practice. Uncertainty is explicitly addressed through the use of expected enterprise utility as the decision criterion.

A case study involving the conceptual design of a Manifold Absolute Pressure (MAP) sensor is used to illustrate the benefits of the PAFD method. Complex trade-offs among engineering, manufacturing, and customer considerations which would result in a difficult synthesis and subsequent decision-making process using QFD are resolved effectively using the PAFD approach. While the PAFD method has been demonstrated as a method for design concept selection, it provides a general design process tool that can be utilized throughout the design process, such as the vehicle target setting case study of Chapter 6. The simple choice model presented here, in which it was assumed that the mapping from qualitative customer-desired attributes to engineering attributes is straightforward, can be replaced with the Bayesian Hierarchical Choice Model of Chapter 6 for a complex system. Also, it is assumed in this study that a single data set is available for choice model estimation, an assumption that is not realistic for a complex system such as a vehicle. Methods for acquiring data are presented in Chapter 4, methods for preparing the data for model estimation are presented in Chapter 5, and methods for model estimation with multiple data sets are presented in Chapter 6.

Chapter 4
OPTIMAL DESIGN OF EXPERIMENTS FOR HUMAN APPRAISALS

In Chapter 3, the PAFD method for making engineering design decisions was presented. A key issue not addressed in that chapter was the acquisition of the data needed to estimate the choice model. For the Bayesian Hierarchical Choice Model introduced in Chapter 1, preference data is required to estimate the hierarchy of discrete choice and ordered logit models. This highlights a general need for the development of a standardized approach for designing experiments to access consumer preferences, using ***Human Appraisal*** experiments. Human appraisals are used to assess consumers' opinions of a given product design, and are unique in that the experiment response is a function of both the product attributes **E** and the respondents' human attributes **S**. In this work, the design of a human appraisal is characterized as a split-plot design, in which the respondents' human attributes form the whole-plot factors while the product attributes form the split-plot factors. The experiments are also characterized by random block effects, in which the design configurations evaluated by a single respondent form a block. An experimental design algorithm is needed for human appraisal experiments because standard experimental designs often do not meet the needs of these experiments. In this chapter, an algorithmic approach to identify the optimal design for a human appraisal experiment is developed, which considers the effects of respondent fatigue and the block and split-plot structure of such a design. The developed algorithm seeks to identify the experimental design which maximizes the determinant of the Fisher Information Matrix, labeled the *D*-criterion of a given design. The algorithm is derived assuming an ordered logit model will be used to model the rating responses. The advantages of this approach over competing approaches for minimizing the number of appraisal experiments and model-building efficiency are demonstrated using an automotive occupant package human appraisal as an example.

The chapter is organized as follows: Section 4.1 provides the definition and challenges in human appraisal experiments, Section 4.2 provides background for the DOE and modeling methodology, Section 4.3 presents the experimental design methodology for human appraisals, Section 4.4 discusses implementation of the methodology, and Section 4.5 provides a case study.

4.1 INTRODUCTION

Human appraisal experiments are used in a variety of contexts in product design to elicit consumer feedback on current or future product designs. The link between consumer preferences and engineering design has received much attention in the literature recently (Li and Azarm, 2000; Besharati et al., 2002; Wassenaar and Chen, 2003; Petiot and Yannou, 2004; Michalek et al., 2005; Wassenaar et al., 2005; MacDonald, 2007). Such design approaches have created the need for methods to assess human preferences for hypothetical or actual product designs to enable the desired linkage between consumer preferences and engineering design. In Chapter 2, a hierarchical choice modeling approach was introduced in which a hierarchy of customer preference models is used to estimate consumer preferences for a given system design. Such an approach requires the collection of customer opinion for given system and sub-system designs. These product designs are generally represented by prototype hardware for human appraisals, more recently by highly flexible, programmable ***hardware-in-the loop*** (Wang et al., 2006), which can assume a wide array of unique configurations for human evaluation. Complementary developments in experimental design are needed to fully exploit such prototype hardware to estimate useful predictive models of customer preferences. The previous approaches to human appraisals in the design literature have generally assumed the customer preference data is readily available, generally from a marketing source (Kumar et al., 2007), or that a standard experiment design (e.g. full factorial or fractional factorial) is given to each respondent for the purpose of collecting the desired preference data (MacDonald, 2007; Michalek et al., 2005). As will be presented in this work, the large number of factors and the experimental structure of a human appraisal for a complex system, such as an automobile, generally preclude the use of standard designs in such experiments. It will be shown that in such cases, it is more efficient, as well as necessary, to provide each survey respondent with a different set of configurations.

4.1.1 Definition of a Human Appraisal Experiment

A human appraisal is characterized by an interaction between the human respondent and the product design; therefore, the set of factors which influence the response from a given respondent for a given product *configuration* are both *product* attributes, denoted by **A**, and respondent *human* attributes, denoted by **S**, as illustrated in Figure 4.1.

$$\text{Customer Response} = f(\mathbf{A}, \mathbf{S})$$

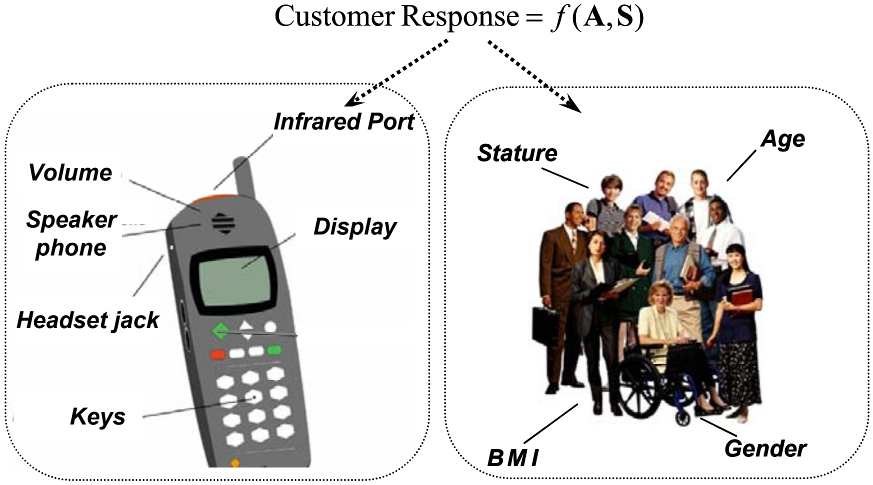


Figure 4.1: Response as a Function of Product and Human Attributes

Human attributes are defined as characteristics, primarily anthropomorphic characteristics such as stature or body mass index (BMI), of a respondent which influence how the respondent experiences the system. In human appraisal experiments, the response for a given experiment could be the identification of a preferred configuration, or *choice*, from the configuration set, a *rank-ordering* of the configurations evaluated, or a *rating* for each configuration (Louviere et al., 2000). In this work, the response considered is in the form of a discrete *rating*, on a scale selected by the survey administrator. The number of rating categories should be limited to between 4 and 11 categories (Cox III, 1980; Green and Rao, 1970) (scales of 0-10, 1-5, and 1-7 are popular in application) to balance the competing desires of maximizing information recovery (i.e. maximize number of categories) versus minimizing scale usage heterogeneity (i.e. minimize number of categories). Rating responses represent an *ordinal* scale, in which higher ratings represent stronger positive preference for a given product configuration. The most popular models for estimating ratings as a function of independent variables are the ordered *probit* (McKelvey and Zavoina, 1975) and *ordered logit* (McCullagh, 1980) models. These models assume a respondent rating is a discrete realization of a continuous underlying opinion, or *utility*, for a given product configuration. In this work, the ordered logit model described in Section 2.3.3 is used; however, the approach presented can easily be adapted to the ordered probit model (or other related models).

4.1.2 Issues in Human Appraisal Experiments

The primary issues with human appraisal experiments are as follows:

- Unique rating style of each respondent.
- Potentially a large number of product and demographic factors to investigate.

- Desire to create a response surface (i.e. quadratic terms) due to non-linearity of the response and the effect of interactions.
- Fatigue of human respondents.
- Desire to specifically include or exclude specific factor combinations.

These issues are described as follows in this subsection.

In human appraisal experiments, a single respondent often evaluates several product configurations in sequence due to time and cost constraints. This implies that human appraisal experiments will naturally have a *random block effect*, as each person's ratings will have some level of correlation depending on the rating style of the respondent. A block is a set of experiments conducted under homogeneous but uncontrolled external conditions. Blocking is necessary since the overall experiment can be quite large, since the number of engineering attributes which potentially characterize a customer-desired attribute, as identified using a process such as the PAFD method of Chapter 3, can be extensive. Also, human appraisals are naturally *split-plot* designs (Box et al., 2005), because it is unrealistic to completely randomize human attributes since a single respondent represents a set of fixed human attributes, and it is the most efficient to have a single respondent evaluate an entire block of experiments, or configurations, at a single time. Split-plot experiments are characterized by one or more factors remaining unchanged for a given set of experiments.

In general, the goal of a human appraisal experiment is to create a response surface model, thus requiring a minimum of three levels of each product attribute (three levels cannot always be achieved for human attributes which are categorical, such as gender). The desire to create a response surface is based upon findings in psychometrics, in which it has been found that the human sensation magnitude to a given stimuli intensity follows a power law relationship

(Stevens, 1986). A three level experiment enables approximation of the power law relationship using linear and quadratic terms in the prediction model (e.g., the ordered logit model).

A key issue to consider in human appraisal experiments is user fatigue (Kuhfeld et al., 1994). Unlike computer or industrial experiments, fatigue will create additional error in the response in a human appraisal experiment. The number of trials or configurations, B, given to each respondent must be managed to ensure the effects of fatigue are limited. Another important issue in human appraisal experiments is the inclusion or exclusion of certain (experimental) design points of interest. The reason for specific inclusion or exclusion of design points is due to the interaction effects of certain factors, which may be theorized to be highly significant and important. If the interaction effect is achievable in the product, it would be of particular interest to study the impact of the interaction, whereas if the interaction is unachievable in the real product, it may be of interest to exclude such a combination. The design of experiments with excluded combinations has been studied previously, e.g. Steckel et al. (1991), but the general case of inclusion or exclusion of certain design points has not been examined.

An example of a human appraisal used throughout this work is the design of an automotive occupant package. A respondent's rating of a particular package configuration is dependent not only upon the product attributes (**A**), such as the amount of headroom, knee room, etc. in the package, but also the human attributes of the respondent (**S**), such as his/her stature, weight, gender, etc. Also, these experiments are characterized by a block effect because, after controlling for the respondents' human attributes, each respondent will retain a certain correlation among their ratings which must be accounted for in the resulting model.

To summarize, the focus of this chapter is the development of a design of experiments methodology for human appraisal experiments, considering the split-plot and block structure of these experiments, and the use of ordered logit (or probit) to estimate the subsequent response model. The developed methodology enables the number of configurations, B, provided to each respondent to be controlled and minimized, and will also allow certain factor combinations to be included or excluded.

4.2 Blocked and Split-Plot Design of Experiments

Blocked and Split-Plot designs have been used extensively in physical experimentation. The difference between block and split-plot designs is illustrated in Figure 4.2. The larger experimental unit (composed of many individual experimental design points or configurations) in a blocked experiment is called a *block*, whereas the larger experimental unit in a split-plot experiment is called a *whole plot*. Each block or whole plot consists of a number of experimental design factors, $\mathbf{x}=(E_1, E_2, \ldots E_j)$, the values of which are determined by a design criterion, such as the D-optimality to be discussed in the next section. The primary difference between a blocked versus split-plot design is that in a split-plot design, *whole-plot factors*, such as a human factor S_1, remain unchanged for a given experimental run. In blocked experiments, there are no corresponding larger experimental-unit, or block-level, factors such as the whole-plot factors. Therefore, the goal of a split-plot design is the selection of the design points *under* each whole plot factor, whereas in a blocked design the goal is the allocation of design points to each block.

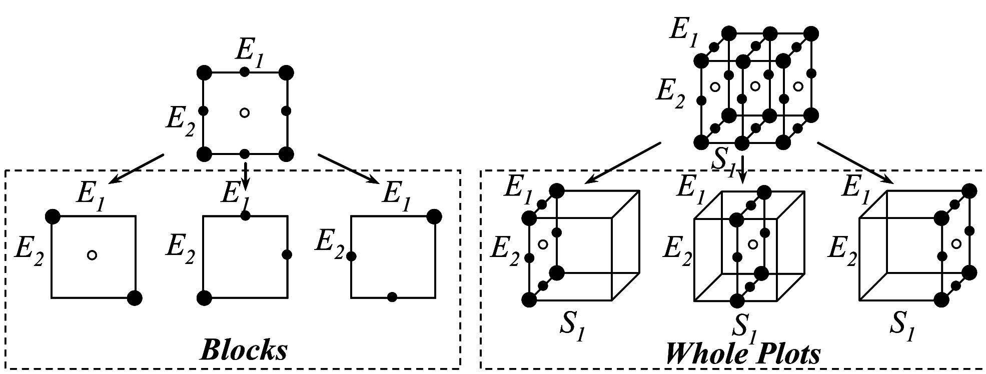


Figure 4.2: The Structure of a Blocked and Split-Plot Experiments (Goos, 2002)

A demonstration experiment with two **E** and one **S**, which is presumably used to estimate a linear regression model, quadratic in **E** and linear in **S**, is used to demonstrate the terminology used in optimal DOE. In the proposed experimental design approach, both **E** and **S** comprise the experimental factor set, **x**:

$$\mathbf{x} = \begin{bmatrix} E_1 & E_2 & S_1 \end{bmatrix}. \tag{4.1}$$

The complete set of terms, **E** and **S**, which appear in the resulting prediction model (e.g., the ordered logit model), such as an intercept and linear, quadratic, and interaction terms, form the extended design point, denoted by **f(x)**:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} 1 & E_1 & E_2 & S_1 & E_1^2 & E_2^2 & E_1E_2 & S_1E_1 & S_1E_2 \end{bmatrix}. \tag{4.2}$$

The matrix of all extended design points in the complete experiment form the extended design matrix, denoted as **F**:

$$\mathbf{F} = \begin{bmatrix} 1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 \\ 1 & +1 & 0 & -1 & +1 & 0 & 0 & -1 & 0 \\ 1 & 0 & +1 & -1 & 0 & +1 & 0 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{4.3}$$

The motivation for split-plot design methodology is the inclusion of "hard-to-change" factors, e.g. a respondent's human attributes, in the experimental design. These hard-to-change factors are the whole-plot factors, which are not completely randomized as with the other design factors, and remain at a fixed level during the completion of a given whole plot experiment. Alternatively, blocked experiments are motivated by the need to minimize the effects of known or theorized uncontrollable factors, such as the rating style of each respondent, not included as a design factor (i.e. **E** or **S**), but believed to have an influence on the experiment response.

Therefore, the goal in blocked experiments is to distribute the experimental design points among homogeneous blocks, or respondents, to minimize the effects of uncontrollable factors.

4.3 Optimal Experimental Design Method for Human Appraisals using Rating Responses

In our proposed experimental design method, the human appraisal experiment is considered as both a split-plot and a blocked experiment. The human attributes **S** form the whole-plot factors because they represent hard-to-change factors. As discussed in the introduction, a single respondent, characterized by a fixed human profile **S**, rates several configurations in succession due to the expense and inconvenience of requiring people to evaluate configurations randomly over time. Also, each whole-plot experiment may be too large for a single respondent to complete due to the fatigue issues discussed in Section 4.1. Each whole-plot may therefore be distributed among multiple survey respondents, each with the same **S**, in the form of blocks. The blocked split-plot design is illustrated in Figure 4.3. In this diagram, the respondent human factors, **S**, are the whole-plot factors, and the product factors, **E**, are the split-plot factors.

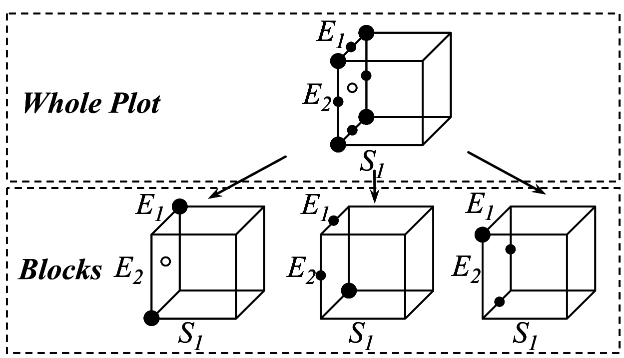


Figure 4.3: The Structure of the Human Appraisal Blocked Split-Plot Experiment (Goos, 2002)

In this work, Optimal Design of Experiments (DOE) methodology will be used to select the preferred human appraisal experiment. In optimal DOE, a *candidate set* of design points *G*, typically the design points of a full factorial experiment in the desired number of factors, is provided to an algorithm which uses a defined *criterion* to select the optimal design points from the set to achieve a design of any arbitrary size, *M*. A key concept in Optimal DOE is that the form of the model to be estimated, i.e. the form of the extended design point **f(x)**, must be specified *a priori* to determine the optimal design which supports the specified model.

4.3.1 Optimal Experimental Design Selection Criterion

Various criteria for selecting the optimal experimental design are available, the most widely used being *D-Optimality*. In general, several criteria exist for selecting a preferred experimental design. Popular criteria in the literature are *D*, *A*, *G*, and *V* (also known as *I*, *Q*, or *IV*) optimality, which are all functions of the Fisher Information matrix, **M,** of the extended design matrix, **F**. The *D* and *A* criteria are related to making precise estimates of the model parameters ($\boldsymbol{\beta}$), whereas the *G* and *V* criteria are concerned with minimizing the overall prediction variance of the resulting model. While any optimality criterion can be used with the approach presented in this work, the approach is presented using the *D*-optimality criterion for several reasons. Firstly, *D*-optimality is widely used as an optimality criterion and is computationally inexpensive for experiment selection compared to some of the other criteria, such as *V*-optimality. Additionally, *D*-optimal designs have been shown to be highly efficient (i.e. provide efficient model building) with respect to the other optimality criteria (i.e. *G*, *A*, and *V*), whereas *G*, *A*, and *V*-optimal designs generally are not efficient with respect to *D*-optimality (Goos, 2002). Also, because the models estimated must be validated in some manner, *D*-optimal designs provide precise estimates of the resulting model parameters ($\boldsymbol{\beta}$) which can be interpreted for expected sign and magnitude as part of the model validation process. *D*-optimality is achieved algorithmically through maximization of the determinant of the Fisher Information matrix, **M,** or the *D*-criterion, of a given experiment design:

$$\max \quad \det(\mathbf{M}). \tag{4.4}$$

The Fisher Information matrix for the Ordinary Least Squares (OLS) fixed-effect model parameters, $\boldsymbol{\beta}$, can be expressed as (Atkinson and Donev, 1992):

$$\mathbf{M} = \sigma_\varepsilon^{-2}\mathbf{F}'\mathbf{F}. \tag{4.5}$$

As seen in Eq.(4.5), **M** for an OLS model is a function of the extended design matrix, **F**, and the random error variance, σ_ε, (which, without loss of generality, can be assumed to be 1 for experiment optimization purposes) both of which are independent of the model parameters $\boldsymbol{\beta}$. In the case of Generalized Least Squares (GLS), Goos (2002) has derived the information matrix for the random block effects model, in which each experiment respondent forms a block. The variance-covariance matrix of the rating observations, **R**, for a single respondent n, $\mathbf{cov}(\mathbf{R}_n)$, is of the form:

$$\mathbf{V}_n = \begin{bmatrix} \left(\sigma_\varepsilon^2 + \sigma_u^2\right) & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \left(\sigma_\varepsilon^2 + \sigma_u^2\right) & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \left(\sigma_\varepsilon^2 + \sigma_u^2\right) \end{bmatrix}. \tag{4.6}$$

where σ_u is the variance at the *respondent* level, and σ_ε is the variance at the *observation* level. The information matrix for all observations can then be written as:

$$\mathbf{M} = \mathbf{F}'\mathbf{V}^{-1}\mathbf{F} = \sigma_\varepsilon^{-2}\left\{\mathbf{F}'\mathbf{F} - \sum_{n=1}^{N}\frac{\rho}{1+\rho\left(B_n - 1\right)}(\mathbf{F}_n'\mathbf{1}_{B_n})(\mathbf{F}_n'\mathbf{1}_{B_n})'\right\}, \text{ where } \rho = \frac{\sigma_u^2}{\left(\sigma_\varepsilon^2 + \sigma_u^2\right)}, \tag{4.7}$$

B_n is the number of configurations in block n (of N blocks), and $\mathbf{1}_{B_n}$ is a square matrix of ones of size B_n. In this case, an estimate of ρ, which is a measure of the ratio of *across*-respondent to *within*-respondent variance, is needed to select the optimal design. For this reason, such experimental designs are referred to as *semi-Bayesian* designs, since they require a prior estimation of ρ. The expression for **M** given in Eq (4.7) is only valid if the model to be estimated is a (least squares) linear regression model. It is therefore not valid for the human appraisal experiments in this work which are to be modeled using ordered logit.

4.3.2 Derivation of Human Appraisal Experiment Selection Criterion

A complementary derivation is proposed in this work to support estimation of the ordered logit model. The ordered logit model can be written as:

$$\Pr\left(R_{ni} = R_{nip}\right) = \pi_{nip}\left(\boldsymbol{\beta}\right) = F\left(k_p - \mathbf{x}'_{ni}\boldsymbol{\beta}\right) - F\left(k_{p-1} - \mathbf{x}'_{ni}\boldsymbol{\beta}\right), \tag{4.8}$$

where R_{ni} is the discrete rating for respondent, or block, n (of N blocks) and configuration i (of B configurations), k is an ordered logit cutpoint, p is a rating category (of P categories, such as 1-10), and F is the Cumulative Distribution Function (CDF) of the logistic distribution (this CDF can be replaced with the standard normal CDF, $\boldsymbol{\Phi}$, if the ordered probit model is to be used).

To enable selection of a D-optimal design to support the ordered logit model, an expression for the information matrix (needed to calculate the D-criterion) that can be estimated without prior knowledge of the resulting model parameters, i.e. $\boldsymbol{\beta}$, is needed. In general, the information matrix for the ordered logit model can be expressed as (Liang and Zeger, 1986):

$$\mathbf{M} = \sum_{n=1}^{N} \mathbf{D}'_n \mathbf{V}_n^{-1} \mathbf{D}_n\,, \tag{4.9}$$

where $\mathbf{V}_n$ is the asymptotic variance-covariance matrix for block n. $\mathbf{D}_n$ is the derivative of $\boldsymbol{\pi}_n$ with respect to $\boldsymbol{\beta}$:

$$\mathbf{D}_n = \mathbf{D}_n(\boldsymbol{\beta}) = d\boldsymbol{\pi}_n(\boldsymbol{\beta})/d\boldsymbol{\beta}, \tag{4.10}$$

where the (P−1) vector of ratings probabilities for a single individual n for configuration i is given as $\boldsymbol{\pi}_{ni} = (\pi_{ni1}, \pi_{ni2}, \ldots \pi_{ni,P-1})$ and $\boldsymbol{\pi}_n = (\boldsymbol{\pi}_{n1}, \boldsymbol{\pi}_{n2}, \ldots \boldsymbol{\pi}_{nB})'$. The asymptotic variance-covariance matrix, $\mathbf{V}_n$, for the ordinal model, such as ordered logit, can be written in block-matrix form as (Williamson et al., 1995):

$$\mathbf{V}_n = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \cdots & \mathbf{V}_{1T} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \cdots & \mathbf{V}_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{V}_{T1} & \mathbf{V}_{T2} & \cdots & \mathbf{V}_{TT} \end{bmatrix}, \tag{4.11}$$

where the on-diagonal matrices are *multinomial* covariance matrices, $\mathbf{V}_{tt} = \text{diag}(\boldsymbol{\pi}_{ni}) - \boldsymbol{\pi}_{ni}\boldsymbol{\pi}'_{ni}$, and the off-diagonal matrices, $\mathbf{V}_{ts}$ ($t \neq s$), are the *within-block* covariance matrices between any two responses in a block. These matrices are generally calculated as part of the model estimation process using collected data; therefore, a method for estimating them for experimental design purposes must be devised.

The on-diagonal multinomial covariance matrices ($\mathbf{V}_{tt}$) can be calculated from knowledge of the ratings probabilities; however, the within-block covariance matrix ($\mathbf{V}_{ts}$) requires additional derivation. In general, the $\mathbf{V}_{ts}$ matrix follows the form (Liang and Zeger, 1986):

$$\mathbf{V}_{ts} = \left(\mathbf{B}^{1/2}\right)' \mathbf{P}_{tn} \left(\mathbf{B}^{1/2}\right), \tag{4.12}$$

where $\mathbf{P}$ is the "working" correlation matrix, and $\mathbf{B}$ is a matrix determined by the correlation structure. The selection of $\mathbf{B}$ and $\mathbf{P}$ depends upon the form of the model to be estimated with the experimental response data (Hines, 1997; Hines, 1998; Zorn, 2001). The proper specification for $\mathbf{P}_{tn}$ for the random-effects ordered logit model has been found to be the "exchangeable" structure. In the "exchangeable" structure, $\mathbf{P}_{tn}$ is a diagonal matrix with all diagonal elements of $\mathbf{P}_{tn} = \alpha$, implying equal correlation among all observations in a given block. In this formulation, α is the pair-wise correlation coefficient between elements in the $\mathbf{V}_{tt}$ matrices, similar to the correlation coefficient ρ applicable for the scalar variance-covariance matrix of Eq (4.6). The recommended specification for $\mathbf{B}$ for the random effects model is $\mathbf{V}_{tt}$ (Hines, 1997; Hines, 1998). Therefore $\mathbf{V}_{ts}$ can be written as:

$$\mathbf{V}_{ts} = \left(\mathbf{V}_{tt}^{1/2}\right)' \mathbf{diag}(\alpha)\left(\mathbf{V}_{tt}^{1/2}\right) \quad t \neq s\,. \tag{4.13}$$

In viewing Eqs.(4.10), (4.11), and (4.13), it can be seen that in order to calculate $\mathbf{M}$, estimates for $\boldsymbol{\pi}_n$ and α are required. The pair-wise correlation coefficient α is not reported in the random-effects ordered logit modeling process, which provides a challenge to determining a reasonable estimate for α from previous experiments or the literature. However, the coefficient ρ is reported in the modeling process, and it has been found that α can be estimated using ρ by the relation $\alpha \approx \rho/P$ to enable calculation of $\mathbf{V}_{ts}$. This estimate is based upon the assumption that ρ should "distributed" over the P ratings categories in the working correlation matrix, such that the influence of α and ρ are equivalent in the respective information matrix calculations of Eqs. (4.7) and (4.9). Because a ratings prediction model is not available before the experiment is conducted, the rating category (e.g., 1-10) probabilities, $\boldsymbol{\pi}_n$, must be estimated directly. They can be estimated from prior knowledge from a previous experiment, or if no prior knowledge is available, an equal probability of each rating category can be assumed. Because estimates of the entire response probability vectors, $\boldsymbol{\pi}_{ni}$, are needed to calculate $\mathbf{V}_n$ and $\mathbf{D}_n$ to compute $\mathbf{M}$, such experimental designs are referred to as *Bayesian* designs (Atkinson and Donev, 1992).

4.3.3 Verification of the Experimental Design Selection Criterion

To verify the formulation of **M** for the ordered logit model and the estimates for $\boldsymbol{\pi}_n$ and α, two test data sets with equal probability of each rating (i.e. ratings 1-10) are created. In one data set, the average correlation ρ of ratings from a single respondent is set to 0, (data set 1) and in the second data set, the ratings were distributed such that the average ratings correlation, ρ, is 0.40 (data set 2). The purpose of this verification is to ensure that the proposed calculation of the information matrix (Eq.(4.9)), in which the ratings probabilities are estimated *a priori* and the correlation of responses is estimated using α, is consistent with the information matrix calculated from actual data. Ordered Logit models are estimated using both data sets in the statistical modeling software Stata™ (Stata Corporation, 1996-2008). The information matrices calculated by Stata (labeled *stat*) are compared to the information matrices calculated using the proposed derivation using estimates for $\boldsymbol{\pi}_n$ and α (labeled *der*). For data set 1, the information matrices calculated by Stata and Eq. (4.9) are identical, and the determinants of **M** identical ($\det_{stata}$ = $\det_{der}$ =1.16×10^{20}). For data set 2, the difference in the determinants is 7.62% ($\det_{stata}$=5.15×10^{17}, $\det_{der}$=5.54×10^{17}), most likely because only the average correlation could be controlled in the created data set and α is approximated as described previously. A study of the sensitivity of the algorithm to misspecification of ρ has been investigated for the GLS algorithm by Goos (2002). He has found that a misspecification of ±50% results in only a 4-8% error in the information matrix. In a further study (Kessels et al., 2008), it was found that the actual experiment design selection was robust to larger misspecifications of ρ (range of 0.1 to 0.9), indicating that an exact estimate of ρ is not needed for design selection purposes.

The challenges of optimal experimental design for a random-effects ordered logit model can be understood through a comparison to the generalized least squares approach presented at the beginning of this section. In the least squares approach, the on-diagonal terms of the $\mathbf{V}_n$ matrix in Eq. (4.6) are scalars of estimated with-in block and across-block variances, whereas in the ordered logit approach, the on-diagonal terms of the $\mathbf{V}_n$ matrix of Eq. (4.11) are matrices which are a function of estimated response probabilities. Comparing the off-diagonal terms of Eq. (4.6) and Eq. (4.11) indicates that the least squares method requires only a scalar estimate of across-block variance whereas the ordered logit requires estimation of a matrix (i.e. $\mathbf{V}_{ts}$). This comparison indicates the difficulties in design optimization for ordinal data models in that the computation is more expensive due to the replacement of scalar quantities with matrices, and that estimates of both $\boldsymbol{\pi}_n$ and α are required.

4.4 Optimal Human Appraisal Algorithmic Implementation

The algorithmic implementation for selecting the optimal blocked split-plot design follows the approach provided in (Goos and Vandebroek, 2004), with the least squares information matrix of Eq. (4.7) used in their approach replaced with that of Eq. (4.9) for the new approach. In general, the experimental design is built sequentially, with points from the candidate set (G) having the highest prediction variance $\mathrm{var}\{\hat{R}(\mathbf{x})\}$ added to the experiment to maximize the D-criterion. The prediction variance for any value of **f(x)** must be calculated to determine the point from G to add to the experiment. For the GLS model, **f(x)** is a vector and the prediction variance can be calculated as $\mathrm{var}\{\hat{R}(\mathbf{x})\} = \sigma_\varepsilon^2 \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}\mathbf{f}(\mathbf{x})$; however in the case of the ordinal model, each **f(x)** results in a matrix of terms for each of the (P−1) rating categories. The prediction variance for any point to be added to the design can be estimated using the delta method for asymptotic variance (Tamhane and Dunlop, 2000):

$$\mathrm{var}\left\{\hat{R}(\mathbf{x})\right\}_p = \mathrm{var}\left\{\pi_{ni}(\boldsymbol{\beta})\right\}_p \approx \left(\frac{d\pi_p(\boldsymbol{\beta})}{d\boldsymbol{\beta}}\right)' \mathbf{M}^{-1}\left(\frac{d\pi_p(\boldsymbol{\beta})}{d\boldsymbol{\beta}}\right) \tag{4.14}$$

As seen in Eq. (4.14), the prediction variance is calculated for each of P ratings categories, leading to a vector of prediction variances for each design point **f(x)**. Therefore, the design point with the highest summed *total* prediction variance is added to the experiment:

$$\mathrm{var}\left\{\hat{R}(\mathbf{x})\right\} = \sum_{p=1}^{P} \mathrm{var}\left\{\pi_{ni}(\boldsymbol{\beta})\right\}_p \tag{4.15}$$

To implement the algorithm, a simplified method of expressing **M** a given in Appendix B. An overview of the algorithm is shown in Figure 4.4 and described as follows:

1. **Generate a set of Candidate points**, G, for the product attributes, **A**, from which to select the optimal set. G is typically the points of a full factorial experiment in the number of factors desired. Specific factor combinations to be specifically excluded from the candidate set, or specifically included in the final experiment design are also specified.
2. **Create an experimental design for the desired human whole-plot factors**, **S**. This design can be a full or fractional factorial in human attributes, depending upon the size of **S** and the number of respondents. Randomly assign the whole plot factors to each block, n.
3. **Create a starting design**. To begin building the experimental design, a starting design is composed of a randomly selected small number of points from the candidate set and randomly assigned to the blocks. Compute the initial information matrix, **M**, and the determinant, det(**M**).

4. **Determine the point in the candidate set *G* with the largest prediction variance**, $\mathrm{var}\left\{\hat{R}(\mathbf{x})\right\}$. Randomly assign this point to a block, and update **M** and det(**M**). Repeat this process until each block receives *B* configurations, forming an experiment design of size *M*.
5. **Evaluate exchanges**. Since the design was started with a random selection of points, there may be points in the candidate set *G* which will increase the *D*-criterion. Each point in the current design is evaluated to determine if its replacement by a point in the candidate set will increase the *D*-criterion. This is continued until no further increases can be established.
6. **Record the *D*-Criterion and repeat steps 3-5**. Steps 3-5 constitute a single *try*, each with a local maximum for *D*-Optimality based on the starting design of Step 3. *T* tries (e.g. 100 tries) can be conducted to search for the global maximum.

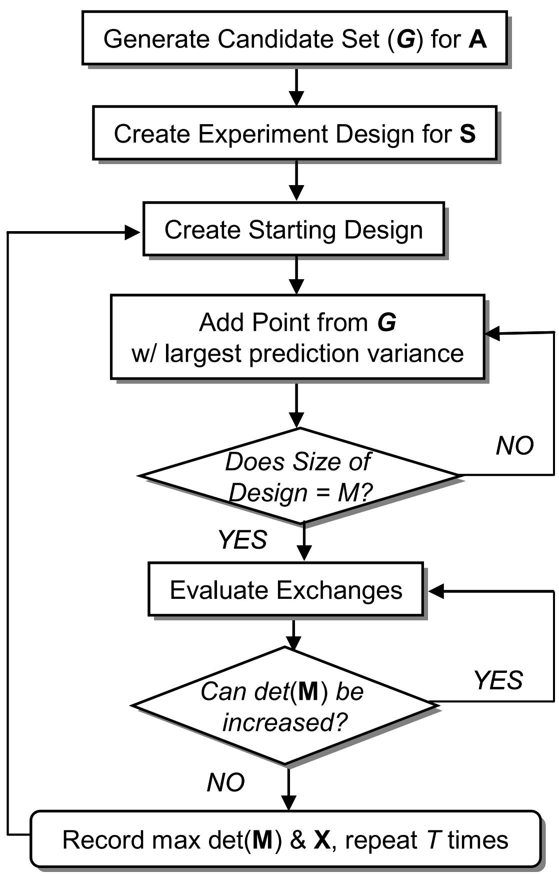


Figure 4.4: Algorithmic Implementation of the Optimal Experimental Design Method

4.5 AUTOMOTIVE OCCUPANT PACKAGING CASE STUDY

A case study using an automotive occupant packaging human appraisal is used to demonstrate the methodology, as well as the advantages of using the blocked split-plot experimental design methodology for human appraisal. The occupant packaging appraisal is performed on a Programmable Vehicle Model (PVM) as shown in Figure 4.5, which is capable of creating a wide range of parametric representations of an occupant package through a computer controlled interface.

Figure 4.5: Programmable Vehicle Model (PVM) (Wang et al., 2006)

4.5.1 Design of PVM Experiments

A human appraisal experiment has been previously conducted by Ford Motor Co. using the PVM to evaluate occupant package design specifically for headroom. In the experiment conducted, headroom design is characterized by three dimensions as defined by the Society of Automotive Engineers (SAE) J1100 (Society of Automotive Engineers, 2002): **L38** (frontal), **W35** (lateral), and **H61** (vertical). These three product factors ($\mathbf{x}^*=[E_1, E_2, E_3]$) were used to create a full $\mathbf{3^2 4^1}$ factorial experiment (i.e. 36 trials) which was given to each of 100 human appraisal respondents, for a total of 3600 ratings responses. The responses were given on a (discrete) scale of 2-10, with 10 representing highest satisfaction with the headroom, and 2 representing the least satisfaction, leading to P=9. Human profile (**S**) factors were not used in the design of the experiment; however, the **S** were treated as *covariates* in that the human profile of each person was recorded, but no attempts were made to control the profiles of the respondents in the experimental design process. The data set with ratings responses were used to create a full quadratic response surface model, used to predict a customer headroom rating for a given occupant package design and a given target market human. This data set is referred to as data set ***Full*** in the case study. Conducting an experiment of this size was very time consuming and costly for Ford, and methods to conduct more efficient experiments are needed. Using this example in which data has already been collected, we will demonstrate that the experimental design methodology presented in this chapter allows selection of an experimental design which can be used to estimate a comparable model with significantly fewer experimental design points than used in the ***Full*** data set. In the new methodology, the $3^2 4^1$ factorial experiment forms the candidate set for the optimization algorithm. Additionally, a set of potentially significant human attributes, **S**, is included in the design of the experiment as whole-plot factors. The human profile attributes included are respondent **gender** (Gen) and **stature** (Stat). An issue to address in the experimental design of **S** is that exact levels cannot be practically achieved for all **S** (e.g. stature) in a real human appraisal design. In this case, human attribute *ranges* are assigned to a level in the design of an experiment, for example statures between 54”-57” are coded as the -1 level and those between 73”-76” are coded as the +1 level. These human attribute “bins” are needed to ensure that the proper respondents are selected for the experiment; however, the actual human measurements (e.g., stature, weight, age) are used in the model estimation process. A criterion for selecting the bins is to ensure that 5 and 95 percentile human-measurement respondents of the target population are included in the bins. If more levels (i.e. bins) can be afforded, respondents closer to the human mean (e.g. 50 percentile) should be included; however, it is most important from a *D*-optimality perspective to include 5 and 95 percentile respondents. At the time of the experiment, additional human and socio-economic attributes of a respondent can be recorded and treated as covariates in the modeling process.

To demonstrate the ability of the new method to manage the size of an experiment, the number of configurations given to each respondent is reduced from 36 to a block size of **18**. The whole-plot experiment design is composed of two levels of gender (i.e. male, female) and four levels of stature (using stature ranges), leading to a $2^1 4^1$ whole plot experiment design. Two respondents (i.e. blocks) will be assigned to each whole plot for a total of 16 respondents (or blocks, n), leading to a total of M=288 total trials, vs. 3600 in the ***Full*** experiment described above. A summary of the experimental design is shown in Figure 4.6.

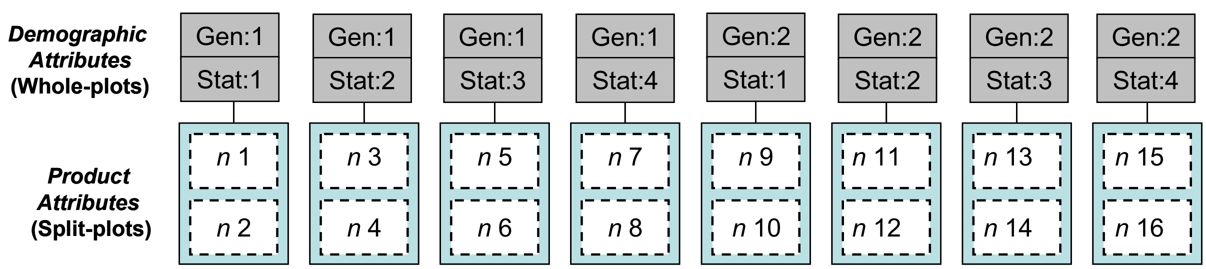


Figure 4.6: Occupant Package Blocked Split-Plot Human Appraisal Experiment

The exact form of the model to be estimated is known for this case study from previous work, enabling specification of model form **f(x)** as defined in Eq. (2). The model form contains full quadratic terms for **E** (linear, squared, interaction) and linear terms for **S** (no **S**·**E** interactions). With **f**(**x**) specified, the algorithm can be used to select the best 18 configurations to give to each of the 16 respondents. As discussed in Section 4.3 a *prior* ratings probability estimate is needed to calculate **M**. For this study, it is assumed that the probability, π_{nip}, of each rating R_p for each respondent n and each configuration i is equally probable, i.e. $\pi_{nip} = 1/9 = 0.11$. Also, it is known from a previous experiment that the correlation among ratings of a single respondent is ρ=0.3. The use of equal ratings probabilities assumes there is *no* prior information about the ratings responses. If prior information is available, (e.g., middle ratings are more likely than extreme ratings) such information can be incorporated to improve the experiment design. In this experiment, the best experiment as selected by the algorithm presents each respondent with a *different* set of configurations, demonstrating that the use of the same 18 point fractional factorial experiment (of the original $3^2 4^1$=36 experiment) for each respondent is not optimal for a human appraisal experiment. The data set with observations based upon this design is labeled ***D-Opt***, with an example of the configurations assigned to the first three respondents shown in Appendix C. For comparison, an additional set of experimental designs is created. In these designs, 16 respondents are randomly selected from the original 100 respondents and 18 observations are randomly selected from the 36 total observations for each respondent. A total of 100 such random experiments are created, such that experimental design comparisons are made to the *mean* random experimental design, to ensure that any comparisons are made based upon a typical random experiment and not an outlying design. This set of experimental designs is labeled ***Rand***.

4.5.2 Results of Random-Effects Ordered Logit Model Estimation

With the three experimental designs established, a random-effects ordered logit model is estimated using each of the three data sets. A summary of the experimental efficiency as measured by *D*-efficiency, model fit as measured by ρ_0^2 (Train, 2003), and average rating prediction error (Johnson and Albert, 1999) are shown in Table 4.1. In the case of the ***Rand*** experiment designs, the model fit is evaluated using experimental data with a mean *D*-efficiency.

Table 4.1: Summary of Experiment and Model Statistics

	Number Experiments	*D*-Efficiency	Model Fit ρ_0^2	Prediction Error
Full	3600	—	0.373	2.80%
D-Opt	288	79.7%	0.485	6.90%
Rand	288	35.8% ± 1.6%*	0.375	14.60%

* The mean and ± 1 standard deviation are shown

D-efficiency is a measure of the relative efficiency of an experiment versus a base experiment, for example the ***Full*** experiment in this work. As seen in the table, the *D*-efficiency of the ***D-Opt*** experiment is high, ensuring low variance estimates of the model parameters, whereas the mean *D*-efficiency of the ***Rand*** experiment is quite low and will result in poor model parameter estimates. The ρ_0^2 statistic varies between 0 and 1 and is a function of the log-likelihood of the estimated model, with higher ρ_0^2 indicating a better "model fit". The ρ_0^2 for the ***D-Opt*** model is significantly higher than that of the ***Full*** model. The explanation for this can be provided by reviewing the assumptions of ordered logit modeling and the nature of ratings. Ratings tend to have higher variance in the middle ratings versus those at the extremes (McKelvey and Zavoina, 1975). *D*-optimality tends to bias towards including those configurations with the most extreme settings. Thus by selecting the *D*-Optimal configurations from the full PVM data set, a more efficient estimation of the model β parameters, and hence utility, is accomplished for the assumed model. The fit of this mean ***Rand*** model is similar to the ***Full*** model, which is consistent with the fact the points were randomly selected, so similar model fits are expected. The prediction error is the ratings misclassification error when using the three models to estimate ratings in the full 3600 observation data set. The effects of the prediction error on the resulting ratings predictions can be seen graphically in Figure 4.7. As shown, the prediction error of the mean ***Rand*** model is significantly higher than the other two models.

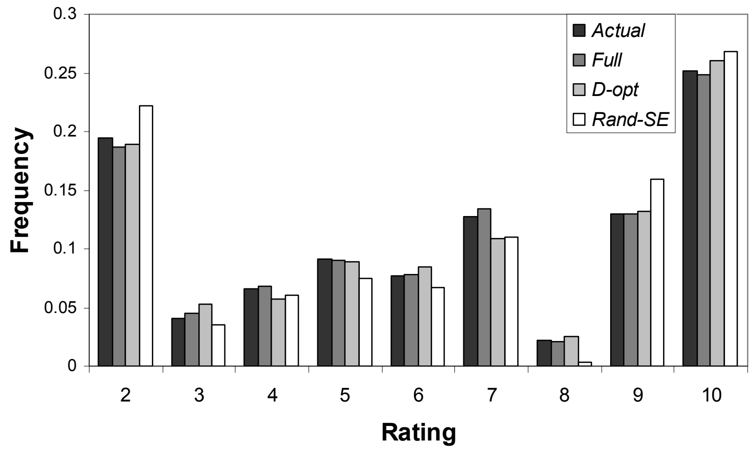


Figure 4.7: Comparison of Ratings Predictions to Actual Ratings

The estimated model parameters, **β**, for the utility function are shown in Table 4.2, along with the standard errors of the parameters (the cut points, **k**, are not shown since these estimates are similar for all three models). We can compare model attributes, such as the relative magnitudes and signs of parameters and the general interpretation of the models, in addition to the model statistics. Considering the model estimated on the ***Full*** data set to be the baseline, it is seen that the model estimated using the ***D-Opt*** data set is close in interpretation. The signs of the parameters agree (except for the insignificant L38*W35 interaction). The ranking of parameter importance as measured by the parameter magnitudes is the same in both models. Vertical headroom clearance (H61) is found to be the most important dimension influencing a respondent's perception of headroom. The next most important dimension is frontal headroom clearance (L38), followed by lateral headroom clearance (W35). The human attributes indicate that taller respondents and female respondents (gender is a dummy variable: 0=male, 1=female) systematically respond with lower headroom ratings (on average) than shorter and male respondents, respectively. The ratio of parameters (e.g., W35/H61) is similar in both models, with the exception of L38 which is more important in the ***D-Opt*** model. The reason for this could be explained by the improved model fit statistic, ρ_0^2, of the ***D-Opt*** model as described previously.

Table 4.2: Summary of Headroom Rating Model Parameters

	Full Model coef.	*std. err.*	***D-Opt Model*** coef.	*std. err.*		***Rand Model*** coef.	*std. err.*	
L38	**2.61**	*0.368*	**4.50**	*0.731*	*	**2.49**	*1.449*	
W35	**2.03**	*0.359*	**2.11**	*0.970*		**2.83**	*1.376*	*
H61	**12.09**	*0.491*	**13.01**	*2.165*	**	**10.61**	*1.838*	*
$L38^2$	**-0.74**	*0.292*	**-0.76**	*0.852*		**-0.23**	*1.111*	**
$W35^2$	**-1.23**	*0.291*	**-1.08**	*1.562*		**-2.14**	*1.104*	*
$H61^2$	**-2.55**	*0.354*	**-2.40**	*1.693*		**-0.89**	*1.325*	*
L38*W35	**0.19**	*0.211*	**-0.16**	*0.820*		**0.13**	*0.826*	
L38*H61	**-0.32**	*0.270*	**-0.16**	*0.949*		**-1.15**	*1.093*	*
W35*H61	**0.49**	*0.261*	**0.20**	*0.857*		**0.85**	*1.010*	
gender	**-0.78**	*0.494*	**-0.56**	*0.726*		**0.14**	*1.115*	**
stature	**-2.24**	*1.008*	**-1.81**	*1.425*		**-0.94**	*2.763*	
resp. σ^2_u	**2.95**	*0.452*	**1.73**	*0.780*		**2.57**	*1.071*	

The model parameters in the ***D-Opt*** and ***Rand*** models are compared to those in the ***Full*** model using a *t*-test, in which the null hypothesis is that the model parameters are not different. The model parameters in which the null hypothesis can be rejected with 95% confidence are marked with *, whereas those rejected with 90% confidence are marked with ** in Table 4.2. As seen in the table, the ***Rand*** model contains significantly more parameters which differ from the ***Full*** model than the ***D-Opt*** model. Such results are expected due to the lower *D*-efficiency of the ***Rand*** experiment, which results in less precise estimates of the model parameters than the higher efficiency ***D-Opt*** model.

While the *D*-optimization algorithm has been shown to be effective for this example, its true utility is in experiments with large numbers of product attribute factors (e.g., 6-9) and several human attributes. In such a case, the candidate set will be several hundred to several thousands of potential points, and the task of choosing the appropriate set of points for each respondent is not as straightforward as in the previous example. To demonstrate, an experiment designed for the PVM to elicit preferences for the roominess and ingress/egress of the vehicle occupant package is used. In this simplified experiment, eight product factors are examined by eight respondents, and it is desired to estimate all linear, quadratic, and all 2-factor **E·E** and **E·S** interactions. Respondents are selected based upon three human factors at two levels (a 2^3 full factorial human experiment). The experiment design for the product attributes is conducted by selecting 18 points from a 3^8 full factorial (i.e. C_{6561}^{18}) for each respondent. In this example, the *D*-optimal experimental design is found with the algorithm, and 100 randomly selected experimental designs are also generated for comparison as in the previous example. In this comparison, the *D*-optimal experiment is the baseline for the efficiency comparison, since comparison to an experiment in which each respondent receives the 3^8 full factorial, product factor experiment (i.e. 6561 configurations) is not a realistic baseline. In Table 4.3, the mean *D*-efficiency of the 8 factor random experiments in this example is compared to the mean *D*-efficiency of the random 3 factor experiments of the previous example. As shown, the efficiency of the random 3 factor experiment has a mean *D*-efficiency of 45.0%, whereas the random 8 factor experiment has a mean *D*-efficiency of 29.4%. The variance of the random 8 factor experiment is higher than the 3 factor experiment as would be expected in selecting 18 points from 6561 (C_{6561}^{18}) versus 36 (C_{36}^{18}) points for each respondent. As shown previously in Table 4.2, reduced *D*-efficiency results in reduced precision in estimating model parameters.

Table 4.3: Comparison of Three Factor to Eight Factor Human Appraisal Experiment

Product Factors	**3** Factors, **3** Levels		**8** Factors, **3** Levels	
Human Factors	$2^1 4^1$		2^3	
	mean	*st.dev.*	*mean*	*st.dev.*
D-Optimal Exp.	2.24E+59		8.10E+140	
Random Exp.	6.17E+52	5.3E+52	9.77E+105	1.1E+107
D-efficiency Random	**45.0%**	**2.1%**	**29.4%**	**4.2%**

4.6 DISCUSSION AND SUMMARY

An algorithmic approach for selection of the optimal design of experiments for human appraisal experiments has been developed, demonstrated, and validated in this chapter. An algorithmic approach is necessary for human appraisals due to the large number of potential design and human attributes, coupled with issues of respondent fatigue in such experiments. Human appraisal experiments have been shown to be unique in that the experiment response is a function of both the product attributes and the human attributes of the respondent. They are characterized as split-plot designs, in which the respondent human attributes form the hard to change whole-plot factors while the product attributes form the split-plot factors. The experiments are also characterized by random block effects, in which the configurations evaluated by a single respondent form a block. The experimental design algorithm presented seeks to identify the experimental design which maximizes the determinant of the Fisher Information Matrix, or *D*-criterion, of a given design, assuming that the model to be estimated is an ordered logit model.

The case study and subsequent discussion demonstrate many of the key features of the optimization algorithm. Most importantly, it was shown that the algorithm allows efficient model estimation with a minimal number of experiment points. For the vehicle headroom appraisal, previous methods had used 3600 experiment points, while a comparable model was estimated using 288 experiment points selected using the proposed algorithm. Also, it was shown that randomly selecting 288 points from the full 3600 point experiment produces an inferior model, and the utility of the algorithm increases as the number of experiment factors increases. The optimization algorithm distributes a different set of experiment points to each respondent, demonstrating that using a standard fractional factorial to reduce the number of trials per person is not the best alterative for human appraisals. This methodology is used to design a human appraisal experiment to understand preferences for automobile occupant package design in Chapter 5. The data collected from this human appraisal is used to build random-effects ordered logit models which are utilized in the Bayesian Hierarchical Choice Model of Chapter 6.

Chapter 5

MULTIVARIATE STATISTICAL ANALYSIS METHODS FOR HUMAN APPRAISALS

In Chapter 4, an algorithmic method to design human appraisal experiments was presented. In this chapter, a human appraisal experiment is designed using the algorithmic method, and subsequently conducted and analyzed to understand preferences for automobile occupant package design. The experiment is conducted on the Ford Programmable Vehicle Model (PVM) to understand preferences for occupant package roominess, ingress and egress. The experiment is conducted to build predictive parametric models of consumer preferences. While the experiment is designed specifically to support random-effects ordered logit modeling using the developed algorithm, several issues must be addressed to obtain useful predictive models. An issue with this class of experiment is that the heterogeneity of the experimental respondents contributes to the response, and this heterogeneity must be understood to separate the influence of design factors from that of human factors. Latent class analysis is used to combine multiple responses of the human appraisal respondents to an appropriate set of measures. Cluster analysis and smoothing spline regression are used to gain an understanding of respondent rating styles and preference heterogeneity. These analyses allow estimation of ordered logit models for prediction of consumer occupant package preferences. Methods from machine learning are also investigated as an alternative to parametric modeling. The methods presented in this chapter are designed to understand consumer heterogeneity and address issues unique to human appraisal experiments.

The chapter is organized as follows: Section 5.1 introduces the problems in analyzing human appraisal experiments; Section 5.2 describes the PVM experiments conducted; Sections 5.3 - 5.5 describe the latent class, clustering, and smoothing spline analyses; Section 5.6 presents the random-effects ordered logit models; and Section 5.7 presents methods from machine learning.

5.1 INTRODUCTION

Chapter 4 outlined the process for designing human appraisal experiments, and used a three factor headroom experiment conducted using the Ford Programmable Vehicle Model (PVM) to demonstrate the methodology. In this chapter, a comprehensive experiment was conducted using the Ford Programmable Vehicle Model (Wang et al., 2006) to determine preferences for automobile occupant package design, specifically regarding the *roominess*, *ingress* and *egress* quality of the package. In the experiment, each respondent is presented with several package configurations, for which they evaluate and express their opinion in the form of a rating (e.g., 1-5, 0-10), a standard method for quantifying preferences for subjective attributes (Keeney and Raiffa, 1993). The intent is to use the data collected in the experiments to build ordered logit models to predict consumer preferences (i.e. ratings) for a given set of consumers and for a given occupant package design.

Analyzing and creating models from data collected from a human appraisal experiment presents unique issues not encountered with data collected from the typical industrial and scientific experiments usually considered in design of experiments methodology (Box et al., 2005; Montgomery, 2005). The key issues in human appraisals are that the responses are more difficult to elicit, respondents may utilize different rating styles, the shape of the response-factor curve may not be approximately linear, and interactions may be highly significant. To address these issues, several analysis and modeling methodologies are employed in this work to combine multiple consumer responses into a set of combined measures, to understand the influence of respondent heterogeneity on rating responses, and to gain further insight into the experiment using alternate data analysis methods. In the human appraisals, multiple responses are often collected from the respondent for a single sub-system design. The reason multiple responses are collected for certain sub-systems is because it can be challenging to devise a single survey question to capture the respondents' true opinion of the subsystem design as a whole, and multiple questions are used to assess opinion for different aspects of the design. To determine a measure to use in the modeling process, *Latent Class Analysis* (LCA) (McCutcheon, 1987) is used to create a combined subsystem measure for each respondent to fully describe his/her overall opinion of the subsystem design. Heterogeneity of the survey respondents has much influence on the rating responses given. The effect of *systematic* heterogeneity, which is heterogeneity that can be captured with a human variable in the model, is investigated using *Smoothing Spline Regression* (SSR) (Wood, 2004); *random* heterogeneity, which is heterogeneity not directly observed but rather captured in a distribution of respondent-specific intercepts, is investigated using *Cluster Analysis* (CA) (Johnson and Wichern, 2002). The previous analyses allow estimation of parametric *Random-Effects Ordered Logit* (RE-OL) models for the prediction of ratings for a given population and given package design, to be used in the Bayesian Hierarchical Choice Modeling approach of Chapter 6. In addition to utilizing the results of the previous analyses, interaction effects are also investigated in the RE-OL modeling process. In addition to the parametric ordered logit models, methods from machine learning are explored. *Decision trees* and *Bayesian networks* (Witten and Frank, 2005) are used to gain insights into the data not easily seen in the previous analyses or parametric modeling methods. The methods developed in this work for the analysis of data collected from human appraisal experiments complement our work in Chapter 4 on human appraisal experimental design. The methods presented in this chapter provide a clear understanding of the heterogeneous preferences within a consumer population, applicable for understanding preferences for system, subsystem, or component design.

5.2 PVM ROOMINESS/INGRESS/EGRESS EXPERIMENTS

While a human appraisal experiment was presented in Chapter 4 for vehicle occupant package headroom, it is desired to complete a comprehensive set of human appraisal experiments to develop models for the overall roominess, ingress, and egress preferences for the vehicle occupant package. The full design of the Programmable Vehicle Model (PVM) roominess/ingress/egress experiment is created using the optimal design of experiments (DOE) methodology of Chapter 4. The combined experiment consists of eight product factors, determined from a mapping of customer-desired attributes (**A**) to engineering attributes (**E**) , to influence roominess, ingress, and egress. The eight factors used in the human appraisal experiment correspond to dimensions defined for control of the Ford Motor Co. PVM (note: SgRP is the Seating Reference Point, which is a fixed point in space within the vehicle interior to serve as a measuring reference point, as define by SAE J1100 specification (Society of Automotive Engineers, 2002)):

1. E_1: SgRP to Hinge ($\mathbf{HNG_X}$)
2. E_2: SgRP to Rocker Y ($\mathbf{ROK_Y}$)
3. E_3: SgRP to Heel Z ($\mathbf{HEL_Z}$)
4. E_4: SgRP to Ground Z ($\mathbf{GRD_Z}$)
5. E_5: Sill to Heel (**StoH**)
6. E_6: SgRP to Roof Z ($\mathbf{HR_Z}$)

7. E_7: SgRP to Front Header X (**HR**$_X$)
8. E_8: SgRP to Side Rail Y (**HR**$_Y$)

The relationship among the product attributes and roominess and ingress/egress is illustrated in Figure 5.1.

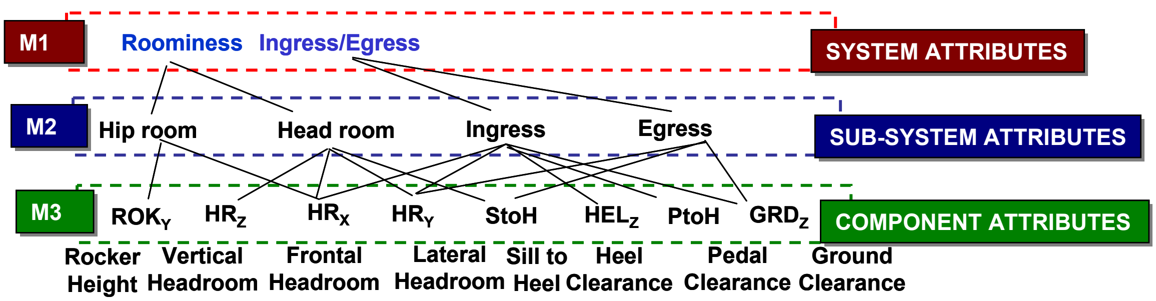


Figure 5.1: Relationship Among Product Attributes and Roominess/Ingress/Egress

All product factors, E_1-E_8, assume three levels to create a response surface ordered logit model, in accordance with the power law response assumption discussed in the previous chapter. The three levels assumed by each product factor (**E**) are shown in Table 5.1.

Table 5.1: Levels of Product Factors (E) used in PVM Experiment (mm)

Level	HNG$_X$	ROK$_Y$	HEL$_Z$	GRD$_Z$	StoH	HR$_Z$	HR$_X$	HR$_Y$
-1	800	380	175	450	0	777	241	122
0	725	450	288	625	70	877	366	197
+1	700	520	400	800	140	977	491	272

Three human attributes have been hypothesized to influence roominess/ingress/egress opinions:

1. S_1: Gender (**Gend**)
2. S_2: Body Mass Index (**BMI**)
3. S_3: Stature (**Stat**)

In this experiment, gender assumes two levels (i.e. male, female), BMI three levels (i.e. low, medium, high), and stature (or height) four levels (i.e. small, medium-small, medium-large, large), and human attribute bins are used for both BMI and stature as described in the previous chapter. The levels used for the human attributes are shown in Table 5.2.

Table 5.2: Levels of Human Factors (S) used in PVM Experiment

	Gend		BMI		Stat: M (in)	Stat: F (in)
-1	Male	**-1**	<24	**-1**	65-68	58-62
+1	Female	**0**	24-30	**-0.33**	68-71	62-65
		+1	>30	**+0.33**	71-73	65-68
				+1	73-78	68-74

It is desired to estimate the following terms in the resulting roominess/ingress/egress models:

- Linear terms for all design factors (**E**) and all demographic factors (**S**) (11 terms)
- Quadratic terms for all **E** (8 terms)
- All 2-factor **E**·**E** and **E**·**S** interactions (28 **E**·**E** and 24 **E**·**S** terms)
- The ordered logit cut points (all cut points count as 1 term in the **X** matrix)

The total number of terms, i.e. the size of the **f(x)** vector, is **72**, which is the number of unique configurations required in the experiment. Based upon previous studies conducted by Ford Motor Co., it has been found that a respondent can evaluate **18** configurations before fatiguing, leading to a block size, *B*, of 18 for each respondent. Based on the number of terms to be estimated and the block size, the minimum number of unique blocks needed in the experiment is **4**. Based on the size of the full factorial demographic design of 24 ($2^1 3^1 4^1=24$) and the desire to have two respondents per demographic class, the experiment requires a total number of respondents (or blocks), *N*, of **48**, each evaluating 18 configurations for a total experiment size, *M*, of **3456**. In this experiment, there are also several factor combinations to specifically exclude:

- no pairing of GRD_Z=450 and HEL_Z=400
- no pairing of GRD_Z =450 and StoH=0
- no pairing of GRD_Z =800 and StoH=140
- no pairing of GRD_Z =800, HEL_Z=175, and ROK_Y>380

With the parameters necessary to run the algorithm of Chapter 4 determined, the experiment is designed. The complete design for all four unique blocks is shown in Appendix D, Table D.1.

An issue encountered with the full experiment is that the resources required for the full experiment could not be secured, and therefore only one half of the experiment is conducted. The *D*-optimal algorithm is used to identify a two-part experiment in which the first two blocks (i.e. blocks 1 and 2) of 18 enable estimation of 36 selected model terms, while the second two blocks (i.e. blocks 3 and 4) are *augmented* to the original two blocks to allow estimation of the remaining 36 terms. The 36 model terms selected which can be estimated with completion of blocks 1 and 2 are as follows:

- Linear terms for design factors (**E**) and all demographic factors (**S**) (8 product terms)
- Quadratic terms for all **E** (8 terms)
- 2-factor **E**·**E** interactions from $E_1 \cdot E_2$ through $E_4 \cdot E_5$ (19 terms)
- All the ordered logit cut points (1 term)

Conducting the experiment in 2 parts reduces the *D*-efficiency to 84.6% of the original experiment. The experimental design for unique blocks 1 and 2 used in this study is shown in Table 5.3. Blocks 3 and 4 are documented in Appendix D, Table D.2; human appraisals for these two blocks were not completed in this work due to PVM resource constraints. Because only blocks 1 and 2 will be completed in this work, a fractional factorial experiment for the human attribute whole-plots (**S**) is needed. The *D*-optimal algorithm is used to identify the most efficient fractional factorial design for the demographic attributes shown in Table 5.4. A total of 30 terms can be estimated from the whole-plot demographic experiment:

- Individual block-effects (24 terms)

- Linear terms for demographic factors (**S**) (3 terms)
- 2-factor **S**·**S** interactions (3 terms)

Table 5.3: Block 1 and 2 Experimental Design for Product Factors (E)

E_1 SgRP to Hinge (HNG_X)	E_2 SgRP to Rocker Y (ROK_Y)	E_3 SgRP to Heel Z (HEL_Z)	E_4 SgRP to Ground Z (GRD_Z)	E_5 Sill to Heel (StoH)	E_6 SgRP to Roof Z (HR_Z)	E_7 SgRP to Frt Hdr X (HR_X)	E_8 SgRP to Side Rail Y (HR_Y)
800	380	175	625	0	807	241	122
700	380	400	800	70	977	241	122
700	380	175	450	140	977	241	122
800	520	400	625	140	977	366	122
800	380	400	625	0	807	491	122
700	520	400	625	0	807	491	122
700	520	175	625	140	877	491	122
800	380	400	625	140	877	491	122
725	450	288	625	140	977	491	122
800	520	400	625	0	877	241	197
700	380	175	800	0	807	241	272
800	380	400	800	70	807	241	272
800	520	288	800	0	977	241	272
800	520	400	625	140	807	366	272
700	450	175	450	140	807	491	272
800	380	175	450	70	977	491	272
700	520	288	450	140	977	491	272
700	380	400	625	140	977	491	272
700	520	400	625	140	807	241	122
800	380	288	450	70	977	241	122
700	380	400	625	140	807	366	122
800	450	288	450	70	877	366	122
800	380	400	800	0	977	366	122
800	520	175	625	0	807	491	122
700	380	175	800	0	807	491	122
800	380	175	800	70	977	491	122
700	520	175	625	0	807	241	197
700	520	400	800	70	807	491	197
700	520	400	800	0	807	241	272
800	380	175	450	140	807	241	272
700	380	175	625	140	807	241	272
725	520	175	625	70	877	241	272
800	520	175	450	140	977	241	272
700	380	288	625	0	877	491	272
700	520	175	625	0	977	491	272
800	450	400	625	0	977	491	272

Table 5.4: Experimental Design for Demographic Attributes

S_1	S_2	S_3
Gender (Gend)	**BMI** (BMI)	**Stature** (Stat)
M	<24	65-68
M	>30	65-68
M	>30	68-71
M	<24	71-73
M	24-30	73-78
M	>30	73-78
F	<24	58-62
F	>30	58-62
F	<24	62-65
F	24-30	65-68
F	<24	68-74
F	>30	68-74

With a complete experimental design for **E** (i.e. split-plot factors) and **S** (i.e. whole-plot factors) complete, the logistics of conducting the experiment are addressed. For each of the 18 configurations presented, the respondent is asked to evaluate the following subsystem designs and to provide ratings as follows (rating scale to be used shown in parentheses):

1. *Ingress*: Acceptability (1-4), Effort (1-5), and Space (1-5)
2. *Interior*: Headroom (1-5), Leftroom (1-5), Kneeroom (1-5), and Roominess (1-5)
3. *Egress*: Acceptability (1-5), Effort (1-5), and Space (1-5)

An example of the questions give to each respondent is given in Figure 5.2. A complete list of the questions asked of each respondent, a definition of the rating scale, and the experimental protocol can be found in Appendix E.

1. How acceptable is this vehicle configuration for ingress? This is rated on a 1 to 4 scale with the following definition for each rating as you can see posted in front of the vehicle: 1 is "very unacceptable", 2 is "somewhat unacceptable", 3 is "somewhat acceptable" and 4 is "very acceptable".

Very unacceptable	**Somewhat unacceptable**	**Somewhat acceptable**	**Very acceptable**
1	**2**	**3**	**4**

2. What is the overall ease of ingress, for the vehicle? This includes evaluation of stepping up and passing through the door opening. This question is rated on a 1 to 5 scale, again as you can see posted in front of the vehicle, with the following definition for each rating: 1 is "very strong effort", 2 is "strong effort", 3 is "moderate effort", 4 is "weak effort", and 5 is "no effort at all".

Very strong effort	**Strong effort**	**Moderate effort**	**Weak effort**	**No effort at all**
1	**2**	**3**	**4**	**5**

Figure 5.2: Example PVM Human Appraisal Questions

An abbreviated example of the recorded ratings for a single respondent is given in Figure 5.3. An example of the responses for a completed trial of 18 configurations can be found in Appendix F.

Name: S34

Recorded Anthropomorphic and Demographic Information

Gender	Height	Seated Height	Weight	Shoe Size	Heel Height	Age*	Current Vehicle	BMI
	in.	in.	lb	in.	in.	Category		
F	66.5	35.8	121	10.25	2.0	3	Taurus X	19.21

Recorded Ratings

		Ingress			Interior				Egress		
#		Acceptable	Ease/effort	Space	Headroom	Left Room	Knee Room	Roominess	Acceptable	Ease/effort	Space
1	6000	1	1	1	1	4	4	2	1	1	1
2	1504	1	1	1	1	4	4	1	2	2	2

Figure 5.3: Example of PVM Ratings from One Respondent

In addition to the human attributes used in the experimental design, several additional human and socio-economic attributes (referred to as the set of respondent demographic attributes) are recorded at the time of the experiment, including seated height, age, and current vehicle ownership. Additionally, as noted in the experimental protocol, the respondent is allowed to adjust the position of the seat, and the respondents' lateral seat positions and seat back angles are also recorded. With completion of the experiments and a complete data set for the 24 respondents, random-effect ordered logit models can be estimated; however, data analysis methods are required to preprocess the data to ensure the most useful models are estimated, as well as to further understand the data collected. In this work, all factor values are normalized on the scale [0, 1] for modeling and analysis, except where noted.

5.3 LATENT CLASS ANALYSIS FOR RESPONSE REDUCTION

It is desired to create predictive preference models for each major sub-system attribute, i.e. ingress, egress and interior roominess, with the collected data; however, in the survey three responses were collected each for ingress and egress (i.e., acceptability, effort, and space) and it is not clear how a single measure of ingress or egress preference can be inferred from the multiple responses. To understand the relationship among the ten responses collected for roominess, ingress, and egress, a correlation matrix (Table 5.5) is estimated, which indicates significant correlation among the responses for each response type.

Table 5.5: Correlation Matrix for Ten PVM Responses

		Ingress			**Roominess**				**Egress**		
		accept	*effort*	*space*	*head*	*left*	*knee*	*room*	*accept*	*effort*	*space*
Ingress	*i_acceptable*	**1**									
	i_effort	**0.828**	**1**								
	i_space	**0.786**	**0.727**	**1**							
Roominess	*headroom*	**0.568**	0.474	**0.664**	**1**						
	leftroom	0.244	0.265	0.259	0.223	**1**					
	kneeroom	0.273	0.294	0.249	0.193	0.522	**1**				
	roominess	**0.562**	**0.527**	**0.660**	**0.774**	**0.549**	0.444	**1**			
Egress	*e_acceptab*	**0.773**	**0.739**	**0.678**	0.442	0.256	0.251	**0.500**	**1**		
	e_effort	**0.700**	**0.779**	**0.629**	0.364	0.216	0.247	0.437	**0.850**	**1**	
	e_space	**0.682**	**0.669**	**0.824**	**0.526**	0.290	0.238	**0.580**	**0.786**	**0.744**	**1**

The following conclusions can be drawn from the correlation matrix and coefficients, *r*:

1. Responses for the three ingress questions are highly correlated ($r > 0.7$).
2. Responses for the three egress questions are highly correlated ($r > 0.7$).
3. Ingress responses are highly correlated to egress responses ($r > 0.6$).
4. Headroom and leftroom are highly correlated to roominess ($r > 0.5$).
5. Roominess and headroom are moderately correlated with ingress and egress ($r > 0.35$).

Based on these observations and the desire to achieve single ingress and egress measures, a formal analysis of the responses is conducted using **Latent Class Analysis** (LCA). LCA is used in this work to identify similarity in rating responses, as opposed to similarity in consumer populations as in the Discrete Choice Analysis literature (Train, 2003). LCA is a general method for data reduction for discrete categorical or ordinal data, analogous to factor analysis used for continuous variables (McCutcheon, 1987). LCA assumes that several discrete variables, such as the three ratings given by each person for ingress or egress, are *indicators* of an overall discrete *latent class* (LC), such as an overall opinion of ingress or egress. LCA provides a single latent class response for each subsystem response (e.g. ingress), based upon the value of the indicators. This predicted LC can be used as the ingress or egress response in a parametric model, such as the discrete choice model, analogous to the use of factor scores resulting from factor analysis for continuous variables.

LCA analysis assumes that the several response indicators are correlated, and seeks to divide the subsystem responses into a number of latent classes such that the indicators are *conditionally independent* within each class. Conditional independence implies that the correlation between the indicators is no higher than “chance” correlation in any class. In order to determine the division of subsystem responses to LCs, the number of LCs must be defined *a priori* for model estimation. The division of subsystem responses is achieved using maximum likelihood estimation to estimate the conditional probabilities of each subsystem response given the LC, and the probability of each LC. A given model can be tested for conditional independence using the likelihood ratio chi-squared test, $L^2 = 2\sum_i n_i \ln\left(\frac{n_i}{\hat{m}_i}\right)$, where n_i is the observed cell frequency in the cross-tabulation table, and $\hat{m}_i$ is the expected cell frequency. The null hypothesis is that the indicators are conditionally independent within each latent class. Another statistic to consider is the index of dissimilarity, *Ds*, given by $Ds = \sum_i abs(n_i - \hat{m}_i)/(2M)$; this measure is the proportion of observations that would have to change cells for the model to fit perfectly, with a generally accepted criterion of $Ds < 0.05$. Among different models (i.e. different assumptions on the *a priori* number of latent classes) which display conditional independence, and $Ds < 0.05$, the Akaike Information Criterion (AIC), which is a function of likelihood and the number of classes (through the remaining degrees of freedom (*df*)), is used for model selection. It is given by $AIC = L^2 - 2df$; the model with the lowest AIC is the preferred model, i.e. the model which balances goodness of fit with the number of model parameters.

LCA is conducted for ingress, assuming the 3 ingress questions (i.e., acceptability, effort, and space) are indicators of each persons overall opinion of the ingress quality. Different numbers of latent classes, between 1 and 10, are assumed. The results of each of the ten models are shown in Table 5.6. Based upon the criteria given for L^2 and *Ds*, models with 1-4 latent classes are not acceptable models (the 5 class model is borderline w.r.t. the *Ds* measure, but will be considered a viable model). The models with between 5 and10 latent classes are compared based on the log-likelihood (*LL*) and the AIC criteria in Table 5.6 and shown graphically in Figure 5.4. The comparison indicates that increases in the number of classes beyond 7 provides no further increase in the *LL*, while the AIC criterion indicates that the 7 class model is preferred considering both the *LL* and the *df*.

Table 5.6: Model Fit Parameters for Differing Class Number Assumptions

LC	*LL*	*df*	L^2	*AIC*	*Ds*
1	-2619.39	88	0.000	1217.79	0.614
2	-2209.27	76	0.000	421.55	0.384
3	-2047.40	64	0.000	121.82	0.258
4	-1969.79	57	0.001	-19.40	0.164
5	-1944.48	50	0.712	-46.01	0.051
6	-1942.11	42	0.592	-44.75	0.046
7	-1936.82	44	0.964	-59.33	0.034
8	-1936.54	38	0.880	-47.91	0.034
9	-1937.36	26	0.279	-22.26	0.035
10	-1937.45	19	0.053	-8.09	0.036

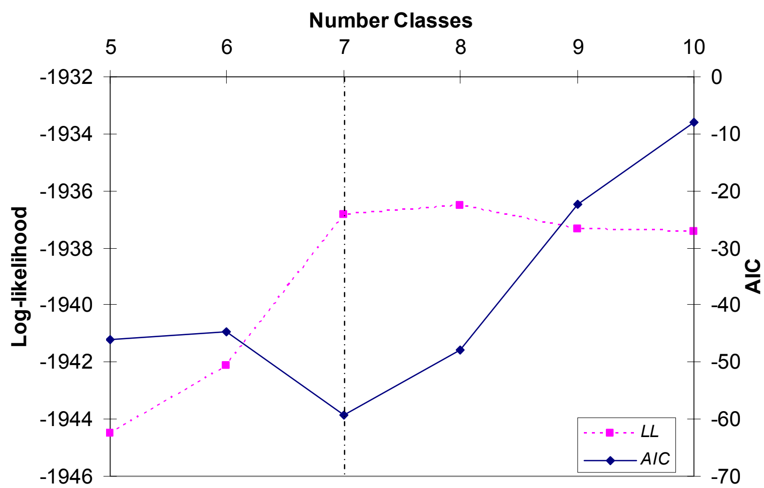


Figure 5.4: Log-likelihood & AIC vs. Number of Classes

The assignment of cases to the latent classes is accomplished using a classification table, in which a case is assigned to the latent class in which it has the highest probability of belonging, as estimated by the conditional probability, π_i, of belonging to each class. Example assignments of select cases of the 3 individual ingress responses to the seven latent classes are shown in Table 5.7.

Table 5.7: Assignment of Cases to Latent Classes

Resp. Number	Ingress Acceptability	Ingress Effort	Ingress Space	Latent Class
101	1	2	1	**1**
103	1	3	2	**2**
105	2	3	2	**3**
108	2	3	3	**4**
110	3	3	3	**5**
115	3	4	4	**6**
120	4	5	5	**7**

The latent class is used as the response variable in the ordered logit model, just as ingress acceptability, effort, or space are used. A comparison using the latent class ingress measure versus the original 3 ingress measures is shown in Table 5.8 using ordered logit models for comparison (numbers in the Table are ordered logit β coefficients)

Table 5.8: Ordered Logit Coefficient Comparison of Ingress Measures

	Ingress Measure				
	acceptability	effort	space	*Range*	*latent class*
HEL_Z	2.017	2.272	1.344	1.34 — 2.27	**1.912**
GRD_Z	-2.026	-2.261	-1.162	-2.26 — -1.16	**-1.875**
StoH	-1.124	-1.268	-0.731	-1.27 — -0.73	**-0.985**
HR_Z	2.270	1.745	2.703	1.75 — 2.70	**2.196**
HR_X	0.550	0.512	0.476	0.48 — 0.55	**0.527**
Stat	-2.814	-2.790	-4.954	-4.95 — -2.79	**-3.150**
Age	3.402	2.878	2.493	2.49 — 3.40	**2.956**
BMI	-2.382	-1.686	-2.003	-2.38 — -1.69	**-2.158**
ρ_0^2	0.1886	0.1873	0.1873		**0.139**

As seen in the models, the parameters in the latent class model are within the *max* and *min* range of the parameters in the models using the three indicators as responses, indicating the latent class is capturing the effect of all three of the ingress indicators. LCA was also conducted for the three egress responses, with a similar result to ingress: the preferred number of classes was found to be 7, with L^2 = 0.99, Ds = 0.045, and AIC = -53.65. LCA was used to create a model for all six ingress/egress responses, assuming ingress and egress responses are indicators of an overall opinion of the vehicle opening. This theory is supported by the fact that high correlation was found between ingress and egress responses; however, a model with an acceptable *Ds* measure is not identified with any number of assumed latent classes. Therefore, it can be concluded that the three ingress responses are indicators of a respondent's opinion of ingress, whereas the three egress responses are indicators of egress opinion.

5.4 Understanding Factor Importance and Rating Style

5.4.1 Analysis of Variation of Rating Responses

In the previous section, latent class analysis was used to understand the relationship among responses, in situations in which multiple responses are assumed to be related to a single unobserved latent factor. In this section, methods are used to understand the relationship among the factors (product and human attributes), respondents, and the responses. In order to understand how the overall variance in the responses is partitioned among the explanatory variables, an **Analysis of Variation** (ANOVA) is conducted. ANOVA analysis is an investigation of how the total sum of squares, SS_T, is decomposed into the sum of squares (SS) contributions from the model, SS_M, and the error, SS_E. The SS_M can be further decomposed to understand the influence of the individual product factors, SS_{TR}, and the individual human factors, SS_R, including the *block effect* attributable to individual respondents. The block effect is the portion of the respondent response not explained by the human factors, with the effect of different configurations and human attributes removed. It is realized in a model as a respondent-specific intercept (i.e. 24 unique intercepts). The magnitude of the sum of squares is a measure of the contribution of each factor and respondent, as well as the error, in explaining the variation in the responses (i.e. the ratings). The main effects ANOVAs for the ingress, egress (the latent classes of Section 5.3 are used for the response), and the four roominess responses are shown in Table 5.9, including the Partial Sum of Squares (*P SS*) and *F* value for each factor.

Table 5.9: ANOVA for the Six PVM Responses

		Ingress		**Headroom**		**Leftroom**		**Kneeroom**		**Roominess**		**Egress**	
		P SS	***F***	***P SS***	***F***	***P SS***	***F***	***P SS***	***F***	***P SS***	***F***	***P SS***	***F***
SS_M	Model	1501.4	*17.42*	1311.8	*27.91*	605.1	*12.88*	369.3	*7.86*	700.5	*14.90*	1269.8	*27.01*
	Gend	2.00	*1.09*	4.22	*9.02*	2.01	*3.32*	0.45	*0.69*	0.39	*0.84*	0.00	*0.00*
	Stat	27.27	*4.96*	**10.44**	*7.44*	**21.86**	*12.06*	9.03	*4.56*	**30.25**	*21.57*	16.40	*3.64*
	BMI	0.01	*0.00*	5.75	*6.14*	**18.66**	*15.44*	14.31	*10.84*	4.41	*4.72*	1.25	*0.42*
	Age	**79.56**	*7.23*	**17.24**	*6.14*	7.97	*2.20*	**48.81**	*12.33*	**25.48**	*9.09*	**95.77**	*10.63*
SS_R	**resp.**	***356.01***	*10.22*	***98.63***	*11.09*	***72.76***	*6.34*	***131.52***	*10.49*	***145.36***	*16.37*	***340.74***	*11.94*
	HNG_X	4.92	*1.34*	0.36	*0.39*	2.47	*2.05*	1.65	*1.25*	0.62	*0.67*	10.69	*3.56*
	ROK_Y	1.06	*0.29*	0.35	*0.37*	**287.97**	*238.3*	**66.86**	*50.66*	46.36	*49.59*	6.71	*2.23*
	HEL_Z	**210.07**	*57.28*	3.22	*3.44*	1.87	*1.55*	31.30	*23.71*	3.55	*3.80*	**275.19**	*91.61*
	GRD_Z	49.27	*13.43*	0.36	*0.38*	1.93	*1.60*	0.57	*0.43*	0.74	*0.80*	75.03	*24.98*
	StoH	18.03	*4.92*	0.46	*0.50*	2.83	*2.34*	2.15	*1.63*	2.20	*2.36*	62.79	*20.90*
	HR_Z	**388.72**	*105.9*	**812.97**	*868.1*	18.10	*14.98*	4.97	*3.76*	**262.16**	*280.4*	**177.86**	*59.21*
	HR_X	46.13	*12.58*	1.66	*1.77*	0.92	*0.76*	3.21	*2.44*	3.00	*3.21*	17.30	*5.76*
SS_{TR}	HR_Y	48.58	*13.24*	8.79	*9.39*	0.75	*0.62*	0.19	*0.14*	9.34	*9.99*	18.44	*6.14*
SS_E	error	*935.3*		*238.8*		*308.1*		*336.6*		*238.4*		*766.0*	
SS_T	total	*2436.6*		*1550.6*		*913.2*		*705.9*		*938.8*		*2035.8*	

- Values unshaded: significant at the 0.05 level

As seen in the ANOVA analysis, not every factor is statistically significant, as measured by the *F*-test (assuming significance at the 0.05 level). This finding serves as a guide to determine which factors to include in the random-effect ordered logit models estimated for each response. The dominant product and human factors are in bold for each response. The analysis also demonstrates the importance of the respondent block effect. The magnitudes of the SS block effect versus the magnitude of the SS human factors are approximately equal, indicating that there is much heterogeneity in responses not captured by the human factors. This unexplained heterogeneity can be attributed to human or socio-economic attributes not recorded and therefore not included in the analysis (e.g. income, usage), or individual *rating styles*. It has been found in previous research that respondents often display distinct rating styles, such as rating systematically high or low, or displaying different scale usage, i.e. scale usage heterogeneity. Systematic high or low rating is related to the mean rating for a given person, μ_i, whereas scale usage heterogeneity is related to the standard deviation of the ratings for a given person, $\sigma_{u,i}$. Attempts have been made to identify these behaviors and control for them in the modeling process (Greenleaf, 1992; Rossi et al., 2001); however, in this work respondents were not given the same set of configurations (i.e. differing **E**) to evaluate due to the blocking (the full experiment contains four unique blocks) and each person is characterized by a different **S**, making comparisons of μ_i and $\sigma_{u,i}$ meaningless.

5.4.2 Analysis of Rating Style using Hierarchical Clustering

For the reasons presented in the previous subsection, a general method to control for rating style must be developed which does not assume respondents have evaluated the same set of configurations, and accounts for the influence of **S**. In the proposed method, the **block effect** will be used as a means of comparison among different respondents. The block effect is the portion of the respondent response not explained by the product or human factors, i.e., it removes the effect of varying **E** and **S** from the analysis. It is estimated in the modeling process as an individual-level intercept, $\beta^0{}_n$, or in a random-effects model as a distribution (typically Normal) of individual-level intercepts with mean 0, and variance τ, i.e. $\beta^0{}_n \sim N(0, \tau)$. A challenge is using the block effect to understand both systematic rating bias, i.e. high rating style vs. low rating style, but also scale usage heterogeneity, i.e. selective use of the provided scale. In this work, these two phenomena are investigated using a *Hierarchical Bayesian* (HB) approach to estimating the block effect for each person, as will be described in the Methodology subsection.

Methodology: The method for using the block-effect resulting from the HB analysis is as follows:

1. Calculate the block-effect for each person for each response using the HB approach.
2. Use the block-effect to calculate ratings *bias, bs_n,* and *scale usage, su_n,* for each person for each response.
3. Perform factor analysis on *bias* and *scale usage* to determine if they are unidimensional and an indicator of rating style, or multidimensional, indicating a missing model parameter (**E** or **S**) or other uncontrolled factor in the experiment.
4. Perform Cluster Analysis on unidimensional rating style terms (i.e. bias or scale usage) to understand the respondent clusters of similar styles (e.g. wide or narrow scale usage).

The key to this approach is the estimation of a *random* block effect for each respondent. In standard MLE, a *fixed* block effect (i.e. individual-level intercept) is estimated for each individual, β^0_n, which contains information about bias only; however, the hierarchical Bayes estimation method allows estimation of a *random* block effect for each person, β^0_{in}, which contains information about both bias and scale usage heterogeneity. In the proposed method, a three-stage hierarchical prior is set for the random block effect as follows:

$$\begin{aligned} prior &= \overbrace{p\left(\beta^0_{in} \mid \beta^0_n, \tau_n\right)}^{level\,1} \overbrace{p(\tau_n) p\left(\beta^0_n \mid \sigma_u\right)}^{level\,2} \overbrace{p(\sigma_u)}^{level\,3} \\ \text{where } \beta^0_{in} &\sim N\left(\beta^0_n, \tau_n\right) \qquad (level\,1) \\ \beta^0_n &\sim N(0, \sigma_u) \qquad (level\,2) \\ \sigma_u &\sim inv.gamma\left(k^0_3, \theta^0_3\right) \quad (level\,3) \end{aligned} \quad . \qquad (5.1)$$

The three levels indicated in the prior are the observation level (*level* 1), the person level (*level* 2), and the population level (*level* 3). This assignment of priors indicates that the block effect for each observation for each person (*level 1*), β^0_{in}, is distributed normally, with mean β^0_n and variance τ_n; the mean block effect for each for person (*level 2*), $\beta^0{}_n$, is normally distributed with mean 0 and variance σ_u; the variance at the population level (*level 3*), σ_u, follows a inverse gamma distribution, with specified parameters k_3^0 and θ_3^0. The hierarchical priors for levels 1 and 2 are illustrated are illustrated in Figure 5.5. The hierarchical prior relaxes the assumption of a fixed block effect for each person, and thus allows for understanding both the mean and variance of the block effect for each person. The mean ($\beta^0{}_n$) provides information about rating *bias* and the variance (τ_n) provides information about the *scale usage*.

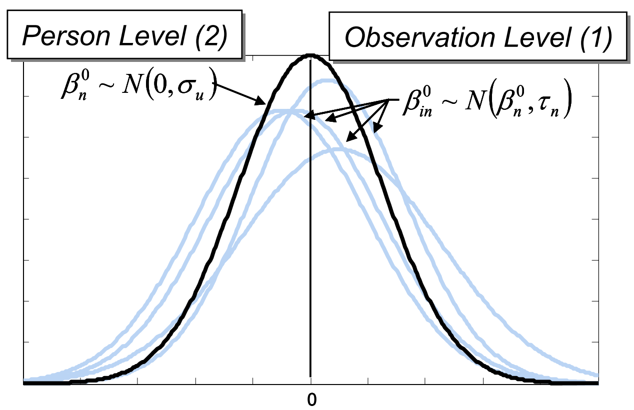


Figure 5.5: Illustration of Bayesian Priors for Block Effects

An issue with such an approach is that the random block term, $\beta^0{}_{in}$, is redundant with the similarly distributed Logistic error term, $\varepsilon_{in} \sim F(0,\sigma_\varepsilon)$, creating potential identification issues. This is overcome in the Bayesian method through the use of the prior which constrains the mean of the $\beta^0{}_n$ to be zero and places a limit on the variance τ_n through the Inv. Gamma prior specification. Although this approach creates models which over-fit the data (i.e. very small error ε_{in}), the intent of such models is to estimate random block effects rather than to predict ratings.

In order to use the information to study rating style, the sign and magnitude of the mean block effect for each person, $\beta^0{}_n$, provides information on rating bias:

$$bs_n = \beta_n^0 - E(\beta_n^0) = \beta_n^0 . \quad (5.2)$$

A positive bs_n indicates a biased high rating style and negative bs_n indicates a biased low rating style compared to the population. To understand scale usage, su_n, a comparison is made between the variance of each individual's set of utilities including the block effect and the utilities with block effect omitted, i.e. the utility variance for the configurations rated by a respondent of a given **S**:

$$su_n = \mathrm{var}(W_n + \beta_n^0) - \mathrm{var}(W_n). \quad (5.3)$$

In this formulation, a positive value for scale usage, sn_u, indicates wider scale usage, while a negative value indicates a narrower scale usage compared to the population of a given **S**.

<u>Analysis of PVM Data:</u> The hierarchical Bayes analysis is used to create models for each of the six responses (using the latent class responses for ingress and egress), and the bias and scale usage is recovered for each respondent for each of the six responses. The bias bs_n and scale usage su_n for each person are investigated to determine if a systematic pattern exists for each person for each of their six responses. Factor analysis is used to determine if bs_n and/or su_n are related to a single latent factor, i.e. the rating style, or if they are related to multiple latent factors, which would be indicative of a missing explanatory variable in the model. The number of latent factors for a given set of indicator variables is determined by the magnitude of eigenvalues of the covariance matrix: the general rule is that only factors with eigenvalues greater than 1.0 be retained (Johnson and Wichern, 2002). In addition to the factor analysis Cronbach's alpha is also calculated, which is a measure of the reliability of indicators to a factor (Cortina, 1993). Cronbach's alpha is a confirmation that the variables are in fact indicators of a single factor, with a value greater that 0.7 generally used as the metric for unidimensionality. The results of the

factor analysis conducted on the bias and scale usage for each of the six responses are shown in Table 5.10. The factor analysis conducted on the bias terms indicates that there is only one latent factor, consistent with the rating style hypothesis, while the analysis of scale usage terms indicates two latent factors, indicating scale usage is not related to an individual rating style. The Cronbach's alpha confirms that the bias indicators are related to a single factor with a high reliability of 0.91.

Table 5.10: Factor Analysis for Block Mean and Variance

	Bias bs_n	**Scale Usage** su_n	
	Factor1	Factor1	Factor2
Ingress	0.789		0.695
Headroom	0.793	0.420	
Leftroom	0.735	0.685	
Kneeroom	0.757	0.566	
Roominess	0.945	0.725	
Egress	0.744		0.783
Average correlation	0.629	0.234	
Cronbach's alpha	**0.911**	**N/A**	

In the case of bias (bsn), cluster analysis is used to identify the dominant rating styles of the respondents, i.e. systematically high or low raters. For the bias terms, cluster analysis is conducted using both the non-hierarchical k-means cluster analysis (assuming three clusters), and complete linkage hierarchical clustering (an a priori assumption of number of clusters is not required) to cluster similar rating styles. Because scale usage does not satisfy the hypothesis of indicating consistent respondent-level rating styles, it is not further analyzed. The results of the k-means cluster analysis, with the bias cluster assignments, are summarized Table 5.11. The results of the complete linkage hierarchical clustering analysis (based upon a Euclidean distance measure) are shown in Figure 5.6.

Table 5.11: *k*-means Cluster Analysis of Bias

	Ingress	Interior Room				Egress	
id	**LC**	**headrm**	**leftrm**	**kneerm**	**room**	**LC**	**Cluster**
resp 1	6.726	6.034	4.812	7.035	6.849	10.450	**2**
resp 2	-1.928	1.073	0.124	1.410	-4.780	-2.846	**1**
resp 3	2.082	0.917	1.367	1.290	1.984	1.775	**2**
resp 4	-2.211	-2.401	-2.132	-5.075	-2.481	-4.880	**1**
resp 5	2.274	-4.292	-3.071	-0.312	3.464	4.473	**2**
resp 6	3.128	10.380	3.181	11.120	6.782	2.186	**1**
resp 7	0.640	2.749	2.033	1.415	-0.327	-0.444	**1**
resp 8	6.022	2.767	2.997	2.121	3.728	3.080	**2**
resp 9	1.482	-6.126	1.515	-5.421	-2.259	-1.126	**1**
resp 10	0.523	0.605	-1.589	-0.844	-1.353	-0.047	**1**
resp 11	-3.777	-0.264	0.044	5.468	-1.913	-6.355	**1**
resp 12	0.743	1.350	-0.527	-0.141	1.956	0.856	**1**
resp 13	0.347	-2.294	3.661	2.345	1.931	-1.827	**3**
resp 14	1.034	-4.605	-4.708	-3.706	-3.308	2.583	**2**
resp 15	-0.453	1.975	-0.992	2.367	0.838	2.701	**3**
resp 16	0.096	1.424	2.201	5.937	6.395	2.631	**2**
resp 17	1.502	-0.477	3.343	-5.648	1.412	1.475	**1**
resp 18	-0.510	3.491	-0.080	-0.828	1.275	-0.796	**1**
resp 19	-3.386	-6.550	-3.246	-4.579	-4.169	-4.461	**1**
resp 20	-3.585	-4.207	-1.401	-4.582	-2.762	-3.193	**1**
resp 21	-0.517	5.074	-2.220	-1.534	0.995	0.673	**3**
resp 22	6.493	7.507	1.612	2.152	3.326	6.292	**2**
resp 23	-2.561	-7.188	-8.716	-10.010	-8.894	-4.011	**3**
resp 24	-2.940	-8.655	-2.681	-1.442	-6.107	-3.180	**1**
resp 25	1.673	6.363	4.902	8.666	6.149	0.533	**2**
resp 26	-5.829	-3.265	0.290	1.714	1.189	-4.731	**3**
resp 27	9.179	2.216	9.650	5.663	4.245	5.409	**2**
resp 28	-0.341	2.761	-1.201	-0.811	-0.791	-1.437	**2**
resp 29	-1.897	-3.244	-4.067	-10.520	-5.406	-0.666	**1**
resp 30	-5.909	-7.334	0.830	0.769	-3.728	-8.125	**3**

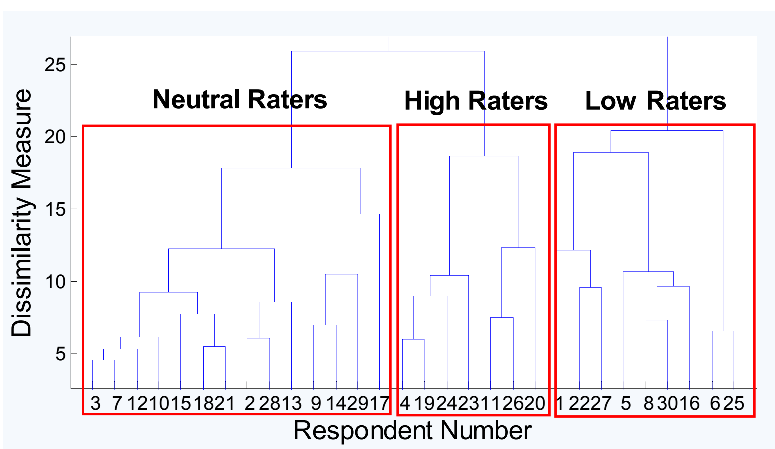


Figure 5.6: Hierarchical Complete Linkage Cluster Analysis

With respect to bias cluster classification, there is strong agreement between the *k*-means and hierarchical clustering, with the hierarchical method confirming the assumption of three unique clusters, and only three cluster classification discrepancies between the two methods. The three cluster model separates the respondents into groups in which each respondent's set of bias terms, bs_n, is close to zero (Neutral Raters), positive (High Raters), or negative (Low Raters).

With the cluster assignments identified for each respondent, a rating style variable, *Styl*, can be defined and added as a respondent level factor in the ANOVA analysis. The style variable is a categorical variable, i.e. [0,1,2], indicating the cluster assignment for each respondent. The result of adding the style factor is shown in Table 5.12. As seen, the random block effect is significantly reduced, and the sum of squares contribution of the style factor is quite large, indicating that a significant portion of the unexplained random respondent effect can be attributed to the rating style of the respondent.

Table 5.12: Inclusion of the Rating Style Variable

	Ingress	**Headroom**	**Leftroom**	**Kneeroom**	**Roominess**	**Egress**
Model	1501.35	1311.83	605.13	369.31	700.45	1269.80
Gend	11.01	15.41	5.82	1.83	11.84	23.07
Stat	22.50	4.52	5.52	2.22	2.68	37.62
BMI	8.25	4.82	16.99	8.75	0.74	14.35
Age	69.77	19.48	11.73	63.06	25.39	69.90
Styl	**69.37**	**23.05**	**19.58**	**45.70**	**51.07**	**61.27**
resp	***143.66***	***56.47***	***61.54***	***95.02***	***61.51***	***106.85***
HNG_X	4.92	0.36	2.47	1.65	0.62	10.69
ROK_Y	1.06	0.35	**287.97**	**66.86**	46.36	6.71
HEL_Z	**210.07**	3.22	1.87	31.30	3.55	**275.19**
GRD_Z	49.27	0.36	1.93	0.57	0.74	75.03
StoH	18.03	0.46	2.83	2.15	2.20	62.79
HR_Z	**388.72**	**812.97**	18.10	4.97	**262.16**	**177.86**
HR_X	46.13	1.66	0.92	3.21	3.00	17.30
HR_Y	48.58	8.79	0.75	0.19	9.34	18.44
error	*935.25*	*238.81*	*308.09*	*336.58*	*238.39*	*766.00*
total	*2436.59*	*1550.64*	*913.22*	*705.89*	*938.84*	*2035.80*

5.4.3 Ordered Logit Model with Rating Style

To better illustrate the use of the style factor, random-effects ordered logit models are estimated with and without inclusion of style variables in Table 5.13, illustrated using the LC Ingress response in Table 5.13. For the modeling process, style is represented using two dummy variables for high rating style, $styl_H$, and neutral rating style, $styl_N$, to represent the three clusters of rating styles.

Table 5.13: Comparison of Ordered Logit Models for LC Ingress

	Without Style		With Style	
	coef.	*t-value*	*coef.*	*t-value*
ROK_Y	0.324	*1.91*	0.320	*1.89*
HEL_Z	2.371	*11.04*	2.371	*11.06*
GRD_Z	-2.207	*-7.40*	-2.217	*-7.42*
StoH	-0.863	*-4.00*	-0.866	*-4.02*
HR_Z	2.822	*13.33*	2.818	*13.32*
HR_X	0.745	*4.22*	0.746	*4.23*
gend	-0.380	*-0.64*	0.619	*1.35*
stat	-0.313	*-1.03*	0.341	*1.39*
BMI	0.229	*0.68*	0.100	*0.40*
age	-1.164	*-1.44*	0.408	*0.64*
$\boldsymbol{styl_H}$			2.307	*5.38*
$\boldsymbol{styl_N}$			1.611	*2.04*
$\boldsymbol{\sigma_u}$	1.48		0.64	
$\boldsymbol{\rho_0^2}$	0.184		0.194	

Random respondent variation is reduced significantly with the inclusion of explanatory variables for ratings style: the fraction of unexplained variance at the respondent level, σ_u, reduces from 1.48 to 0.64 with the inclusion of style variables. In addition, the goodness-of-fit of the model, ρ_0^2 (a measure between 0 and 1), improves from 0.184 to 0.194. This indicates there is less unexplained ratings heterogeneity among respondents with inclusion of the style terms. The benefit of including the style term in the predictive model is a reduction in the variation of the block effect distribution, which results in smaller standard errors in the human/socio-economic model terms and improved understanding of the heterogeneity in rating responses. Assuming the population sampled in the experiment is representative of the population as whole, controlling for the rating style explicitly in the model will provide better predictions than those obtained by integrating over the respondent variance. Also, by knowing people have certain ratings styles, a pre-experiment calibration technique could be used to determine a respondent's rating style before the appraisal is conducted to ensure better consistency in rating style in future experiments (Greenleaf, 1992).

5.5 SMOOTHING SPLINE REGRESSION TO UNDERSTAND RESPONSE BEHAVIOR

With a set of responses determined in Section 5.4 and an understanding of the factor/response relationship determined in Section 5.5, the modeling process can be conducted. A remaining issue is an understanding of the functional relationship between the factors and responses. As was noted in Section 4.1, it has been generally found that a human response to stimuli follows a power law relationship, which provides guidance for determining the form of the product factors in the model. However, in the case of human or socio-economic attributes, such a general theory does not exist. In addition, as noted in Section 5.2, the actual human attributes of each person were collected during the experiment and will be used in model estimation, such that higher ordered terms can be estimated for these terms. A general method to understand the relationship between the response and a factor is the use of **Smoothing Spline Regression**. Smoothing spline regression is similar to piecewise linear regression; however, the breakpoints are connected with polynomials as opposed to lines. Smoothing spline regression is used to better understand the relationship between response and factor, and decide upon the factor forms (e.g., linear, quadratic, cubic) to include in the subsequent random-effects ordered logit models. In this work, smoothing spline linear regression models will be fit to the PVM data and the results will be used to provide guidance in determining factor forms for the ordered logit modeling, in which the utility function is linear additive. This is due to the fact that readily available smoothing spline software is available for linear regression but not ordered logit modeling.

Plots of representative smoothing spline regression relations are shown in Figure 5.7 a), b), and c) (dashed lines represent 95% confidence intervals). These three plots represent the three dominant types of relationships found in the modeling process:

1. *Linear Relationship*: As illustrated in Figure 5.7 a), using the SgRP to Ground factor (GRD_Z) as an example, many of the factors, both product and demographic, have a linear relationship with the rating response.
2. *Power Law Relationship*: As illustrated in Figure 5.7 b), using the SgRP to Roof Z factor (HR_Z) as an example, several of the product factors exhibit a power law relationship. In such a relationship the rate of increase of the rating response decreases as the magnitude of the stimuli increases. This is important to capture in the modeling process and for the vehicle level optimization presented in Chapter 6 because increasing the magnitude of these dimensions, such as HR_Z, results in a diminishing rate of increase in the expected rating.
3. *Critical Level Relationship*: As illustrated in Figure 5.7 c), using Seated Height as an example, several of the demographic attributes display a critical level relationship. In such a relationship, the rating response is constant over certain factor levels, such as very small or very large seated heights, but displays a linear (or higher) relationship over other levels of the factor, such as medium statures. It is important to capture such relationships in the modeling process, particularly if the model is to be used in optimization, since the demographic of the target population for the product may fall in different portions of the plot (e.g. medium seated height), which will determine if a significant relationship exists.

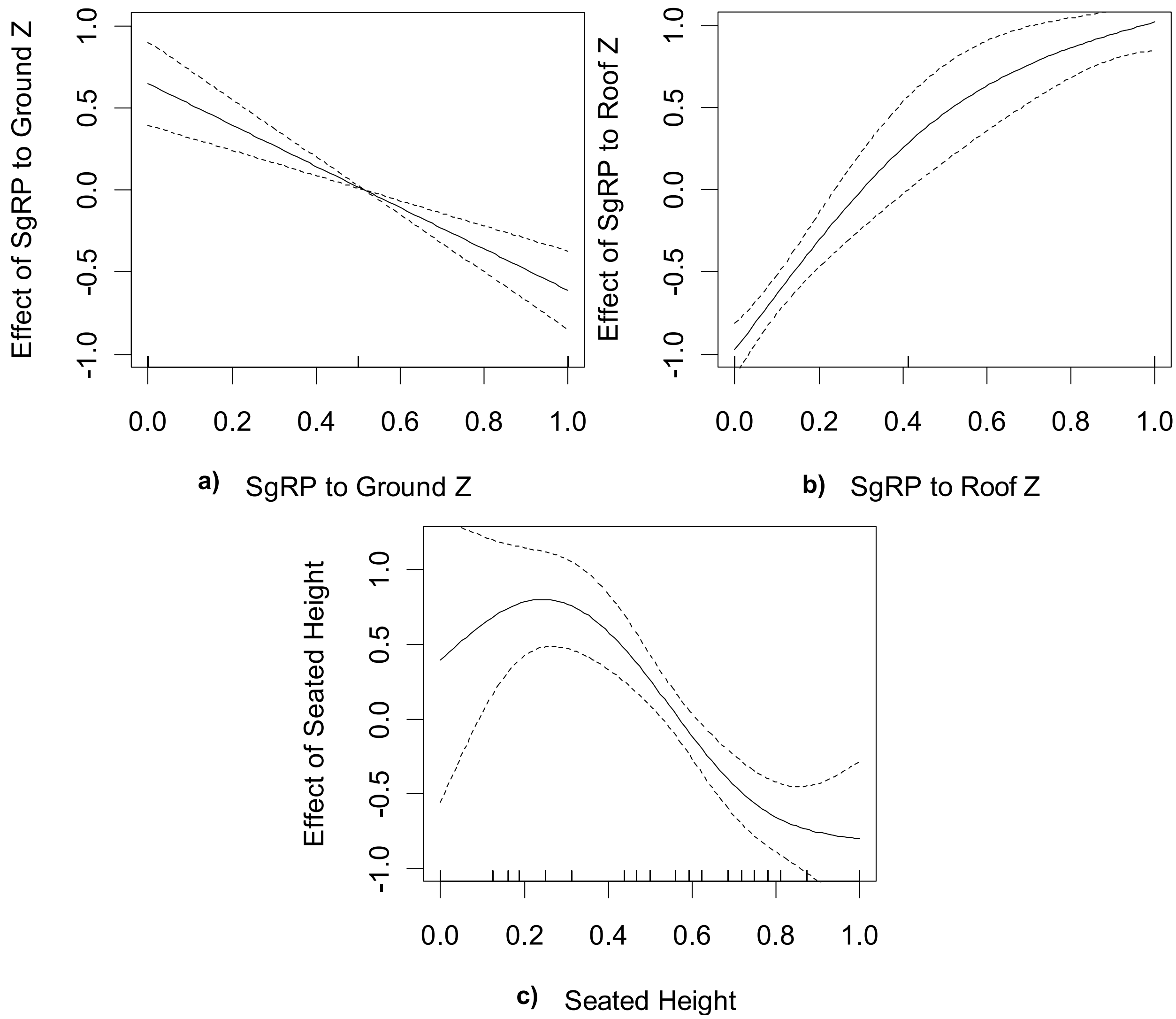


Figure 5.7: Examples of Linear, Power Law, and Critical Level Attributes

With an understanding of the various relationships created using polynomial splines, a straightforward method is required to approximate these relationships in the random-effects ordered logit models described in Section 2.3.3. The three behaviors identified can be approximated closely through combinations of linear, quadratic, and cubic terms. The linear relationship only requires a linear term, the power relationship a linear and quadratic term, and the critical level relationship a linear, quadratic, and cubic term (and thus can only be implemented for the demographic attributes). This method is utilized and demonstrated in a random-effects ordered logit model for the ingress rating response (using the latent class ingress response created in Section 5.3). The results of the model are shown in Table 5.14.

Table 5.14: Random-Effects Ordered Logit for LC Ingress Response

	coef.	*t-value*
ROK_Y	0.29	1.71
HEL_Z	6.95	2.83
HEL_Z^2	-4.75	-1.92
GRD_Z	-1.87	-6.03
StoH	-0.79	-3.58
HR_Z	35.95	3.72
HR_Z^2	-33.15	-3.45
HR_X	5.43	1.81
HR_X^2	-4.73	-1.57
$styl_H$	3.05	4.72
$styl_N$	1.41	2.04
gender	1.25	1.61
age	-5.02	-1.78
age^2	5.12	1.61
BMI	1.37	1.88
seated	832.33	1.68
$seated^2$	-1636.61	-1.66
$seated^3$	805.13	1.65

Using the coefficients from Table 5.14, the effect of three factors shown in Figure 5.7 (i.e. GRD_Z, HR_Z, Seated Height) are plotted in Figure 5.8, and compared to the smoothing spline regression plots to determine if the proposed modeling approximations are close to the smoothing spline regression results, and if the effects of the three factors are similar in the random-effects ordered logit model. In the plots of Figure 5.8, the actual linear, quadratic, and cubic terms of the factor are shown in the legend. In general, the shapes of the factor responses in the random-effects ordered logit model match the shapes of those in the smooth spline linear regression; however, the overall scale is different, since the linear regression model is on the scale of ratings, whereas the RE-OL model is on the scale of utility. Additional higher-ordered terms were tested in the RE-OL model, but the relationships identified in the smoothing spline regression were found to be applicable for the RE-OL, and thus no other higher ordered terms were found to be significant. Similar findings were made with the other collected PVM responses, i.e. headroom, leftroom, kneeroom, roominess, and egress. Based upon this study, smoothing spline regression is an effective method for guiding the selection of terms to be included in the prediction model.

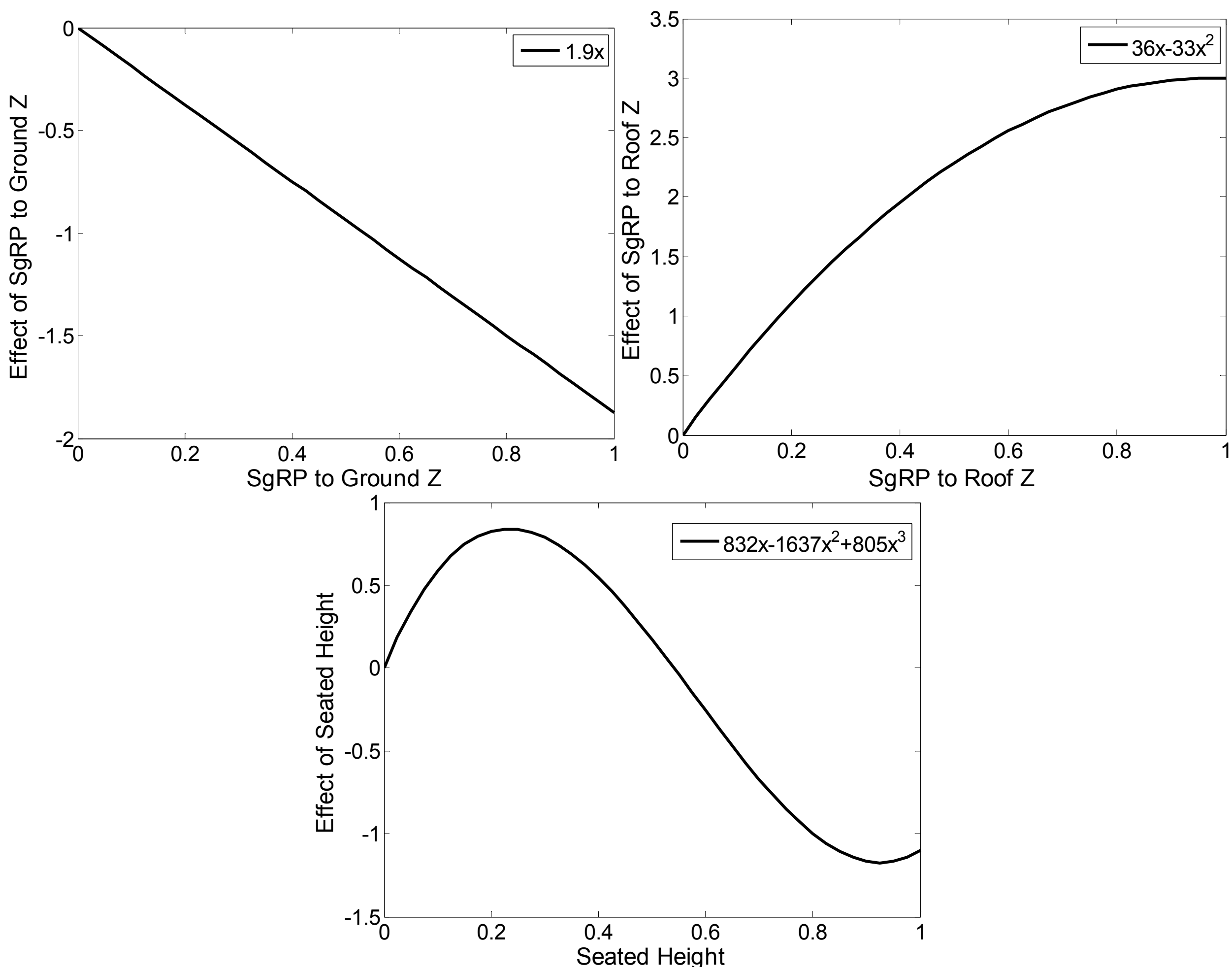


Figure 5.8: Model Factors using Linear, Quadratic, and Cubic Terms

5.6 Random-Effects Ordered Logit Models for Roominess and Ingress/Egress

With the set of responses determined using Latent Class Analysis in Section 5.3, an understanding of the significant responses and the effect of rating style in Section 5.4, and an understanding of the shape of the factor-response relationship in Section 5.5, random-effects Ordered Logit models are fit to the data. The previous methods did not study the effect of the interactions, which will be investigated in the modeling process.

5.6.1 Comparison of Ordered Logit to Linear Regression

Before fitting the ordered logit models, a comparison will be made to linear regression modeling to illustrate the benefits of ordered logit modeling. As an example, the PVM based headroom model was estimated using linear regression by maximum likelihood estimation (MLE), instead of the more common least squares method, to allow comparison of the goodness of fit, ρ_0^2, measures. Using this approach yields ρ_0^2=0.257 for the linear regression (LR) model, compared to ρ_0^2=0.364 for the ordered logit model (OL), indicating the ordered logit model better fits the data from the PVM-based survey. This is illustrated in the histogram of Figure 5.9, comparing LR and OL predictions to the *actual* ratings distribution in the PVM survey.

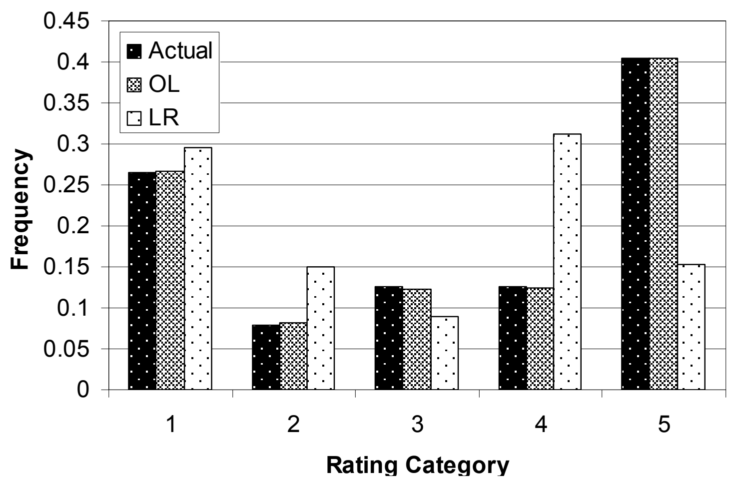


Figure 5.9: Comparison of Ordered Logit and Linear Regression Model Fit

5.6.2 Random-Effects Ordered Logit Models and Interpretation

With confirmation that the RE-OL model is the proper model specification for the collected rating data, models are created with the specific task to include the effects of significant interactions. The two RE-OL models for the latent class ingress and egress responses with significant terms, including interactions, are shown in Table 5.15 (OL cut-points omitted). RE-

OL models for the roominess responses, i.e. headroom, leftroom, kneeroom, and roominess, are shown in Table 5.16 (OL cut-points omitted).

Table 5.15: Ingress-Egress RE Ordered Logit Models

LC Ingress

	coef.	*t-value*
Gend	**-33.53**	-2.35
Seated	**-786.25**	-1.43
$Seated^2$	**1399.74**	1.29
$Seated^3$	**-632.79**	-1.18
Age	**-51.67**	-3.33
Seated·Gend	**35.65**	2.37
Seated·Age	**55.64**	3.37
ROK_Y	**0.28**	2.18
HEL_Z	**-16.75**	-3.80
HEL_Z^2	**-5.43**	-2.01
GRD_Z	**-1.75**	-1.33
StoH	**-4.09**	-5.53
HR_Z	**45.78**	4.20
HR_Z^2	**-47.01**	-4.33
HR_X	**1.07**	5.48
HR_Y	**7.71**	2.12
HR_Y^2	**-10.49**	-2.99
$ROK_Y \cdot HEL_Z$	**2.78**	1.39
$ROK_Y \cdot GRD_Z$	**-5.48**	-2.52
$ROK_Y \cdot HR_Y$	**4.84**	2.97
$HEL_Z \cdot GRD_Z$	**6.21**	2.25
$HEL_Z \cdot StoH$	**4.66**	5.09
$HEL_Z \cdot HR_Z$	**22.94**	7.69
$Styl_H$	**2.16**	4.72
ρ	***0.155***	
$\rho^0{}_2$	***0.262***	

LC Egress

	coef.	*t-value*
Age	**-6.64**	-1.26
Age^2	**6.57**	1.23
ROK_Y	**3.53**	2.27
HEL_Z	**-11.52**	-2.57
HEL_Z^2	**-5.23**	-1.79
GRD_Z	**-0.74**	-0.34
StoH	**-4.27**	-2.54
HR_Z	**41.41**	3.64
HR_Z^2	**-41.13**	-3.64
HR_X	**1.48**	1.05
$ROK_Y \cdot HEL_Z$	**3.93**	1.83
$ROK_Y \cdot GRD_Z$	**-7.73**	-1.95
$ROK_Y \cdot StoH$	**-2.32**	-1.28
$ROK_Y \cdot HR_X$	**-3.17**	-1.64
$HEL_Z \cdot GRD_Z$	**6.77**	2.4
$HEL_Z \cdot StoH$	**5.74**	6.04
$HEL_Z \cdot HR_Z$	**12.89**	4.34
$HEL_Z \cdot HR_X$	**2.95**	2.55
$Styl_H$	**2.70**	6.12
ρ	***0.214***	
$\rho^0{}_2$	***0.274***	

Table 5.16: Roominess RE Ordered Logit Models

Roominess

	coef.	*t-value*
Seated	**-11.08**	-2.82
Age	**-44.67**	-2.99
Age^2	**16.17**	2.90
Seated·Age	**30.42**	2.10
ROK_Y	**15.47**	1.79
ROK_Y^2	**-29.31**	-3.53
HEL_Z	**-2.33**	-2.54
HR_Z	**50.53**	4.49
HR_Z^2	**-57.09**	-5.18
HR_Y	**6.93**	1.91
HR_Y^2	**-6.25**	-1.72
$ROK_Y \cdot HEL_Z$	**-3.48**	-2.26
$ROK_Y \cdot HR_Z$	**28.76**	6.34
$HEL_Z \cdot GRD_Z$	**-5.11**	-2.65
$HEL_Z \cdot HR_Z$	**7.44**	3.97
$Styl_H$	**2.51**	5.07
ρ	***0.216***	
ρ^0_2	***0.408***	

Headroom

	coef.	*t-value*
Gend	**6.65**	2.81
Seated	**901.88**	2.03
$Seated^2$	**-1896.38**	-2.16
$Seated^3$	**990.29**	2.28
Age	**-16.52**	-2.39
Age^2	**18.07**	2.67
BMI	**2.13**	2.06
Gend·Age	**-10.23**	-2.52
HNG_X	**-2.76**	-1.74
ROK_Y	**1.63**	2.04
HEL_Z	**-7.72**	-2.29
HR_Z	**83.34**	6.75
HR_Z^2	**-74.99**	-6.15
HR_Y	**1.12**	4.53
$HNG_X \cdot ROK_Y$	**-3.92**	-1.69
$HEL_Z \cdot HR_Z$	**9.58**	2.18
$Styl_H$	**2.17**	3.51
ρ	***0.251***	
ρ^0_2	***0.536***	

Leftroom

	coef.	*t-value*
Seated	**1360.97**	1.83
$Seated^2$	**-2694.32**	-1.83
$Seated^3$	**1332.40**	1.82
BMI	**-8.49**	-1.32
BMI^2	**7.80**	1.17
ROK_Y	**4.84**	16.40
GRD_Z	**-0.45**	-1.61
StoH	**-0.47**	-2.10
HR_Z	**1.26**	6.31
ρ	***0.263***	
ρ^0_2	***0.307***	

Kneeroom

	coef.	*t-value*
Age	**-11.99**	-1.66
Age^2	**15.12**	2.06
BMI	**-1.31**	-1.27
ROK_Y	**2.95**	4.64
HEL_Z	**5.37**	2.09
HEL_Z^2	**-3.88**	-1.51
HR_Z	**0.59**	2.81
HR_X	**1.79**	1.37
$ROK_Y \cdot HR_X$	**-2.17**	-1.12
$Styl_H$	**1.95**	3.45
ρ	***0.333***	
ρ^0_2	***0.225***	

As seen in comparing among the models for ingress-egress and roominess, factors thought to be primarily associated with roominess, such as HR_Z, HR_X, and HR_Y, appear in the ingress-egress models, and factors thought to be associated with ingress-egress, such as HNG_X, appear in the roominess models. The reason for this could be two-fold: respondents' opinions of ingress-egress also influence their opinions of roominess, or the factors actually contribute to the ingress-egress or roominess experience directly. Different demographic attributes and demographic attribute interactions appear in the models. For example, gender, seated height, and age appear in the ingress model, whereas only age appears in the egress model. This could be explained by the fact that it is generally easier for respondents to exit the vehicle than enter the vehicle, and thus factors such as seated height and anthropomorphic gender differences do not influence the rating for egress as they do for ingress.

5.6.3 Effect of Explicitly Modeling Heterogeneity

The effect of including both systematic (**S**) and random heterogeneity (σ_u) on the rating predictions can be seen using a simple example in which the headroom model is re-estimated without **S**, without σ_u, and without both **S** and σ_u. The models estimated with different representations of heterogeneity are compared in terms of their ability to match the first four moments of the actual ratings distribution, as shown in Table 5.17.

Table 5.17: Comparison of Inclusion of Heterogeneity in Model

	OL without S		***OL with S***		***RE-OL without S***		***RE-OL with S***	
	Sample	*Error*	*Sample*	*Error*	*Sample*	*Error*	*Sample*	*Error*
Mean	**3.321**	-0.15%	**3.322**	-0.14%	**3.315**	-0.35%	**3.318**	-0.25%
Variance	**2.049**	-26.41%	**2.315**	-16.86%	**2.203**	-20.85%	**2.389**	-14.20%
Skewness	**-0.109**	-68.19%	**-0.249**	-27.35%	**-0.168**	-50.93%	**-0.264**	-22.94%
Kurtosis	**1.178**	-18.92%	**1.349**	-7.11%	**1.276**	-12.14%	**1.360**	-6.36%
ρ_0^2	***0.380***		***0.483***		***0.518***		***0.536***	

A primary difference among the models can be seen in the goodness of fit, ρ_0^2, which increases as either systematic, random, or both, types of heterogeneity are included in the model. The effect of the improved model goodness of fit results in improved moment matching, as can be seen in the decreasing error in each moment as heterogeneity is more explicitly represented. An exception to this finding is the ability of each of the models to match the mean, since all models are unbiased estimates of the mean; improvements resulting from modeling heterogeneity are only seen in matching the higher moments. The improved model fit can be seen graphically using a comparison of histograms of the OL model without **S** and σ_u versus the OL with **S** and σ_u (i.e. RE-OL) in Figure 5.10. It can be seen that the OL model without **S** and σ_u does a poor job of matching the actual ratings distribution, whereas the OL model with **S** and σ_u is much better at matching the actual ratings distribution.

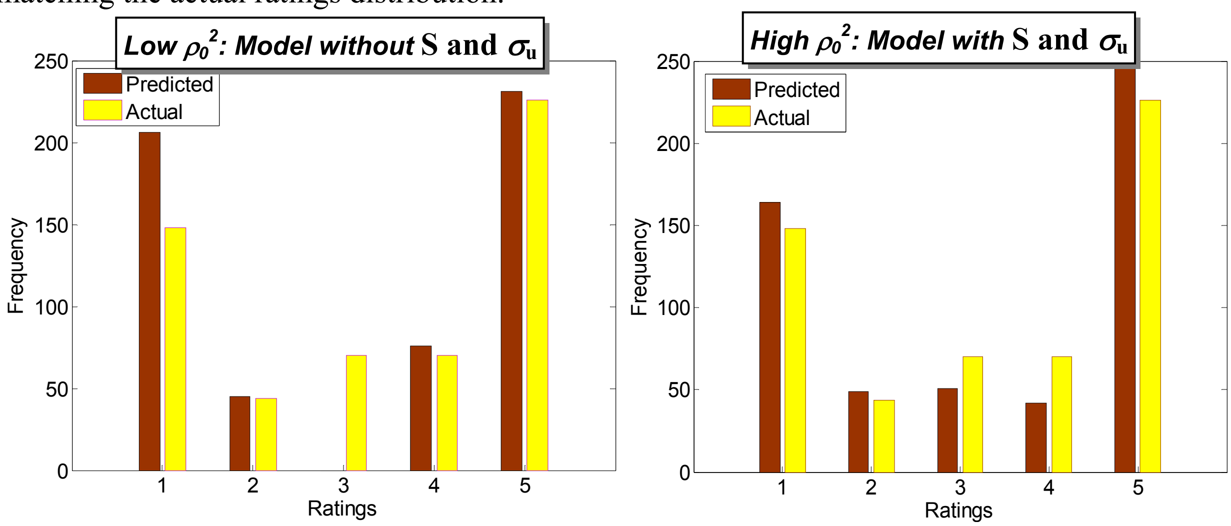


Figure 5.10: Comparison of Lowest to Highest Goodness of Fit Model

5.7 ALTERNATE DATA ANALYSIS METHODS

While the focus of this work is to estimate parametric models to be used in the hierarchical choice modeling approach, other model forms are also investigated to gain further insight into the data, and to confirm the RE-OL modeling approach. Data mining **machine learning** methods are investigated to determine if such methods can aid or replace traditional statistical modeling methods, such as the ordered logit model. The data mining methods investigated in this work are classification methods, i.e. methods to predict the ratings class (i.e. 1-5 rating) based upon the attribute values. Two applicable approaches to classification data mining are investigated: a C4.5 Decision Tree and a Bayesian Network. The five classes to be estimated are the five (or 4) ratings: 1, 2, 3, 4, 5. An issue with these approaches is that the mainstream implementation of the Decision Tree and Bayesian network is based upon the assumption that attribute values, **Z**, including both product attributes **E** and human attributes **S**, are *discrete* variables. This is not a significant issue for the PVM product factors, which only assume three levels and therefore can be considered discrete; however, they will be treated as nominal as opposed to interval (or ratio) level variables in these analyses. The demographic attributes are generally continuous interval level variables (except gender), and thus will be divided into discrete categories based upon their continuous values.

5.7.1 Decision Tree for Ratings Classification

A Decision Tree is created using the PVM dataset. A decision tree is created through a process in which a number of observations or cases, *s*, within a training data set, *Tn*, are *classified* into a number subsets with respect to a class variable (i.e. a response), R_p, based upon a rule concerning a "splitting" attribute value, **Z** (i.e. a product or demographic attribute). The tree building process continues to add branches until no further information can be gained. The decision tree is then pruned using a cost criterion to maximize the classification accuracy relative to the complexity of the tree (Witten and Frank, 2005). The goal is to create a non-parametric model capable of predicting the class (rating) based on the value of the attributes. In this respect, a decision tree is similar to the ordered logit model, in that the goal is to predict a rating based upon attribute values (e.g. HR_Z, Stature). Therefore, a decision tree can be viewed as a non-parametric alternative to the ordered logit model.

The rule for selecting a splitting attribute is determined by selecting the attribute which maximizes the information gain for a given split, *gain*(Z), based on a measure of information, *info*:

$$\max\{gain(Z) = info(s) - info_Z(s)\}, \tag{5.4}$$

The average information, *info*(s), is expressed in the units of bits and can be calculated from the number of occurrences of a particular class R_p in s, given as *freq*(R_p,s):

$$info(s) = -\sum_{p=1}^{P}\left(\frac{freq(R_p, s)}{|s|}\right)\cdot\log_2\left(\frac{freq(R_p, s)}{|s|}\right)[bits]. \tag{5.5}$$

The average information is calculated over the entire training set, $s = Tn$, for the first split, and over number of cases at the root attribute for all subsequent splits. The information associated with a split on attribute Z, $info_Z(s)$, is given by:

$$info_Z(s) = \sum_{j=1}^{J}\frac{|s_j|}{|s|}\cdot info(s_j) \tag{5.6}$$

where s_j is a subset of cases created by performing J splits on attribute Z. The tree building process continues to add branches until no further information can be gained. The decision tree is then pruned using a cost criterion to maximize the classification accuracy relative to the complexity of the tree. As an example, a simplified decision tree is built for the *headroom* response as shown in Figure 5.11 (variables un-normalized for clarity), and the model goodness-of-fit statistics are shown in Table 5.18. All units in the figure are *cm* except for BMI in standard units kg/m^2; the number in the box is the rating class, and the number below the rating class is the number of predicted observations belonging to each rating class based on the classification rule.

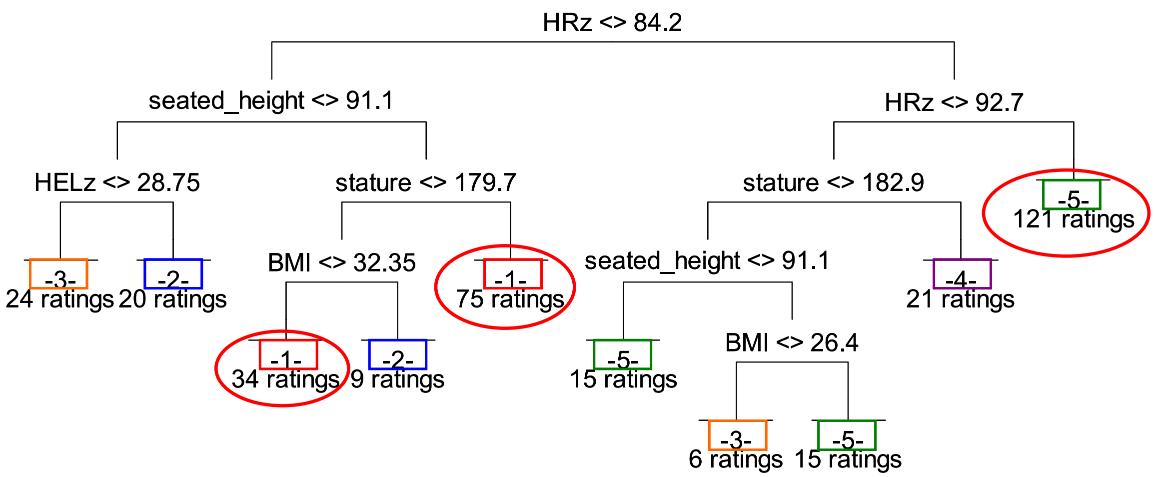


Figure 5.11: C4.5 Decision Tree for Headroom Rating

Table 5.18: Summary Statistics for the C4.5 Decision Tree

1	2	3	4	5	<-- classified as
141	0	7	3	1	**1**
20	1	14	8	2	**2**
22	8	22	14	8	**3**
3	1	12	25	35	**4**
0	0	3	4	228	**5**

Correctly Classified	**417 (71.65%)**
Incorrectly Classified	**165 (28.35%)**
Root mean squared error	**0.293**
Kappa statistic	**0.595**

The decision tree can provide insights not gained readily in traditional parametric modeling methods, such as ordered logit modeling. One such observation is that 85% of configurations receiving a rating of 5 occur when HR_Z (E_6) is at its maximum value, regardless of other product or demographic attribute values. This indicates that increasing HR_Z is a straightforward method for achieving a high headroom rating. While the HR_Z attribute is dominant in the ANOVA analysis, the decision tree provides information regarding how specific attribute values influence specific rating frequencies. Another interesting observation is that the combination of low values of HR_Z coupled with respondents of large seated height and overall stature account for the majority of the low ratings (69%). A more enlightening finding is that HR_Z at its minimum value coupled with high seated height, low stature, and low BMI account for 31% ratings of 1. This could possibly be explained by the seating position of low BMI respondents versus high BMI respondents, because low BMI respondents may position their seat differently in terms of lateral position and tilt angle, leading to a different experience of headroom for a given configuration for respondents of the same stature. This can be captured in a model through the inclusion of a BMI-seated height interaction term, which should be positive in sign. In conclusion, these findings indicate that HR_Z, BMI, seated height, stature and a BMI-seated height interaction are important variables in the parametric modeling process.

A decision tree was also conducted for *Ingress Effort*, as shown in Appendix G. As seen previously in the ANOVA analysis for ingress, there is not a dominant attribute in explaining ingress ratings, as for HR_Z in the headroom model. The tree indicates that ROK_Y, HEL_Z, GRD_Z, StoH, HR_Z, and HR_X are important to classifying the ratings, consistent with the ordered logit model for ingress. Seated height and gender appear to be the most important demographic attributes, with age, stature and gender appearing to be less important. Interactions of seated height and gender and seated height and BMI should be tested in the modeling process. In general, large levels of HEL_Z (E_3) and HR_Z (E_6) lead to branches with high ratings for ingress, which is expected because these two variables control the vertical height of the door opening.

Decision trees can be created for the other attributes as well to better understand the relationship between factors and rating responses. The decision tree is useful for understanding important variables to consider in the modeling process, and can also point to trends not seen easily in a parametric modeling process. On the other hand, the decision tree does not account for individual rating styles, or allow for standard statistical interpretation of the factors, which are added to the model based upon the information gain criterion of Eq. (5.4). Another issue is that the decision tree does not provide continuous functions of the factors, but rather classifies based on threshold values of the factors (i.e. <, >), making it an inefficient tool to study the effect of changing attribute values upon ratings. Additionally, the hierarchical choice modeling approach to be developed in Chapter 6 relies upon the use of parametric models in the framework. Based upon the advantages and disadvantages of the decision tree, it is considered a preprocessing tool for the ordered logit modeling process rather than a competing modeling methodology.

5.7.2 Bayesian Network for Ratings Classification and Associations

The use of Bayesian networks in analyzing and modeling the PVM data is investigated in this subsection. The Bayesian network can be used in two distinct implementations, **supervised** and **unsupervised**. In the supervised implementation, the Bayesian network is used a classifier in which attribute values are used to predict a class, e.g. a rating. In the unsupervised implementation, no assumption is made regarding responses (dependent variables) and factors (independent variables), but rather the network identifies dependent and independent variables. The two implementations of the Bayesian network will be investigated.

Supervised Bayesian Network: The supervised Bayesian network is a classifier in which a class R_i, such as a rating, is predicted based upon the conditional probability of the attribute $\mathbf{Z}$ values. In the supervised network, the class to be predicted is defined *a priori*. Therefore, the Bayesian network is used as a method to determine the probability of being in each class R_i, (i.e. each rating category) for each observation (i.e. each respondent), just as in the ordered logit model. The probability of the class assuming a certain value R_i (i.e. rating of 1,2,3,4 or 5), given a set of attribute values $\mathbf{Z}$, is determined using Bayes law:

$$\Pr\left[R = R_p \mid Z_1, \ldots, Z_J\right] = \frac{\Pr\left\{Z_1, \ldots, Z_J \mid R = R_p\right\} \cdot \Pr\left\{R_p\right\}}{\Pr\left\{Z_1, \ldots, Z_J\right\}} \tag{5.7}$$

The Bayes network uses the assumption of conditional independence. Conditional independence requires that each attribute, Z_j, is conditional only on the immediate, or parent, attributes and not upon the distant relative attributes (i.e. grandparents, great-grandparents, etc.). Using this assumption, the conditionally independent probabilities can be multiplied to find the joint probability of **Z**:

$$\Pr\left\{Z_1, \ldots, Z_J \mid R = R_p\right\} = \prod_{j=1}^{J} \Pr\left\{Z_j \mid R, \text{parents}\left(Z_j\right)\right\} \tag{5.8}$$

Using the assumption of conditional independence, Eq. (5.7) can be written (omitting the normalizing term $\Pr\left\{Z_1, \ldots, Z_J\right\}$):

$$\Pr\left[R = R_p \mid Z_1, \ldots, Z_J\right] = \Pr\left\{R_p\right\} \cdot \prod_{j=1}^{J} \Pr\left\{Z_j \mid R, \text{parents}\left(Z_j\right)\right\} \tag{5.9}$$

Eq (5.9) demonstrates that the supervised Bayesian network is a form of non-parametric regression. The Bayesian network ratings predictions can therefore be directly compared to the ordered logit regression predictions of Eq.(4.8):

$$\Pr\left[R = R_i \mid Z_1, \ldots, Z_J\right] = F\left(k_p - \boldsymbol{\beta}'\mathbf{Z}\right) - F\left(k_{p-1} - \boldsymbol{\beta}'\mathbf{Z}\right)$$

where F is the cumulative logistic distribution. An advantage of the Bayesian network is that no assumptions are made on the error distribution (i.e. logistic or normal distribution) because it is non-parametric. The rating predictions from both the Bayesian network and the ordered logit model are compared in Figure 5.12.

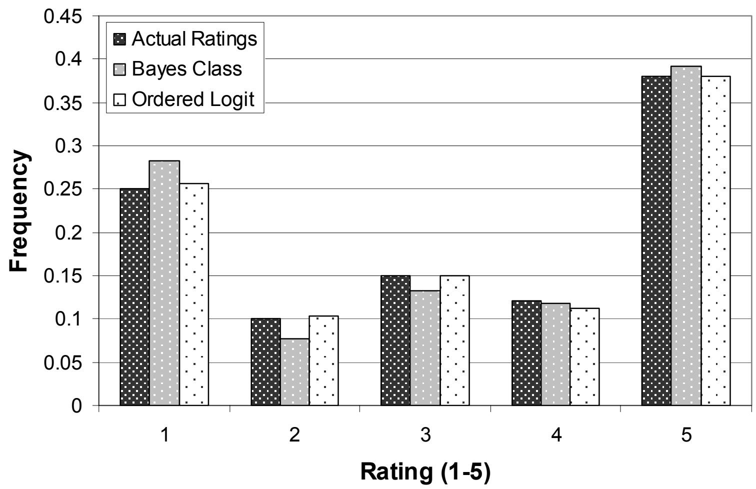


Figure 5.12: Comparison of Ratings Predictions

As seen in the histogram, the Bayesian network results in similar ratings classification to the ordered logit model. The superior performance of the ordered logit model can be attributed to the enforcement of the ordinal constraint (i.e. adjacent ratings are correlated), as opposed to the nominal assumption of the Bayesian network, and the discretizing of attributes to nominal categories in the Bayesian network. The Bayesian network also identifies the conditional relationships as shown in Figure 5.13, with arrows going from the parent attributes to the child attributes. In this case the effect of E_6 (i.e. HR_Z) is conditional on the value of seated height, indicating that an interaction term of HR_Z·Seated Height should be investigated.

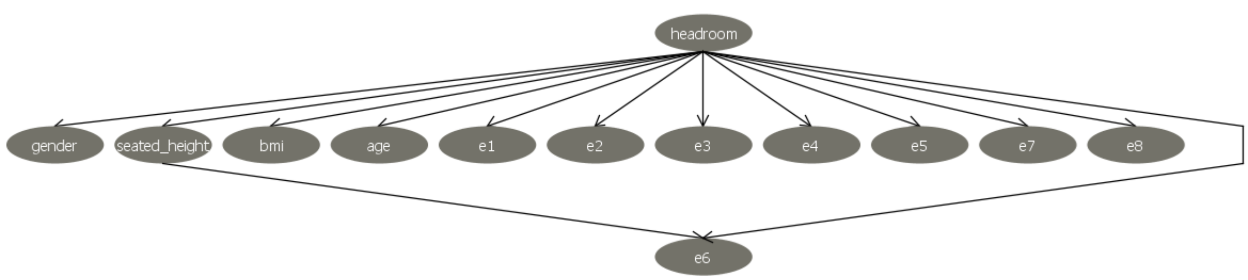


Figure 5.13: Supervised Bayesian Network Graph for Headroom

For datasets in this work (e.g., the PVM data) which are structured for random-effects ordered logit modeling and do not have missing values, the Bayesian network offers few advantages over ordered logit modeling. Because the supervised Bayesian network is non-parametric and is a form of machine learning, like the decision tree, it can be viewed as a pre-processing tool to better understand relationships in the data.

Unsupervised Bayesian Network: As opposed to the supervised Bayesian network which can be viewed as an alternative to the ordered logit model, the unsupervised Bayesian network is used to understand relationships in the data. In the unsupervised Bayesian network, no distinction is made between responses and factors with the goal of understanding the relationships within the data sets rather than predicting a class. For this reason, the focus is upon identifying the joint distribution of attributes **Z** (in this case the rating response is considered another attribute) in terms of the conditional distributions:

$$\Pr[Z_1,\ldots,Z_J] = \prod_{j=1}^{J} \Pr\{Z_j \mid \text{parents}(Z_j)\} \tag{5.10}$$

An unsupervised Bayesian network for the PVM dataset (with only the headroom response included) is shown in Figure 5.14, with the arrows going from parent to the child attributes.

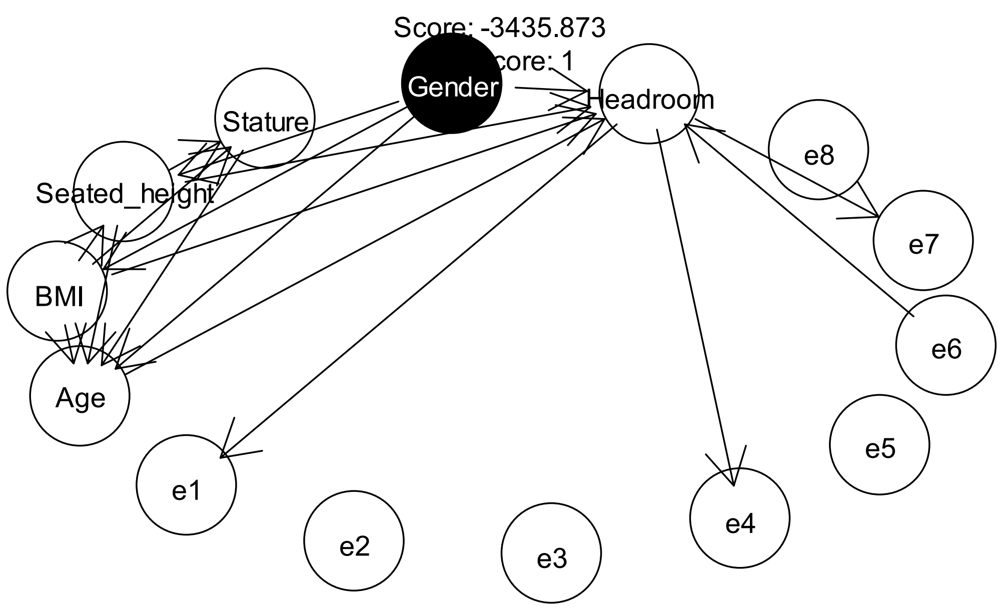


Figure 5.14: Unsupervised Bayesian Network Including Headroom

The joint distributions of attributes **Z** are expressed as the product of the conditional probabilities. The conditional probabilities identified to have parent attributes are as follows:

- [Headroom|Gender;Seated height;BMI;Age;E_6]
- [Age|Gender;Stature;Seated height;BMI]
- [Stature|Seated height;BMI]
- [Seated height|Gender;BMI]
- [BMI|Gender]
- [E_1|Headroom]
- [E_4|Headroom]
- [E_7|Headroom]

The relationships identified indicate that not all demographic attributes collected are independent, i.e. there are correlations among the demographics. For example, age is conditional upon the values of gender, height, seated height, and BMI, which can be confirmed using a regression analysis as shown in Table 5.19.

Table 5.19: Regression of Age on Other Demographic Attributes

Age	coef.	*t-value*
gender	**-0.364**	*0.00*
height	**-0.232**	*0.03*
seated height	**-0.610**	*0.00*
BMI	**0.177**	*0.00*
constant	0.985	*0.00*

Such a regression is hard to interpret from the cause-effect standpoint as assumed in regression, since the prediction of age based upon other demographic attributes does not make intuitive sense. On the other hand, prediction of height based upon seated height and BMI is more plausible. In general, the Bayesian network is identifying *correlations* within the data, which may or may not represent proper cause-effect regressions.

The issues created by correlation among demographic attributes in the modeling process are *redundancy* and *suppression*. Redundancy and suppression occur when certain correlation patterns are present among multiple independent (**Z**) and dependent variables (**Y**). An example of redundancy can be seen in Figure 5.15 in which two *Z* and one *Y* are positively correlated. When redundancy is present and one of the *Z* is removed, the magnitude of the remaining *Z* increases. An example of suppression is also shown in the figure, in which two *Z* are positively correlated with the *Y*, but the *Z* negatively correlated with each other. When suppression is present and one of the *Z* is removed, the magnitude of the remaining *Z* may decrease or the sign may change. If there are more than two *Z*, the patterns are very difficult to diagram, and both redundancy and suppression may occur.

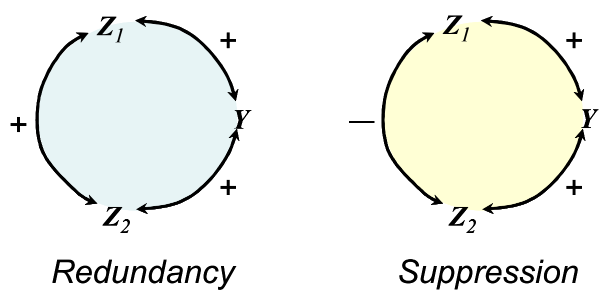


Figure 5.15: Correlation Patterns Creating Redundancy and Suppression

In the context of the hierarchical modeling framework, the predictive ability of a model is not hurt by redundancy or suppression, but these phenomena make it difficult to interpret the effect of the demographic attributes individually. Coefficient interpretation is important for model validation, ensuring that model coefficients match or can be explained from an understanding of the problem. Another issue is that in the hierarchical choice modeling approach developed in

Chapter 6, multiple data sets may be merged with different combinations of redundancy and suppression, leading to issues in model estimation with the combined data.

An approach to the problem is to assume the demographics are related to a smaller number of uncorrelated latent factors. Therefore, a latent variable analysis is conducted using the demographic attributes collected in the PVM appraisal. It is found that there are two significant factors. Using the iterated principle factor (IPF) method of solution and performing an orthogonal rotational (so that the 2 factors are uncorrelated) gives the result shown in Table 5.20.

Table 5.20: Factor Analysis for Collected Demographic Attributes

	Factor 1	Factor 2	Uniqueness
gender	**-0.765**	-0.420	0.335
height	**0.861**	-0.196	0.169
seated height	**0.970**	-0.072	0.033
BMI	0.059	0.269	**0.929**
age	-0.092	**0.724**	0.447

The interpretation is that gender, height, and seated height are related to a latent factor related to size, primarily vertical size. Factor 2 consists primarily of age, while BMI is primarily unique. Possible modeling solutions include either using factor scores for correlated demographic attributes, or using either height or seated height, but not both, in a model together with BMI and age, which are not highly correlated with height or seated height. Gender can be used in the modeling since it is not as highly correlated as height and seated height; however, it may cause some level of suppression or redundancy in the model and may not be easily interpreted. Factor scores can be used in the ordered logit models as well, but in the Bayesian hierarchical modeling framework, in which multiple data sets are to be used and models to be updated, the use of factor scores will complicate the process and is not recommended in this application.

5.8 DISCUSSION AND SUMMARY

Methods for the analysis of human appraisal experiments to understand and predict consumer preferences for new or existing product designs were developed in this chapter. The methods developed are for the purpose of preprocessing data, reduction of data, capturing respondent heterogeneity, and creating random effects ordered logit models to understand consumer preferences and enable prediction of preferences for new product offerings. Latent Class Analysis is shown to be effective for combining several responses given by a consumer during an appraisal into a smaller number of latent classes related to their overall opinion of key product features. ANOVA analysis is used to understand the relative importance of the product and human attributes on the different rating responses provided in the survey. In these analyses, the respondent block effect, or unexplained respondent heterogeneity is found to be large. Cluster analysis of the block effect is used to identify systematic ratings styles of the respondents, which explain a significant portion of the unexplained heterogeneity. Adding new variables to control for rating style in the modeling process significantly reduces the unexplained heterogeneity. The use of smoothing spline regression is demonstrated to be an effective tool to understand the shape of the response-factor curve and guide the form of factors (i.e. linear, quadratic, cubic) to be introduced in the subsequent ordered logit modeling.

With data preprocessing, response reduction, and an understanding of respondent heterogeneity, random effects ordered logit models are estimated for each response. The importance of interactions and the benefits of explicitly modeling systematic heterogeneity and random heterogeneity are demonstrated in the ability of the distribution of the predicted ratings to match the actual distribution of ratings, an important feature of a model to be used to predict preferences for different populations and different designs. Machine learning methods from data mining are also applied to the PVM data. The decision tree provides additional insights into the relationship among the product factors, human factors, and rating responses not easily identified in the parametric ordered logit model. The unsupervised Bayesian network provided insights into the relationships among the human factors not easily seen in methods such as correlation analysis. The methods developed in this section are crucial to not only understanding consumer heterogeneity in human appraisal experiments, but also creating predictive models for use in the Bayesian Hierarchical Choice Model introduced in Chapter 6.

Chapter 6

BAYESIAN HIERARCHICAL CHOICE MODELING FOR ENGINEERING DESIGN

In Chapter 3, the PAFD tool was presented for engineering design decision making, built upon Decision-based Design principles; however, the case study design example was a component design characterized by quantitative choice attributes. A more general choice modeling approach is required to accommodate a complex system, with both quantitative and qualitative choice attributes. The hierarchical choice modeling approach for complex system design presented in Section 2.3 utilizes multiple model levels to create a link between qualitative attributes considered by consumers when selecting a product and quantitative attributes used for engineering design. In this chapter, the approach is expanded to the Integrated Bayesian Hierarchical Choice Modeling (IBHCM) framework, estimated using an integrated multi-stage solution procedure. This approach utilizes choice data as well as other preference data, such as that collected using the Ford Programmable Vehicle Model presented in Chapters 4 and 5, to create a comprehensive choice model to support complex system design. This new framework addresses the shortcomings of the previous method while providing a highly flexible modeling framework to address the needs of complex system design. In this framework, both systematic and random consumer heterogeneity is explicitly considered, the ability to combine multiple sources of data for model estimation and updating is significantly expanded, and the integrated estimation method is introduced to mitigate error propagation throughout the model hierarchy. In addition to developing the new choice model approach, the importance of including a complete representation of consumer heterogeneity in the model framework is provided. The new modeling framework is validated using several metrics and techniques. The benefits of the IBHCM method are demonstrated in the design of an automobile occupant package.

This chapter is organized as follows: Section 6.1 introduces the hierarchical choice modeling approach, Section 6.2 presents the Integrated Bayesian Hierarchical Choice Model framework, Section 6.3 provides a vehicle design case study, and Section 6.4 provides validation of the IBHCM.

6.1 INTRODUCTION

A large-scale design problem is characterized by attribute hierarchies in demand model estimation, a hierarchy of consumer demographic descriptors (**S**), and data from multiple sources with varying degrees of richness (e.g., in-house marketing surveys, purchase data). Existing demand modeling approaches in the design literature require that product attributes considered in the choice model be quantitative. However, many criteria used by customers to choose between complex engineering systems tend to be qualitative, especially those at the system level. Also, as noted in Section 1.1, existing demand modeling approaches used in engineering design do not adequately account for consumer heterogeneity, nor do they adequately consider multiple data sources. To deal with the attribute hierarchy inherent in the design of a complex system (e.g., automotive design), the **Hierarchical Choice Modeling** framework has been developed as described in Section 2.3 (Kumar et al., 2009) and illustrated in Figure 6.1.

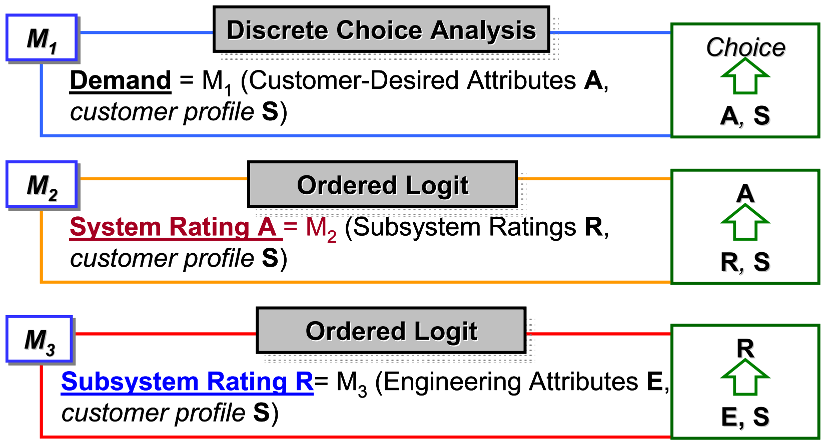


Figure 6.1: Hierarchical Choice Modeling Method

The proposed approach uses customer *ratings* **R** for qualitative customer-desired attributes (**A**) in the choice model, which are expressed in terms of quantitative engineering attributes through the hierarchy of *linking* models. Qualitative attributes in the top-level discrete choice analysis (DCA) model, labeled $\mathbf{M_1}$, are linked to engineering attributes through a series of ordered logit (OL) ratings prediction models for the subsystems, labeled $\mathbf{M_2}$ and $\mathbf{M_3}$, as illustrated in Figure 6.1. In this framework, the top level choice model only contains a reasonable set of system-level customer-desired attributes **A** (including price *P*) and demographic attributes, **S**. For example in vehicle design, **A** can include interior *roominess* and *exterior styling*, and **S** can include *income* and *age*. The lower level ratings models at the component and subsystem levels establish the relationships between qualitative customer perceptual attributes **A** as functions of quantitative engineering design attributes **E** and **S,** i.e., $\mathbf{A}=f(\mathbf{E}, \mathbf{S})$**.** For example, the lower level models can link **E**, such as *occupant package* and *styling* dimensions, and **S**, such as *stature* and *gender*, to the **A** defined at the top level. This structure ensures a more manageable model at each level, and mitigates the model estimation issues that accompany an all-in-one approach. The hierarchical choice model framework is used as the basis for an enterprise-driven engineering design decision making process to set targets for key system attributes, as illustrated in Figure 6.2.

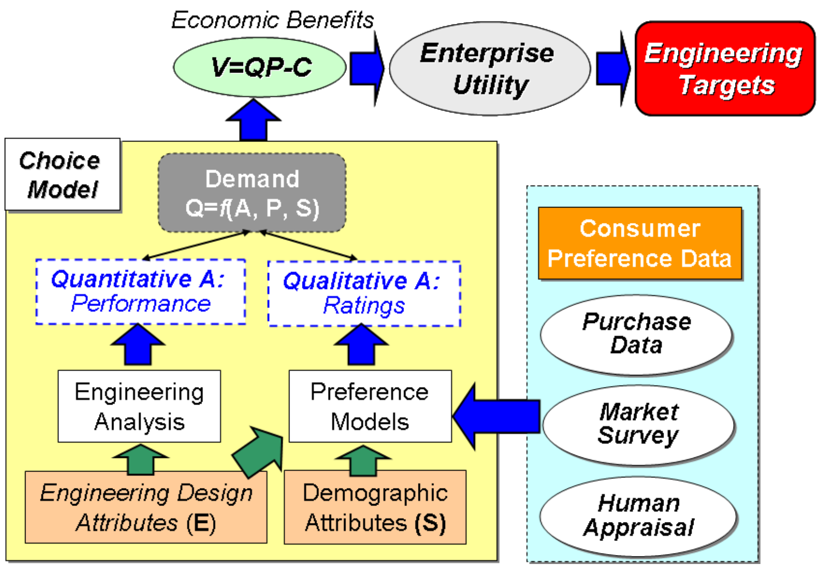


Figure 6.2: Overview of the Hierarchical Choice Modeling Method

In this framework, engineering and demographic attributes are mapped to the attributes in the choice model (M_1 level), which enables estimation of product demand Q. Creating an enterprise utility function as a function of Q, as well as Price, P, and Cost, C, allows target levels for the engineering attributes to be determined through maximization of enterprise utility.

While Section 2.3 laid out the general hierarchical modeling framework, several issues exist. A primary issue is the **lack of a mechanism to mitigate error propagated** in the hierarchy, since each of the models, i.e., M_1, M_2, and M_3, is estimated separately in the current implementation. This limitation is a significant issue because it inhibits the quantification of uncertainty at the top level choice model needed for decision making, and also provides no mechanism to ensure that the model accurately captures consumer preferences. Another issue identified in the hierarchical modeling approach is the **challenge of data collection** to enable model estimation over the entire model hierarchy. As noted, in the design of a complex system, it is unrealistic to expect that the necessary data for the complete model estimation process is present in a single data set. Our previous work introduced the Nested Logit (NL) method to combine multiple subsystem surveys to estimate the M_1 level choice model; however, this method does not fully address the problem because it is only valid at the choice level (i.e., M_1 level) and does not address how to combine data collected at different times. Another issue is that the current framework only captures **systematic heterogeneity** (illustrated in Figure 2.3), and lumps **random heterogeneity** into the overall error term. In this work, systematic heterogeneity is observed and described by an attribute of the customer, **S**, to explain his/her choice behavior. Random heterogeneity is not observed and is captured assuming the model parameters are random, as opposed to fixed, across respondents, as illustrated in Figure 6.3. Random heterogeneity accounts for the fact that two people with the same **S**, facing the same product attributes **A**, can make different choices. It was seen in the ANOVA analysis (Section 5.4) that random heterogeneity can be significant.

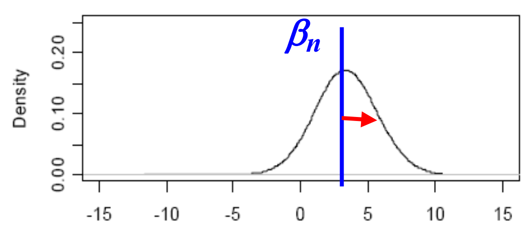


Figure 6.3: Example of Parameter Distribution Associated with Random Heterogeneity

Modeling the heterogeneity of customer preferences in a complete way is a challenge in choice modeling for complex engineering system design. Most existing approaches in the design literature do not consider heterogeneity of preference in modeling (i.e. systematic and random heterogeneity do not appear in the demand model). Li and Azarm (2000), and Michalek et al. (2005) used conjoint analysis, in which systematic heterogeneity was not considered; Michalek et al. (2005) considered random heterogeneity only in using a mixed logit DCA model. Cook (1997) used a linear model derived from Taylor Series expansion which used product value and price to estimate demand. Wassenaar et al. (2003; 2005) considered systematic heterogeneity only by including a limited number of demographic attributes (e.g., age, gender) in a DCA model.

To address the issues described, a unified **Integrated Bayesian Hierarchical Choice Model (IBHCM) framework** is developed to capture both systematic and random heterogeneity at all levels of the hierarchy, as well as to provide a method to estimate the predictive models from multiple data sources. An **integrated multi-stage** model solution methodology is introduced to mitigate error propagated through the model hierarchy and quantify uncertainty. Bayesian choice modeling has been applied primarily for estimating the mixed logit choice model to capture random heterogeneity (Train, 2003), and has been developed for this purpose in a variety of product marketing contexts, such as to model repeated purchase behavior (Rossi and Allenby, 2003; Rossi et al., 2005). Limited investigation of the use of Bayesian methods for combining multiple information sources has been conducted (Neelamegham and Chintagunta, 1999; Erdem and Keane, 1996), but not specifically in the choice modeling context. Specifically, the use of the Bayesian estimation method to estimate a complete hierarchy of random parameter or “mixed” models from a variety of data sources has not been presented in the literature. The automobile vehicle occupant packaging problem of Section 1.3 is used to demonstrate the methodologies developed in this research. The occupant packaging problem contains the proper level of complexity to demonstrate the features of the methodology.

6.2 INTEGRATED BAYESIAN HIERARCHICAL CHOICE MODELING APPROACH

Bayesian estimation methods offer many advantages over classical methods in estimating the hierarchical choice model. Bayesian estimation differs from classical methods in that the posterior distribution of the parameters is identified in the solution process, as opposed to point estimates specified model parameters (i.e. β, Σ). Bayesian estimation uses Gibbs sampling to sample from the posterior distribution. The Bayesian paradigm is also well aligned with the challenges of creating a design decision tool. Throughout the product design cycle and the product life, new information about demand may become available. This information may result from additional product surveys (SP) conducted or new actual purchase data (RP) acquired. With regard to actual purchase data, the growth of the internet, and the resulting growth in information, points to a future in which new information will be obtained at an almost continuous rate (Varian, 1995). Incorporating this increased knowledge must be considered in the product planning phase and throughout the product life to ensure products will be competitive and profitable throughout the lifecycle in which they compete. Such considerations point to the use of a Bayesian methodology for estimating the choice model.

6.2.1 Formulation of Choice and Ratings Models Incorporating Heterogeneity

The Bayesian Hierarchical Choice Model framework proposed in this work is a system of predictive models which captures consumer heterogeneity at all levels in the hierarchy, allows for estimation using multiple data sources, and provides a method for mitigating error propagated and quantifying uncertainty using integrated model estimation. The following models are used in the hierarchy of Figure 6.1:

- M_1 (*choice*): Mixed Logit (**MXL**)

- M_2, M_3 (*ratings*): Random-Effects Ordered Logit (**RE-OL**)

The MXL and RE-OL models are random parameter models which capture the effect of system design attributes, as well as both systematic and random heterogeneity, in modeling consumer choices or ratings. The MXL model (Train, 2003) is used to model consumer choice as a function of both system customer desired attributes **A** and consumer demographic descriptors **S**. The RE-OL models (Hedeker and Gibbons, 1994) express consumer ratings as a function of engineering attributes **E**, or sub-system ratings **R**, and **S** (Figure 6.1).

The MXL and RE-OL models assume the choice or rating is a discrete expression of an unobserved, latent consumer utility for a product or system design. The concept of choice (M_1) or rating utility (M_2, M_3) is derived assuming that the individual's, n, true choice utility, u, for a design alternative, i, consists of an observed part W, and an unobserved random disturbance ε (unobserved utility):

$$u_{in} = W_{in} + \varepsilon_{in} \ . \tag{6.1}$$

Observed utility, W_{in}, is parameterized in terms of model coefficients, $\boldsymbol{\beta}$, and **A**, **E**, and **S**. As noted in the previous section, **S** accounts for systematic taste heterogeneity. Random taste heterogeneity is accounted for using random model coefficients. This is achieved by allowing each individual person, n, to have his/her own set of model coefficients, β_n (Train, 2003; Train and McFadden, 2000; Rossi et al., 2005). In the MXL model, all customer-desired model parameters, $\boldsymbol{\beta}_{A,n}$, are random, while the $\boldsymbol{\beta}_S$ are fixed to avoid the identification problems caused by allowing a β to vary over alternatives i and people n (Train, 2003). The observed utility, W_{in}, in the MXL is given by:

$$W_{in}\left(\mathrm{MXL}\right)=\boldsymbol{\beta}'_{\mathrm{A},n}\mathbf{A}+\boldsymbol{\beta}'_{\mathrm{S}}\mathbf{S}+\boldsymbol{\beta}'_{\mathrm{A\cdot S}}\left(\mathbf{A}\cdot\mathbf{S}\right). \tag{6.2}$$

In the RE-OL model, the random intercept term, $\beta^0{}_n$, captures random consumer heterogeneity. The observed utility, W_{in}, in the RE-OL model is given by:

$$W_{in}\left(\mathrm{RE\text{-}OL}\right)=\beta_n^0+\boldsymbol{\beta}'_{\mathrm{E}}\mathbf{E}+\boldsymbol{\beta}'_{\mathrm{S}}\mathbf{S}+\boldsymbol{\beta}'_{\mathrm{E\cdot S}}\left(\mathbf{E}\cdot\mathbf{S}\right). \tag{6.3}$$

To prevent identification issues created by confounding of the random intercept and the error term, the random intercept must be estimated on *multiple* observations for each person n (i.e. panel data is required).

The MXL choice probability is expressed as (Train, 2003):

$$\mathrm{Pr}_n(i)=\int\left(\frac{e^{W_{in}(\beta)}}{\sum_{j=1}^{J}e^{W_{jn}(\beta)}}\right)pdf\left(\boldsymbol{\beta}_{\mathrm{A}}\right)d\boldsymbol{\beta}_{\mathrm{A}}. \tag{6.4}$$

The directly analogous random-effects ordered logit model to the choice model of Eq. (6.4) is formulated as:

$$\mathrm{Pr}_n\left(R_p\right)=\int\left(\frac{e^{k_p-W_{in}(\beta)}}{1+e^{k_p-W_{in}(\beta)}}-\frac{e^{k_{p-1}-W_{in}(\beta)}}{1+e^{k_{p-1}-W_{in}(\beta)}}\right)pdf\left(\beta^0\right)d\beta^0. \tag{6.5}$$

where R_p is a rating and k_p is an ordered logit cutpoint (Hedeker and Gibbons, 1994). Using Bayesian model estimation (as described in Section 2.2.2), the individual-level model coefficients ($\boldsymbol{\beta}_{\mathbf{A},n}, \beta_n^0$) are estimated, with uncertainty in the estimate decreasing as more choices or ratings per respondent are observed. In this work, random taste heterogeneity is considered as a form of uncertainty in preference behavior, since model predictions are generally not made for the sampled population used in the training data set, but rather the target population as a whole.

Therefore, the calculated distribution of β_n is of interest in this work, as opposed to individual-specific values of β_n.

6.2.2 Importance of Modeling Heterogeneity

The importance of accounting for heterogeneity, **S** and Σ, throughout the hierarchical choice modeling process results from the non-linear relationship between observed customer choice utility, W_{in}, and choice, $\Pr_n(i)$, (or rating, $\Pr_n(R_p)$) probability. Probabilistic choice modeling was developed within mathematical psychology (Luce, 1959) to capture the probabilistic nature of individual choice behavior, i.e. individuals do not always select the alternative with highest expected utility. The non-linear, S-shaped (i.e. logistic) relationship between W_{in} and $\Pr_n(i)$ implies that an equal change to W_{in} for a given design alternative for all individuals n, such as that resulting from a design change (e.g. a change in value of **A**), results in a *different* change in $\Pr_n(i)$ for each individual. This behavior can be interpreted as individuals with strong preferences (positive or negative) for a particular alternative are not as likely to modify their choice behavior when design changes are made, as are individuals with weaker preferences.

The role of demographic descriptors (i.e. systematic heterogeneity), **S**, is to capture individual-level attributes which influence utility, W_{in}, to enable a better estimate of individual-level choice probability, $\Pr_n(i)$. The effect of preference heterogeneity is demonstrated graphically in Figure 6.4, in which a MNL model without **S** (black line) and a MNL with **S** (gray lines) are estimated. For the training data set with a given set of demographics, both estimate a choice share for an alternative i of 0.5. If a change is made to the design (i.e. $\Delta\mathbf{E}$=0.75), this change results in an equal utility change for all consumers in the data set; however, the aggregate logit method overestimates the increased choice probability, and hence choice share (i.e. 0.2

increase), vs. the more accurate estimate from the disaggregate method (i.e. [0.1+0.05]/2 = 0.075 increase).Inclusion of **S** explicitly in the choice model also allows for choice predictions to be made for a new target market with a different demographic distribution than the survey market used for model estimation.

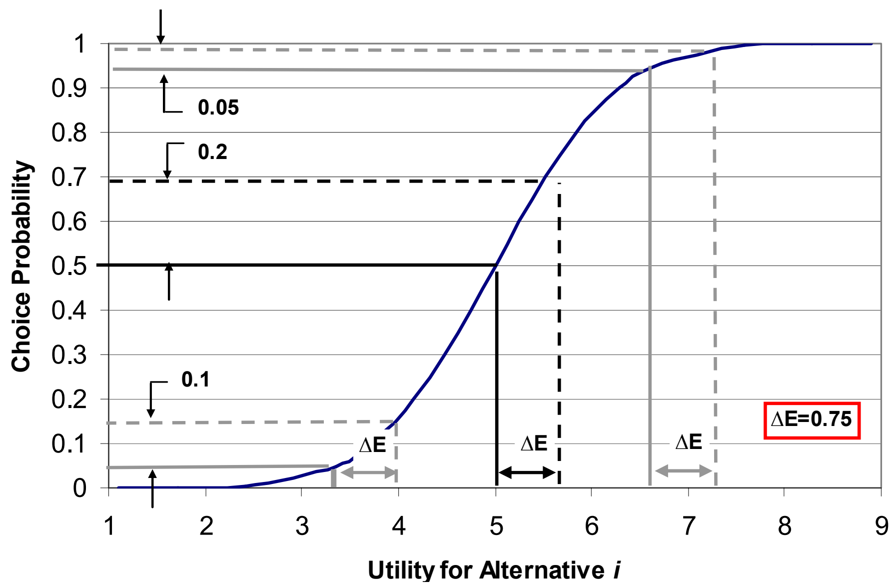


Figure 6.4: Effect of S upon Choice Probability

The effect of including random heterogeneity is that it provides both a more rigorous representation of preference heterogeneity as well as a relaxation of the Independence of Irrelevant Alternatives (I.I.A.) property of the discrete choice model (Train, 2003). Unlike a fixed parameter model, a random parameter model requires that the expected value of the choice probabilities be determined; therefore, each respondent utility function is an integration over β to find the expected value of $\Pr_n(i)$. This expected value of choice probability is different than the probability calculated using the mean value of β (i.e. the plug-in approach (Rossi et al., 2005)) due to the non-linear utility vs. probability curve described previously. The other benefit of considering random heterogeneity is that the mixed logit (MXL) choice model relaxes the I.I.A. property of the multinomial logit (MNL) choice model. As was shown in Brownstone and Train

(1998), the MXL model results in much different predictions of choice shares for a given design change than the MNL model due to the relaxation of I.I.A.: I.I.A. restricts the pair-wise choice probability ratios for each set of alternatives in the choice set to remain unchanged for a given design change. When using the MXL model to make choice estimates for a new market, the random heterogeneity of the new population is assumed to exhibit the same random heterogeneity as that in the model training data.

6.2.3 Model Fusion and Updating

In addition to capturing heterogeneity, use of the IBHCM allows the models within the hierarchy to be estimated from several data sources. There are two methods to combine these multiple data sources: *fusion* is used when no single survey contains the complete information necessary to estimate all the desired model parameters, and *updating* is used when new information becomes available to update all model parameters. As illustrated in Figure 6.5, fusion is associated with creating a model at a single time period from multiple data sources, whereas updating is associated with updating a model as new data becomes available.

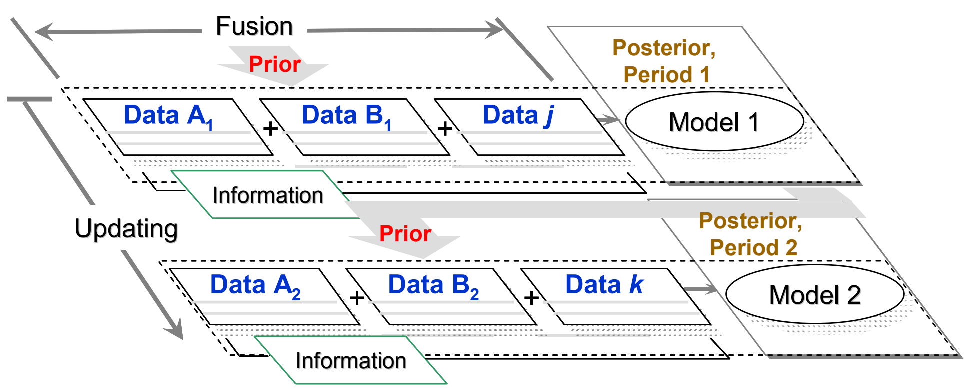


Figure 6.5: Data in the Hierarchical Choice Model Approach

This framework is based upon the Bayesian tradition of estimation, in which a *prior* distribution is assumed and the new data form the *information*, allowing estimation of a *posterior* distribution. The prior may be a non-informative prior and the information may be several fused data sets. The methodologies for model fusion and updating are described in this subsection.

Model Fusion: Whenever data from multiple surveys are fused (e.g., data sets labeled *1* and *2*) to create a single model with pooled utility function W_{pooled}, the error variances in each of the datasets may in fact be different, thus violating the assumption of an independently and identically distributed (I.I.D.) error term in the resulting model. Differences in error variance affect the scale of the model parameters because only differences in utility matter in the utility function, and thus the scale of the utility function is set based on the variance associated with a given dataset. The scale is set by scaling the overall model error variance, $\mathrm{var}(\varepsilon_{in})$, to a given value $\mathrm{var}_S(\varepsilon_{ni}) = \sigma_k$ (e.g. $\sigma_k = \pi^2/6$), by dividing W_{in}, and hence model parameters β, by a scale factor μ_k (e.g. $\mu_k = \mathrm{var}(\varepsilon_{in})/(\pi^2/6)$) to achieve σ_k. Thus the β coefficients in a choice or ratings model are confounded with the scale factor (i.e. $\beta = \beta^*/\mu_k$) and cannot be separately identified. This presents an issue when estimating a choice model with multiple data sets in that the error variance, and thus μ_k, will differ for each data set. Thus a method is needed to ensure the scale factors from all data sets are equal (i.e. $\mu_1 = \mu_2 = \mu_k$) to ensure β coefficients estimated from different data sets are on the same scale in the pooled utility function, W_{pooled}.

The Nested Logit (NL) methodology has been adapted previously to combine multiple data sets with different error variances as described in Kumar et al. (2009). A method to combine data from multiple sources using the MXL or RE-OL model is formulated in this section. To enable use of multiple data sets in the MXL methodology formulated in Eq. (6.4) or the RE-OL method of Eq. (6.5), a random term, η_k, of mean 0 and variance τ_k (i.e. $\eta_k \sim N(0, \tau_k)$) is assigned to each dataset-specific u_{in} to account for the different error variances associated with each dataset to enable a common model error term, ε_{in}. This is expressed as follows (Brownstone et al., 2000):

$$u_{1,in} = W_{1,in} + \eta_1 + \varepsilon_{in} \quad i \in 1, \quad u_{2,in} = W_{2,in} + \eta_2 + \varepsilon_{in} \quad i \in 2\,, \tag{6.6}$$

where η_1 and η_2 are the survey-specific error component terms which are estimated together with the other model parameters (note: the η_k associated with the data set with the lowest variance is set to 0 for model identification). These additional error terms, $\eta_1 \ldots \eta_k$, account for the differences in error variances in different data sets, and ensure that the overall common model error, ε_{in}, is I.I.D. (i.e. var($\varepsilon_{1,in}$) = var($\varepsilon_{2,in}$)= var(ε_{in})). This equivalence is achieved by allowing the additional error term, η_k, to contain the additional variance greater than the base variance, var(ε_{in}); for example:

$$\mathrm{var}(\varepsilon_{2,in}) = \mathrm{var}(\eta_2 + \varepsilon_{in}) = \mathrm{var}(\eta_2) + \mathrm{var}(\varepsilon_{in}), \quad s.t.\ \mathrm{var}(\varepsilon_{2,in}) > \mathrm{var}(\varepsilon_{in}). \tag{6.7}$$

This approach therefore relies on the ability to separately estimate var(η_1),…, var(η_k) such that the error variance associated with each data set is var(ε_{in}). To estimate the η_k, it is necessary that each survey, and thus each observed utility function W_1 and W_2, shares some common attributes, given by $\mathbf{A}_{com}$ or $\mathbf{S}_{com}$ to determine the survey-specific error component terms, base on the fact that model parameters for shared attributes indicate differences in model scale. Estimation of $\eta_1 \ldots \eta_k$ is enabled by the condition that the β coefficients for shared attributes, $\mathbf{A}_{com}$ or $\mathbf{S}_{com}$, are equivalent and that $\eta_1 \ldots \eta_k$ are positive.

This approach is unlike the previous Nested Logit approach (Kumar et al., 2009) in which the individual utility functions, W_1 and W_2, must be scaled by a scaling factor, μ, to create W_{pooled}. In the MXL and RE-OL approach, the pooled observed utility is simply the sum of W_1 and W_2 with a single set of common parameters, indicated by the subscript *com*:

$$W_{pooled} = \left(\boldsymbol{\beta}'_{\mathbf{A},1}\mathbf{A}_1 + \boldsymbol{\beta}'_{\mathbf{S},1}\mathbf{S}_1\right) + \left(\boldsymbol{\beta}'_{\mathbf{A},2}\mathbf{A}_2 + \boldsymbol{\beta}'_{\mathbf{S},2}\mathbf{S}_2\right) + \boldsymbol{\beta}'_{\mathbf{A},com}\mathbf{A}_{com} + \boldsymbol{\beta}'_{\mathbf{S},com}\mathbf{S}_{com}. \tag{6.8}$$

Model Updating: The Bayesian framework provides a convenient means for updating the hierarchical model as new data becomes available, for example as new model year surveys are conducted for an automobile. In this case, the current model parameters are the *prior* distribution and the new survey is the *information* necessary to calculate the new *posterior* distribution of the model parameters. Updating can be useful when the initial model is estimated using survey data acquired from prototype hardware. In Bayesian estimation, model parameters (including ordered logit cut points **k**) are updated according to the ratio of variances in the prior distribution versus that in the information (i.e. vehicle) data set (Johnson and Albert, 1999). Therefore, the prototype based model can be updated with the limited vehicle survey to both update the model parameters, $\boldsymbol{\beta}$, to account for the influence of actual vehicle preferences, and to update the cut points, **k**, to ensure that OL model rating predictions are reflective of actual vehicle ratings.

6.2.4 Integrated Choice Model Formulation

With mixed formulations for the M_1 level choice model (Eq. (6.4)) and the M_2 and M_3 level ratings models (Eq. (6.5)), an integrated formulation of the hierarchical choice model is derived for estimation of the model. The general integrated model framework is shown in Figure 6.6. The theoretical advantages of such a framework are as follows:

1. Mitigate error and quantify uncertainty by propagating the distribution of $\boldsymbol{\beta}$ throughout the model hierarchy.
2. Track the respondent effect for a single person throughout the model hierarchy.

3. Quantify uncertainty to create the enterprise level utility function for decision making.

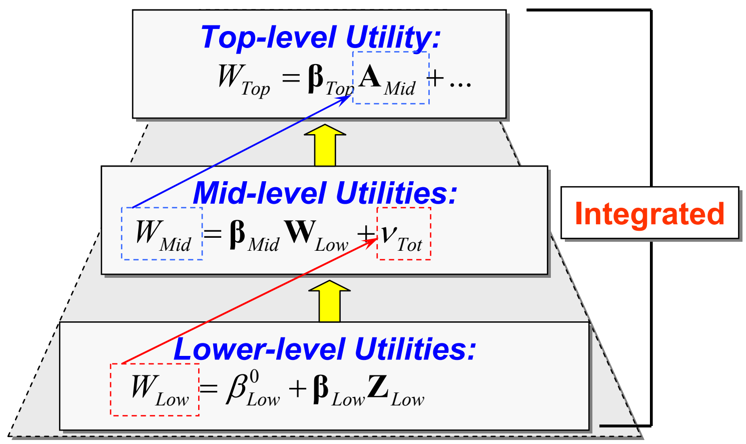


Figure 6.6. Integrated Choice Model Estimation

The framework uses a propagation of cumulative respond-level utility (*W*), including cumulative respondent-level error ν, through a complete hierarchy of models from the bottom level (*Low*) to the top level (*Top*). The complete hierarchy of models uses integrated multi-stage estimation to fit model parameters (i.e. $\boldsymbol{\beta}$) in upper-level models to model predictions from lower level models to minimize prediction error when using the system of models to make choice share predictions.

The method for propagating utility, including respondent-level error through the model hierarchy is formulated. To quantify error, the error distribution at each level of the model hierarchy must be accounted for in the final choice prediction. This problem has been solved for linear regression modeling, using instrumental variable techniques. Specifically, two-stage least squares regression (Greene, 2002) has been used to account for error propagation in a 2-level linear regression model system. The approach has been generalized by Lancaster (2004) using Bayesian solution methods for linear regression systems with more than two models. In the approach of Lancaster, error from the lower model level, denoted as M_{Low}, is propagated to the top level model, M_{Top}, such that the total error due to both models, $\varepsilon_{Tot} = f(\varepsilon_{Low}, \varepsilon_{Top})$, is estimated. This approach is demonstrated assuming a 2-level model hierarchy:

$$\begin{aligned} x &= \gamma + \delta z + \varepsilon_1 \\ y &= \xi + \beta x + \varepsilon_2 \end{aligned} . \tag{6.9}$$

The lower-level equation for *x* is substituted into the upper-level equation for *y*, and *y* is rewritten as follows:

$$\begin{gathered} y = \xi^* + \beta\delta z + \nu_2 \\ \xi^* = \xi + \beta\gamma,\ \nu_2 = \varepsilon_2 + \beta\varepsilon_1 \end{gathered} . \tag{6.10}$$

In this new form, the posterior distribution of ξ^* and ν_2 is found, thus identifying the cumulative intercept and error term directly in the upper-level model. This method can be extended to any *j*-level model hierarchy.

The issue with applying this approach to the hierarchy of *choice* and *ratings* models is that the error terms are not directly estimable in such models: the error variance is confounded with the β terms. This confounding occurs because only *differences* in utility matter in the choice/ratings model, and thus β and Var(ε) in each model cannot be separately identified (Train, 2003; Lancaster, 2004). Thus the method of finding the posterior distribution of ε_{Tot} developed for least squares regression cannot be applied directly for the IBHCM problem.

To account for the error propagated in the hierarchical choice model, an error components interpretation of the random term is applied (Train, 2003). In the random-effects ordered logit model, the β^0 term is added to the utility expression to capture the random respondent effect (Eq. (6.3)). The β^0 term is the random intercept and can be interpreted as the portion of the overall model error which is attributed to individual respondents (Hedeker and Gibbons, 1994). Thus, the part of the error attributed to respondent-level variation is observable, and we use a formulation analogous to that presented in Lancaster (2004) for least-squares regression, with modifications as required by the model form. The respondent-level error is propagated through all levels of ratings models (M_2 and M_3 in the current discussion) and the total observed error for each system rating quantified at the M_1 level choice model.

As shown in Eq. (6.5), the expected rating predicted by the ordered logit model is a function of the utility, W_{in}, with expected ratings predicted by the M_3 and M_2 level models expressed as a function of engineering and demographic attributes, $\mathbf{Z}$ ($\mathbf{Z} = \{\mathbf{E}, \mathbf{S}\}$):

$$\begin{aligned} R_{M3} &= f\left(W_{M3}\right) = f\left(\beta^0_{M3} + \boldsymbol{\beta}'_{M3}\mathbf{Z}_{M3}\right) \\ R_{M2} &= f\left(W_{M2}\right) = f\left(\beta^0_{M2} + \boldsymbol{\beta}'_{M2}\mathbf{R}_{M3}\right) \end{aligned}. \tag{6.11}$$

As seen in Eq. (6.11), the upper-level equation for R_{M2} is a function of expected ratings predicted by the lower level model, $\mathbf{R}_{M3}$. In order to enable the error term to be propagated through the model hierarchy, utility, W, is propagated through the model hierarchy instead of the expected rating, R:

$$\begin{aligned} W_{M3} &= \beta^0_{M3} + \boldsymbol{\beta}'_{M3}\mathbf{Z}_{M3} \\ R_{M2} &= f\left(W_{M2}\right) = f\left(\beta^0_{M2} + \boldsymbol{\beta}'_{M2}\mathbf{W}_{M3}\right) \end{aligned}. \tag{6.12}$$

Using utility instead of expected ratings allows an approach analogous to two-stage least squares regression to be used to estimate a total error term, given by ν_{M2}:

$$W_{M2} = \beta^0_{M2} + \boldsymbol{\beta}'_{M2}\mathbf{W}_{M3} = \boldsymbol{\beta}'_{M2}\left(\boldsymbol{\beta}'_{M3}\mathbf{Z}_{M3}\right) + \overbrace{\beta^0_{M2} + \boldsymbol{\beta}'_{M2}\left(\beta^0_{M3}\right)}^{\nu_{M2}} \tag{6.13}$$

The posterior distribution of ν_{M2} thus captures the cumulative respondent-level error (β^0) from all preceding levels in the model hierarchy. The posterior distribution of ν_{M2} is sampled directly in the solution process, therefore simplifying model estimation. With this formulation, the predicted rating for each subsystem for each vehicle alternative, *i*, for each person, *n*, is estimated for customer-desired attributes, **A**, appearing in the M_1-level choice model:

$$\mathbf{A} = f\left(\boldsymbol{\beta}'_{M2}\boldsymbol{\beta}'_{M3}\mathbf{Z}_{M3} + \mathbf{v}_{M2}\right). \tag{6.14}$$

With expressions for each level in the model hierarchy and a method for integrated multi-stage model parameter estimation given in Eq. (6.13), the integrated choice model framework is formulated. Because the models are estimated simultaneously, β parameters from one model, such as the M_3 headroom model, can be correlated with β parameters in another model, such as the M_2 roominess model or the M_1 choice model. The set of models is solved simultaneously using Markov Chain Monte Carlo (MCMC) sampling methods (Rossi et al., 2005).

When using the estimated model to calculate choice probabilities, the logit probabilities for each person *n* and each alternative *i* must be integrated over the full distribution of $\boldsymbol{\beta}$ or $\boldsymbol{\nu}$ in each model of the hierarchy (i.e. M_1, M_2, M_3):

$$\Pr_n(i) = \int L_{ni}\left(\boldsymbol{\beta}_{\mathrm{M1}}, \mathbf{v}_{\mathrm{M2}}\right) g\left(\boldsymbol{\beta}, \mathbf{v} \mid \mathbf{b}, \boldsymbol{\Sigma}\right) d\boldsymbol{\beta} d\mathbf{v}. \tag{6.15}$$

Integration over the distribution of model parameters is used instead of the individual-level β_n, which are recoverable from the Bayesian solution, because the model training data is a population sample and the *distribution* of β_n is assumed to be representative of the population as a whole. The random parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$, are defined by a mean vector, **b** (*b*=0 for $\boldsymbol{\nu}$ parameters), and a full variance-covariance matrix, $\boldsymbol{\Sigma}$, of the parameters estimated in the IBHCM solution process ($\boldsymbol{\beta}, \boldsymbol{\nu} \sim N(\mathbf{b}, \boldsymbol{\Sigma})$).

6.3 Case Study: Vehicle Occupant Package Design

6.3.1 Integrated Bayesian Hierarchical Choice Model Estimation

Vehicle occupant package design is used as an example to demonstrate the use of the IBHCM in system design. The focus is to present the features and benefits of the hierarchical choice modeling approach in an illustrative manner, rather than completing a comprehensive design optimization of the entire vehicle package. The scope of the case study is restricted to the driver's occupant package. The IBHCM framework and vehicle dimensions for this problem are shown in Figure 6.7. The vehicle dimensions considered are the eight dimensions used in the Programmable Vehicle Model (PVM) human appraisal experiment described in Section 5.2.

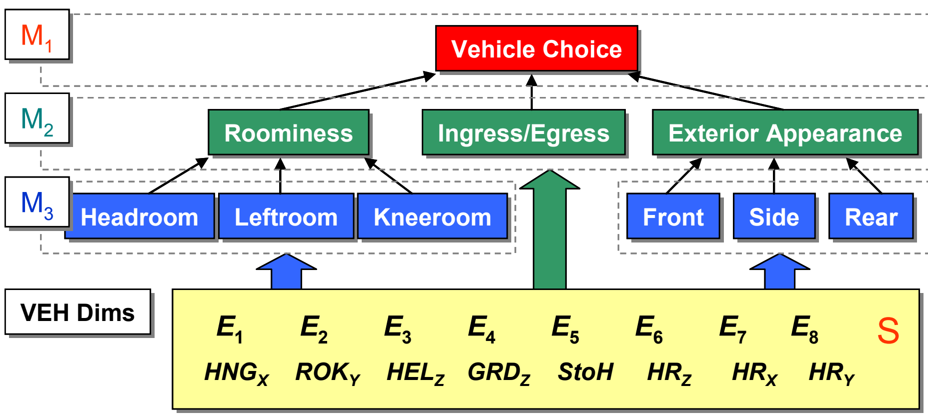


Figure 6.7. Case Study Hierarchical Model Structure

<u>Data Available for Model Estimation</u>: Three data sets are available for model estimation: two clinical studies—an interior packaging-based survey (***DS***$_1$) and an exterior styling-based survey (***DS***$_2$)—and a combined roominess, ingress, and egress (***DS***$_3$) study performed on the Ford Programmable Vehicle Model (PVM). The interior and exterior clinical surveys were conducted on four vehicles in the full-size luxury segment. In the interior package survey, 73 respondents are asked to rate package attributes at both sub-system (e.g., overall roominess, ingress/egress) and component levels (e.g., head room, knee room) for four vehicles. In the exterior survey, the

same 73 respondents are asked to rate exterior appearance attributes and choose between the same set of vehicles. In addition to the packaging attributes, demographic attributes $\mathbf{S}_1$ and $\mathbf{S}_2$ (age, income, stature, gender) are recorded. Five vehicle attributes are included in the M_1 choice model: (1) roominess (*Room*), (2) ingress (*Ing*), (3) quality of materials (*Q Mat.*), (4) exterior styling (*Ext*), and (5) willing-to-pay (*W to P*). The attribute values are in the form of ratings on a 1-10 scale for each attribute for each respondent. The following demographic descriptors (**S**) are used in the M_1 model: (1) gender, (2) stature (i.e. height), (3) income, (4) age.

The Ford PVM, described in Section 4.5, is a computer controlled vehicle model which can simulate a large number of vehicle configurations, and hence is efficient for gathering preference data. A comprehensive combined roominess, ingress, and egress human appraisal was conducted using the PVM, consisting of 30 respondents each rating 18 vehicle configurations. The appraisal was designed using the optimal experiment design method for human appraisals of Chapter 4. In the designed experiment**,** the **8** engineering attributes of Figure 6.7 and **3** demographic attributes were varied as described in Section 5.2. The PVM data is used to estimate M_2 and M_3 level roominess, ingress, and egress RE-OL models in the hierarchy.

Models to be Estimated: The hierarchical model used to model consumer choices and preferences for the occupant package **roominess** and **ingress/egress** used in the case study is the random parameter IBHCM described in Section 6.2. This model links the preferences for roominess and ingress/egress at the choice level with the vehicle variables which determine the roominess and ingress/egress design (i.e. E_1–E_8). This model estimation is presented later in this section. RE-OL M_2 and M_3 level models linking the vehicle variables to the preferences for **exterior styling** are also created. For these models, a height-to-width ratio variable, (GRD_Z +

HR_Z)/ ROK_Y, called H/W, and a height-to-length variable, ($GRD_Z + HR_Z$)/ HNG_X, called H/L are created to capture the finding that respondents view styling in terms of the ratio of dimensions, rather than the absolute dimensions. Also, HEL_Z, GRD_Z, and StoH are expressed in terms of H130, the Step Height to the driver's door, as $GRD_Z - HEL_Z +$ StoH. The exterior styling M_2 and M_3 models, estimated using an integrated multi-stage approach, are shown in Table 6.1.

Table 6.1: Exterior M_2 and M_3 Models

M_3: Front Appearance

	coef.	*t-value*
H/W	**-0.799**	*-10.15*
Gend	**-0.304**	*-3.65*
age	**0.318**	*1.95*

M_3: Side Appearance

H130	**-0.219**	*-2.19*
H/L	**-0.732**	*-8.07*
Gend	**-0.262**	*-3.08*
age	**0.281**	*1.75*

M_3: Rear Appearance

H/W	**-0.840**	*-10.52*
Gend	**-0.318**	*-3.69*
age	**0.644**	*3.61*

M_2: Exterior Appearance

front	**0.45**	*19.59*
side	**0.66**	*8.76*
rear	**0.35**	*13.35*

Variance-Covariance Matrix of Random Effects

Front (M_3)	**0.99**			
Side (M_3)	**0.90**	**0.98**		
Rear (M_3)	**0.86**	**0.93**	**1.07**	
Exterior (M_2)	**0.69**	**0.76**	**0.72**	**0.83**

<u>RE-OL Model Updating</u>: An issue to address is that the M_2 and M_3 model estimation data for the occupant package is collected using the PVM, while the resulting model estimated is to be used for predicting ratings for actual vehicle designs. The PVM lacks vehicle-specific features and styling. This is important because PVM-based ratings models will not include the influence of

customer *perceptions* created by unique styling and layout features, in addition to the influence of the purely dimensional PVM features. Using actual vehicles to gather survey data for model estimation is challenging because it is difficult to achieve the necessary factor randomization and variation to achieve efficient model parameter estimates as described in Chapter 4. However, the IBHCM framework can be utilized for updating the PVM-estimated ordered logit model with the limited actual vehicle survey data (i.e. $\boldsymbol{DS_1}$) to include the influence of vehicle styling features in the model and ensure the predicted ratings are relative to the actual vehicles.

IBHCM Estimation: The model hierarchy and the use of the data sets for the roominess/ingress/egress model estimation are shown in Figure 6.8.

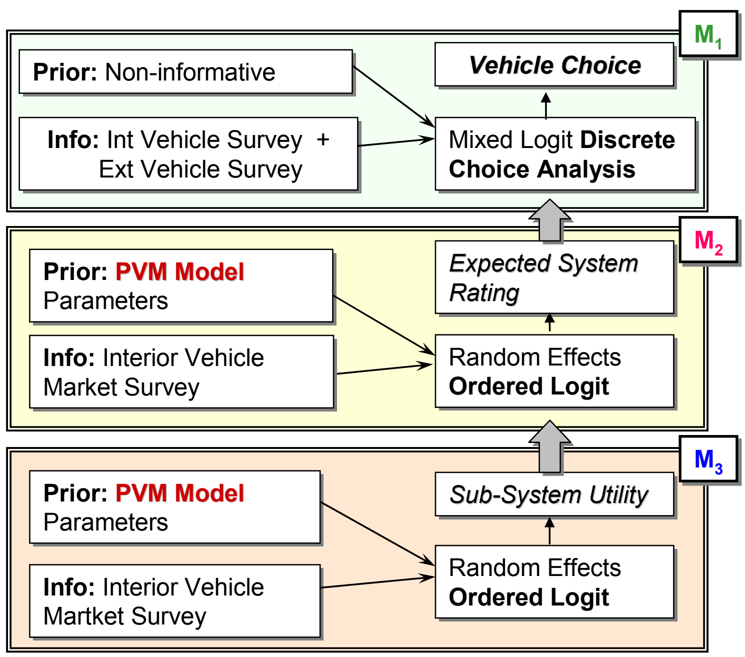


Figure 6.8. Integrated Bayesian Hierarchical Choice Model

In order to illustrate the benefits of the IBHCM approach, four versions of the hierarchical choice model are estimated. The alternative model versions use fixed (i.e. no random heterogeneity) versus random coefficients, and separate versus integrated model estimation in these combinations:

- *Scenario 1*: Fixed parameters, each model individually estimated (**SEP**).
- *Scenario 2*: Fixed parameters, integrated approach (**INT**).
- *Scenario 3*: Random parameters, each model individually estimated (**SEP**).
- *Scenario 4*: Random parameters, integrated approach (**INT**) (as in Section 6.2.4).

In Scenarios 1 and 3, each of the three models (M_1, M_2, and M_3) is estimated independently, whereas Scenarios 2 and 4 utilize the integrated estimation; however, only Scenario 4 utilizes the full error propagation method described in Section 6.2.4. Additionally, **Vehicle 4** is the vehicle under design, and thus is the only vehicle in the choice set linked to the M_2 and M_3 level models. The IBHCM was estimated using *WinBUGS* (Spiegelhalter et al., 2003), interfaced with *R-Project* (Ihaka and Gentleman, 1996) for data pre- and post-processing.

The results (i.e., model β coefficients) of the 4 model scenarios are shown in Table 6.2 (note: the ordered logit cut points are not shown for simplicity, see Appendix H for full model results). The variance and variance-covariance matrices for the two random effects models are shown in Table 6.3. For the Random SEP model, covariance between parameters in different models cannot be estimated since each model is estimated separately. In the Random INT model, covariance can be estimated between parameters in different models. All variance-covariance values in the M_2 and M_3 models are significant at greater than the 99% confidence level. While significant covariance was found among the parameters in the M_3 and M_2 level models, significant covariance was not found between the ratings models at the M_3 and M_2 levels and the parameters in the M_1 level model. Studies of the M_1 level choice model indicated that the covariance between M_1 model parameters is statistically insignificant, and thus a diagonal variance matrix can be specified to aid in convergence.

Table 6.2: Results of Four Model Scenarios

	Fixed SEP		Fixed INT		Random SEP		Random INT	
	coef.	*t-value*	coef.	*t-value*	coef.	*t-value*	coef.	*t-value*
M_3: Headroom								
ROK_Y	**1.53**	*8.26*	**1.50**	*8.10*	**1.55**	*8.53*	**1.60**	*6.74*
GRD_Z	**-0.19**	*-0.97*	**-0.18**	*-0.89*	**0.11**	*0.54*	**0.09**	*0.46*
HR_Z	**4.02**	*18.80*	**3.90**	*16.81*	**4.64**	*19.91*	**4.67**	*16.05*
HR_Y	**1.36**	*9.01*	**1.29**	*8.50*	**1.67**	*10.43*	**1.67**	*10.97*
Stat	**-2.16**	*-6.84*	**-2.20**	*-7.23*	**-3.27**	*-8.00*	**-3.13**	*-7.92*
BMI	**1.46**	*4.84*	**1.53**	*5.23*	**1.76**	*4.56*	**1.62**	*4.43*
Age	**-0.84**	*-3.47*	**-0.87**	*-2.71*	**-1.18**	*-3.72*	**-1.07**	*-3.67*
M_3: Leftroom								
ROK_Y	**3.48**	*16.83*	**3.45**	*18.27*	**3.54**	*17.16*	**3.51**	*18.16*
GRD_Z	**-1.53**	*-8.24*	**-1.64**	*-9.77*	**-1.44**	*-7.67*	**-1.37**	*-8.00*
StoH	**0.28**	*1.63*	**0.25**	*1.59*	**0.27**	*1.55*	**0.29**	*1.51*
HR_Z	**1.72**	*10.30*	**1.63**	*9.65*	**1.71**	*10.59*	**1.69**	*9.87*
Stat	**-2.17**	*-6.88*	**-2.13**	*-6.15*	**-3.17**	*-8.07*	**-3.07**	*-8.62*
BMI	**-1.21**	*-4.33*	**-1.08**	*-3.90*	**-1.87**	*-5.29*	**-1.78**	*-5.07*
Age	**-0.36**	*-1.44*	**-0.32**	*-1.22*	**-0.63**	*-2.03*	**-0.30**	*-0.92*
M_3: Kneeroom								
ROK_Y	**1.38**	*9.62*	**1.44**	*8.88*	**1.36**	*8.98*	**1.36**	*8.55*
HEL_Z	**1.01**	*6.70*	**1.03**	*6.62*	**1.05**	*6.93*	**1.04**	*6.80*
StoH	**-0.36**	*-2.22*	**-0.34**	*-2.32*	**-0.34**	*-2.06*	**-0.38**	*-2.24*
HR_Z	**0.81**	*5.04*	**0.86**	*5.51*	**0.84**	*5.06*	**0.88**	*4.82*
Stat	**-1.40**	*-4.52*	**-1.38**	*-4.41*	**-1.02**	*-2.37*	**-1.67**	*-4.25*
BMI	**-1.01**	*-3.55*	**-1.06**	*-4.19*	**-1.42**	*-4.13*	**-1.42**	*-4.25*
Age	**1.02**	*4.13*	**0.92**	*3.56*	**1.70**	*5.15*	**1.08**	*3.50*
M_2: Roominess								
Head	**0.38**	*6.93*	**0.27**	*2.90*	**0.52**	*9.53*	**0.21**	*2.02*
Left	**0.48**	*6.50*	**0.39**	*2.50*	**0.66**	*7.63*	**0.42**	*1.32*
Knee	**0.59**	*8.00*	**-0.18**	*-1.15*	**0.58**	*7.47*	**-0.53**	*-1.23*
M_2: Ingress/Egress								
ROK_Y	**0.83**	*5.39*	**0.86**	*5.55*	**0.80**	*5.26*	**0.81**	*5.77*
HEL_Z	**1.86**	*10.31*	**1.93**	*12.81*	**1.87**	*10.56*	**1.85**	*10.91*
GRD_Z	**-2.03**	*-10.83*	**-2.09**	*-11.53*	**-2.26**	*-11.57*	**-2.23**	*-12.67*
StoH	**-2.79**	*-14.97*	**-2.79**	*-15.32*	**-2.83**	*-15.30*	**-2.84**	*-16.53*
HR_Z	**1.87**	*10.23*	**1.81**	*11.46*	**1.86**	*9.95*	**1.88**	*10.90*
Stat	**-2.28**	*-6.82*	**-2.03**	*-5.98*	**-2.92**	*-7.41*	**-2.87**	*-8.06*
Age	**-1.04**	*-3.53*	**-1.01**	*-3.88*	**-1.75**	*-4.83*	**-1.66**	*-6.68*
Gend	**-0.09**	*-0.60*	**-0.01**	*-0.04*	**-0.31**	*-1.11*	**0.02**	*0.10*

M_1: Choice Model

Roominess	**0.53**	*4.11*	**0.90**	*1.30*	**1.17**	*4.45*	**1.90**	*2.30*
Ingress/Egress	**0.43**	*3.75*	**0.44**	*0.95*	**1.25**	*2.99*	**1.10**	*2.74*
Quality Materials	**0.41**	*3.90*	**0.51**	*4.82*	**1.45**	*4.05*	**1.30**	*3.14*
Ext. Appearance	**1.48**	*5.22*	**1.52**	*4.87*	**2.68**	*3.70*	**2.69**	*5.03*
Willing to Pay	**0.39**	*2.06*	**0.34**	*1.65*	**0.87**	*1.83*	**0.89**	*1.49*
Gend×Alt2	**0.03**	*0.06*	**-0.10**	*-0.18*	**0.35**	*0.37*	**0.09**	*0.12*
Gend×Alt3	**-0.03**	*-0.05*	**-0.01**	*-0.02*	**0.11**	*0.12*	**0.08**	*0.11*
Gend×Alt4	**-0.24**	*-0.43*	**-0.31**	*-0.62*	**-0.37**	*-0.39*	**-0.35**	*-0.48*
Stat×Alt2	**-0.54**	*-0.54*	**-1.57**	*-1.52*	**-2.02**	*-1.08*	**-3.09**	*-1.96*
Stat×Alt3	**-0.89**	*-0.79*	**-1.14**	*-0.99*	**-2.34**	*-1.31*	**-3.11**	*-2.17*
Stat×Alt4	**-0.18**	*-0.16*	**0.00**	*0.00*	**-1.77**	*-1.00*	**-1.16**	*-0.77*
Inc×Alt2	**0.87**	*0.77*	**-0.01**	*-0.01*	**1.59**	*0.77*	**-0.06**	*-0.04*
Inc×Alt3	**0.23**	*0.20*	**0.00**	*0.00*	**0.84**	*0.43*	**0.45**	*0.31*
Inc×Alt4	**-0.17**	*-0.13*	**-1.18**	*-1.18*	**1.51**	*0.67*	**-1.16**	*-0.75*
Age×Alt2	**0.93**	*0.94*	**2.16**	*2.06*	**2.34**	*1.44*	**3.76**	*1.97*
Age×Alt3	**0.71**	*0.71*	**1.22**	*1.10*	**1.72**	*1.10*	**3.16**	*1.70*
Age×Alt4	**1.29**	*1.24*	**1.51**	*1.75*	**2.33**	*1.49*	**3.00**	*1.57*
$\sigma_{packaging}$	**1.31**	*2.44*	**0.98**	*3.28*	**1.27**	*2.26*	**0.95**	*3.67*

Table 6.3: Variance-Covariance Matrix for Random Effects Models

M_2 & M_3 Variance (-Covariance) Matrices

Random SEP	
Headroom (M_3)	**1.94**
Leftroom (M_3)	**2.31**
Kneeroom (M_3)	**2.16**
Roominess (M_2)	**1.99**
Ingress/Egress (M_2)	**1.15**

Random INT					
Headroom (M_3)	**3.96**				
Leftroom (M_3)	**3.54**	**5.40**			
Kneeroom (M_3)	**3.26**	**4.35**	**4.64**		
Roominess (M_2)	**3.32**	**3.86**	**3.71**	**4.05**	
Ingress/Egress (M_2)	**3.08**	**3.52**	**3.22**	**3.35**	**3.94**

M_1 Variance (-Covariance) Matrices

Random SEP	
Roominess	**0.93**
Ingress/Egress	**2.00**
Quality Materials	**1.54**
Ext. Appearance	**1.00**
Willing to Pay	**1.61**

Random INT					
Roominess	**1.29**				
Ingress/Egress	**-0.10**	**1.19**			
Quality Materials	**-0.08**	**-0.21**	**1.38**		
Ext. Appearance	**—**	**—**	**—**	**0.98**	
Willing to Pay	**—**	**—**	**—**	**0.09**	**1.81**

The actual distributions of the random parameters for the Random INT model (i.e., random effect for the M_2 and M_3 models and random betas for the M_1 model) are shown in Appendix I.

6.3.2 Vehicle Occupant Package Design Optimization

A vehicle package optimization formulation is used to demonstrate the benefits of the estimated IBHCM in setting package design targets. The package optimization problem is to select the HNG_X, ROK_Y, HEL_Z, GRD_Z, StoH, and HR_Z dimensions to maximize choice share for Vehicle 4, of the four vehicles in the clinical surveys, while meeting vehicle-level requirements (i.e., fuel economy and weight). The six dimensions to be optimized are shown in Figure 6.9 (solid line oval), as well as the overall vehicle dimensions which will be used in the constraint functions (dashed line oval).The height, a function of GRD_Z and HR_Z, is limited by the overall vehicle height, given by H100, assumed to have a limit of 58 inches. Also, it is assumed that the weight of the vehicle is limited to 3500 lbs maximum for overall performance reasons, and that the fuel economy must be at least 24 mph to meet federal standards.

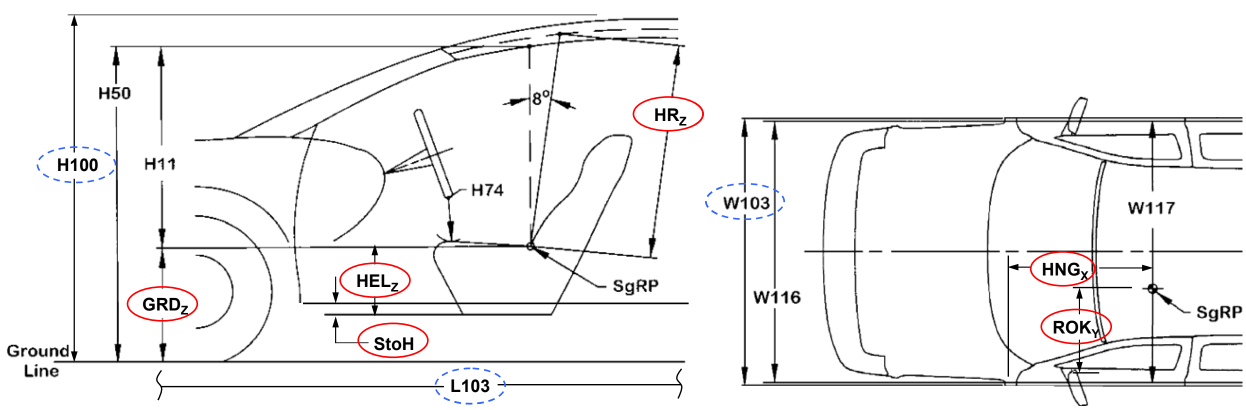


Figure 6.9: Relationship among Vehicle Packaging Dimensions (Society of Automotive Engineers, 2002)

The optimization problem is summarized in Table 6.4. The optimization is conducted using both the integrated random parameter model, and the integrated fixed parameter models to compare the difference in results between the two approaches. While the choice model captures consumers' trade-offs among interior and exterior attributes, it is necessary to mathematically express the relationships between packaging dimensions and other vehicle design performances (e.g., weight, fuel economy) to capture vehicle-level design trade-offs. Data to estimate linear regression models is collected for a total of 77 vehicles from the automotive website Edmunds.com (Edmunds Inc, 1995-2007), based on the segment of the vehicles in the occupant-based packaging survey $\boldsymbol{DS}_1$. From Kumar (2007), fuel economy is modeled as a function of weight, engine size, and width/length ratio, and the weight is expressed as a function of vehicle dimensions as summarized in Table 6.4.

Table 6.4. Vehicle Choice Share Optimization Problem

GIVEN

a) Vehicle Dimensions (**L103, W103, H100**) for Vehicle 4
b) Preference Models for $\mathbf{M_1}$, $\mathbf{M_2}$, and $\mathbf{M_3}$ (
c) Table 6.2, Table 6.1)
d) Models for fuel economy and weight
e) Target Market Demographics (**S**)

FIND:

***Dimensions*: HNG_X, ROK_Y, HEL_Z, GRD_Z, $StoH$, HR_Z**

to MAXIMIZE

Choice share, Q, for Vehicle 4

Subject to:

- H100 ≤ 58 in.
- Fuel Economy f(engine size, L103, W103, H100, *weight*) ≥ 24.0 *mpg*
- Weight =*f*(L103, W103, H100) ≤ 3500 *lbs*

Relationships:

- $L103 = 165.4\text{ in.} + HNG_X$
- $W103 = 41.8\text{ in.} + 2 \cdot ROK_Y$
- $H100 = 2.0\text{ in.} + GRD_Z + HR_Z[\cos(8\text{ deg})]$

The results of the optimization are shown in Table 6.5, with the current values of the six dimensions to be optimized listed under *stating value*. The initial choice share estimated using the hierarchical choice model is 32.65%. The optimization converges with the final optimum values for the six dimensions listed in the table, leading to a choice share increase to 40.07 % using the *Random INT* model vs. 41.97% for the *Fixed INT* model. The final values of fuel economy, weight, and vehicle height are also presented below, with GRD_Z and StoH reaching the lower constraint on their values. The optimum values of the variables and the maximum choice share solution are different when using the fixed parameter model. Because, as will be shown in Section 6.4, the goodness of fit of the random parameter model is significantly higher than the fixed parameter model and the choice share prediction accuracy higher, the choice share prediction accuracy is higher for the random parameter model.

Table 6.5. Optimization Results for the Package Design

		Optimum Value	
Attribute	**Starting Value**	**Random INT**	**Fixed INT**
HNG_X	**29.5 in.**	**31.5 in.**	**31.5 in.**
ROK_Y	**16.1 in.**	**15.4 in.**	**16.1 in.**
HEL_Z	**12.2 in.**	**13.0 in.**	**13.0 in.**
GRD_Z	**22.0 in.**	**21.6 in.**	**21.6 in.**
StoH	**5.6 in.**	**5.3 in.**	**5.3 in.**
HR_Z	**32.2 in.**	**36.3 in.**	**33.6 in.**
Fuel Economy	24.2 mpg	24.2 mpg	24.0 mpg
Weight	3480 lb	3460 lb	3475 lb
Vehicle Height (H100)	56.9 in.	56.7 in.	55.2 in.
Vehicle Choice Share	**32.65%**	**40.07%**	**41.97 %**

6.4 Validation of the Integrated Bayesian Hierarchical Choice Model

The IBHCM is validated both to ensure convergence of the model as well as to test the fit of the model and its ability to accurately predict choices within the data set. The four model scenarios presented in Section 6.3.1 are used for comparison purposes.

IBHCM Convergence: Convergence of the Monte Carlo Markov Chains (MCMC) is assessed to determine if the posterior distribution is stationary and thus is a reasonable approximation of the actual posterior distribution. The most popular practical measure of MCMC convergence is the Gelman-Rubin $\hat{R}$ statistic (Brooks and Roberts, 1998). In order to test for convergence, at least two chains must be utilized and two intermediate measures must be determined to calculate the

$\hat{R}$ statistic. A measurable quantity of the each iteration of each chain for each model parameter (i.e. b, Σ) is defined as $\boldsymbol{\omega}$. The first measure is the *between-chain* variance, B, and the second measure is the *within-chain* variance, W, as illustrated in Figure 6.10. The expressions for B and W are given in Eq. (6.16) as:

$$\begin{aligned} B &= \frac{n}{m-1}\sum_{k=1}^{m}\left(\overline{\boldsymbol{\omega}}_k - \overline{\boldsymbol{\omega}}\right)^2 \\ W &= \frac{1}{m(n-1)}\sum_{j=1}^{n}\sum_{k=1}^{m}\left(\omega_{jk} - \overline{\boldsymbol{\omega}}_k\right)^2 \end{aligned}, \tag{6.16}$$

where m is the number of parallel chains, n is the number of realizations of each chain, $\overline{\boldsymbol{\omega}}$ is the overall mean value of all $m{\cdot}n$ realizations, $\overline{\boldsymbol{\omega}}_k$ is the mean of the n realizations of chain k, and ω_{jk} is the j^{th} realization of chain k.

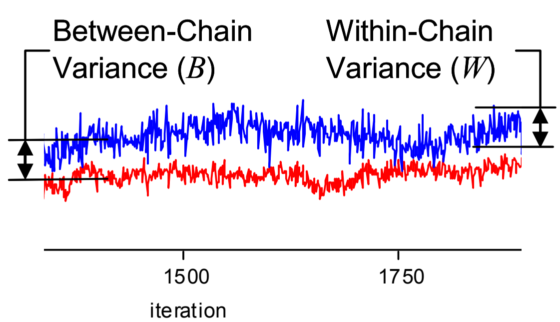


Figure 6.10: Example of Between vs. Within Chain Variance

Using the measures B and W, the $\hat{R}$ statistic is defined as:

$$\hat{R} = \sqrt{\frac{(1-1/n)W + (1/n)B}{W}}\,. \tag{6.17}$$

The within-chain variance, W, will initially be small as the sampler will not fully explore the state space, whereas the between-chain variance, B, will initially be large before the j chains have converged to the posterior distribution. Therefore, $\hat{R}$ will initially be large, but will converge to 1.0 as the j chains converge to the posterior distribution (Lancaster, 2004). While there is no formal definition for acceptable convergence, it is recommended by Gelman-Rubin that a value of $\hat{R}$ of less than 1.2 is a reasonable measure of convergence for a chain (Gelman et al., 2004).

Histograms of $\hat{R}$ for each of the model parameters (i.e. b, Σ) are shown in Figure 6.11. As seen in the figures, each of the models generally converged with $\hat{R}$ less than 1.2; however, each of the models has a few (less than 5%) of the chains with $\hat{R}$ greater than 1.2. These outliers have been investigated and found to be related either to the ordered logit cut points, or the variance components, and not the model parameters in the utilities functions (i.e. $\boldsymbol{\beta}$). The cut points have more difficulty in converging, possibly due to the ordinal constraint upon them.

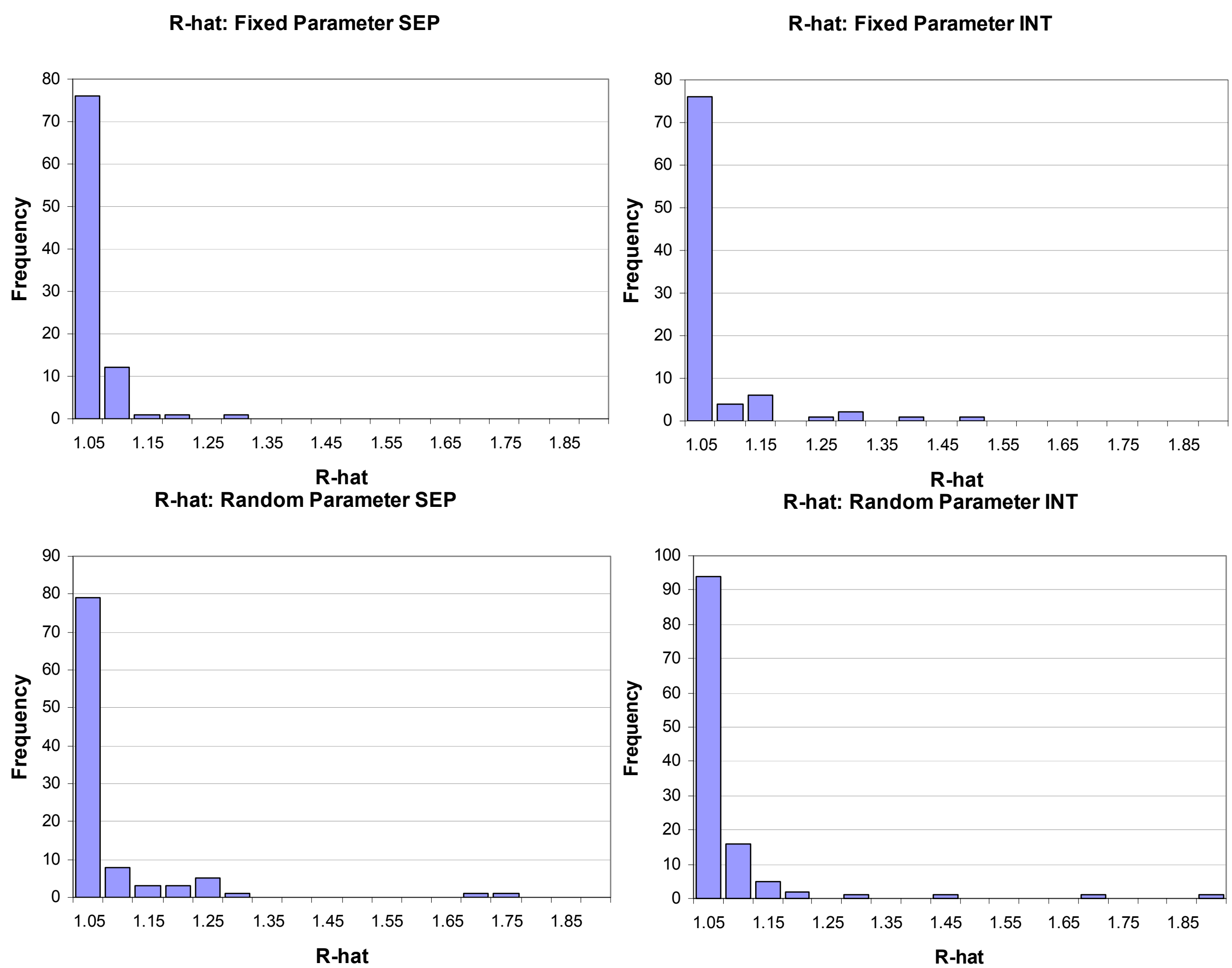


Figure 6.11: R-hat Statistic Distribution for Parameters in Each Model Scenario

<u>IBHCM Model Fit and Prediction Tests</u>: Unlike physics-based models, model validation for behavioral models is challenging in that a physical experiment may be difficult to conduct to validate the model. Validation must be done utilizing the same data available for model estimation in most situations. The following validation techniques are used for behavioral models:

1. *Goodness-of-Fit Measures*: Goodness-of-fit measures based upon the log-likelihood of the converged model, such as the likelihood ratio index ρ_0^2, are measures of how well the estimated model predicts actual *individual* choices in the data set. Higher values of ρ_0^2 indicate better prediction of the choices.
2. *Comparison Bayesian MXL to MLE NL*: As described in the previous section, the method for data fusion used in the Bayesian MXL method is different than that used in the Maximum Likelihood Nested Logit method.
3. *Choice Share/Segment Prediction Tests*: Due to the hierarchical nature of the model, prediction errors in the lower level models propagate to the choice level model, creating inaccuracies in choice prediction. Therefore, a test is conducted to determine overall vehicle choice share prediction accuracy, as well as a test of predictions on specific segments of the market, for example predictions on several segments of human stature (Ben-Akiva and Lerman, 1985).
4. *Confirmation of Effect of Modeling Heterogeneity*: The effect of including both systematic and random heterogeneity in the model is shown.

Hierarchical choice models were estimated for the four scenarios outlined previously. The log-likelihood and the statistic ρ_0^2 for each of the models within the hierarchical model framework are reported in Table 6.6.

Table 6.6: Comparison of the Model Fits of the 4 Scenarios

	Fixed SEP		Fixed INT		Random SEP		Random INT	
	L-L	ρ_0^2	*L-L*	ρ_0^2	*L-L*	ρ_0^2	*L-L*	ρ_0^2
M_1 Model	-209.99	**0.353**	-228.88	**0.294**	-67.00	**0.793**	-135.71	**0.582**
M_2 Ingress/Egress	-1198.18	**0.316**	-1197.03	**0.317**	-906.63	**0.483**	-924.24	**0.473**
M_2 Roominess	-867.99	**0.471**	-1086.98	**0.338**	-733.56	**0.553**	-810.70	**0.506**
M_3 Headroom	-1104.35	**0.339**	-1101.91	**0.340**	-845.64	**0.494**	-865.22	**0.482**
M_3 Leftroom	-1146.00	**0.320**	-1144.64	**0.320**	-804.25	**0.523**	-827.82	**0.509**
M_3 Kneeroom	-1204.12	**0.307**	-1205.65	**0.307**	-889.15	**0.489**	-907.28	**0.478**
	-5730.6235		**-5965.0935**		**-4246.233**		**-4470.96**	

Goodness-of-Fit: As seen in the table, significantly higher log-likelihood and subsequently ρ_0^2 values are achieved using the random parameter models versus the fixed parameter models. This is to be expected as the random parameter model captures random taste heterogeneity in addition to the systematic taste heterogeneity of the fixed parameter models. The inclusion of random heterogeneity provides the largest improvement to the M_1 level choice model, indicating that there is much taste variation at the choice level not accounted for by the choice of model parameters. Additionally, the models estimated separately have better goodness-of-fit statistics than the integrated estimated models. This is due to the fact that the parameters in the integrated models are fit to models in which the response is predicted by a model as opposed to the actual responses.

Comparison Bayesian MXL to MLE NL: A comparison of the Bayesian MXL versus the Nested Logit method is shown in Table 6.7 for the customer-desired attribute (**A**) in the model. The MLE NL results are scaled to be on the same scale as the Bayesian MXL results, because, as discussed earlier in this chapter, MLE identifies the *mode* of the parameter distribution versus

Bayesian which identifies the *mean* of the distribution. As seen in the table, there is strong agreement between the two methods.

Table 6.7: Comparison Bayesian MXL vs. Nested Logit Data Fusion

	Bayesian MXL		MLE NL		Scaled NL	Difference
Roominess	**0.53**	*4.11*	**0.51**	*3.78*	**0.53**	0.07%
Ingress/Egress	**0.43**	*3.75*	**0.41**	*3.55*	**0.43**	0.81%
Quality Materials	**0.41**	*3.90*	**0.40**	*3.77*	**0.42**	2.83%
Ext. Appearance	**1.46**	*5.22*	**1.36**	*5.06*	**1.43**	2.10%
Willing to Pay	**0.39**	*2.06*	**0.38**	*1.92*	**0.39**	1.49%
Sigma/scale	**1.31**	*2.44*	**1.15**	*0.14*		

The NL or MXL method of combining data can be compared to a MNL method in which the multiple data sets are merged into a single set and a single scale, μ, based upon the pooled variance of both data sets is used. Because all parameters in such an approach are scaled by the same pooled scale factor, the MNL approach will over or under-estimate the importance of model coefficients. For example, estimating the M_1 model using the single scale MNL approach and comparing to the NL approach demonstrates that the exterior styling-to-interior packaging ratio decreases from $1.46/.53 = 2.76$ estimated using the MXL approach, to $1.36/.54 = 2.52$ using the MNL approach. As seen, the importance for interior packaging would be overestimated while the importance for exterior styling is underestimated using the MNL approach.

Choice Share Predictions: The models are compared based upon the error in **choice share predictions** for the four different vehicles in the choice set as shown in Table 6.8.

Table 6.8: Comparison of Choice share Predictions for the 4 Scenarios

		Fixed Parameter C.S. Predictions				*Random Parameter C.S. Predictions*			
	Actual	**SEP**	***% Error***	**INT**	***% Error***	**SEP**	***% Error***	**INT**	***% Error***
Veh. 1	0.1410	**0.1412**	*0.124*	**0.1370**	*2.855*	**0.1353**	*4.060*	**0.1441**	*2.180*
Veh. 2	0.3803	**0.3579**	*5.900*	**0.3677**	*3.324*	**0.3764**	*1.036*	**0.3758**	*1.194*
Veh. 3	0.1795	**0.1753**	*2.333*	**0.1958**	*9.089*	**0.1727**	*3.781*	**0.1774**	*1.163*
Veh. 4	0.2991	**0.3256**	*8.843*	**0.2995**	*0.119*	**0.3157**	*5.534*	**0.3027**	*1.188*

The error for both fixed parameter models is relatively high; however, the integrated approach successfully minimizes the error of the vehicle under design (Veh. 4) from 8.84% to 0.12%. Introducing random parameters has the effect of more evenly distributing the error among the C.S. predictions, thus reducing the maximum choice share errors. The integrated estimation with random parameters has a similar effect as the integrated fixed parameter estimation in that it significantly reduces the error of the vehicle under design (Veh. 4), but also lowers the C.S. prediction error for the other vehicles as well.

The **market segment prediction test** is conducted for three segments of Stature (small, medium, large) and three segments of Age (low, medium, high). The results of the Stature market segmentation test are shown in Table 6.9 and the results of the Age market segment test are shown in Table 6.10. In order to determine a 95% confidence interval for the segments, the variance of the observed choice share is calculated using the binomial proportion confidence interval (Ben-Akiva and Lerman, 1985):

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_s}}, \tag{6.18}$$

where $\hat{p}$ is the observed choice share proportion, $z_{\alpha/2} = 1.96$ for a 95% confidence interval, and n_s is the number of people in each market segment.

Table 6.9: Stature Market Segment Test

Small Stature Segment

	Fixed Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Fixed Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.094**	**0.348**	**0.226**	**0.332**	**0.084**	**0.384**	**0.248**	**0.284**
Observed	0.100	0.388	0.238	0.275	0.100	0.388	0.238	0.275
95% C. I.	0.034 0.166	0.281 0.494	0.144 0.331	0.177 0.373	0.034 0.166	0.281 0.494	0.144 0.331	0.177 0.373
% Error	*6.00%*	*10.09%*	*4.84%*	*20.58%*	*15.70%*	*0.95%*	*4.42%*	*3.24%*

Medium Stature Segment

	Fixed Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Fixed Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.159**	**0.405**	**0.146**	**0.291**	**0.157**	**0.408**	**0.172**	**0.264**
Observed	0.125	0.513	0.175	0.188	0.125	0.513	0.175	0.188
95% C. I.	0.053 0.198	0.403 0.622	0.092 0.258	0.102 0.273	0.053 0.198	0.403 0.622	0.092 0.258	0.102 0.273
% Error	*26.88%*	*21.00%*	*16.86%*	***55.20%***	*25.36%*	*20.47%*	*1.60%*	***40.59%***

Large Stature Segment

	Fixed Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Fixed Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.174**	**0.317**	**0.153**	**0.357**	**0.173**	**0.307**	**0.165**	**0.355**
Observed	0.203	0.230	0.122	0.446	0.203	0.230	0.122	0.446
95% C. I.	0.111 0.294	0.134 0.326	0.047 0.196	0.333 0.559	0.111 0.294	0.134 0.326	0.047 0.196	0.333 0.559
% Error	*14.41%*	*38.09%*	*25.66%*	*20.05%*	*14.85%*	*33.74%*	*35.77%*	*20.36%*

Small Stature Segment

	Random Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Random Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.080**	**0.386**	**0.238**	**0.296**	**0.086**	**0.396**	**0.240**	**0.278**
Observed	0.100	0.388	0.238	0.275	0.100	0.388	0.238	0.275
95% C. I.	0.034 0.166	0.281 0.494	0.144 0.331	0.177 0.373	0.034 0.166	0.281 0.494	0.144 0.331	0.177 0.373
% Error	*20.00%*	*0.41%*	*0.25%*	*7.64%*	*13.70%*	*2.14%*	*1.18%*	*0.95%*

Medium Stature Segment

	Random Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Random Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.147**	**0.428**	**0.128**	**0.297**	**0.150**	**0.438**	**0.153**	**0.259**
Observed	0.125	0.513	0.175	0.188	0.125	0.513	0.175	0.188
95% C. I.	0.053 0.198	0.403 0.622	0.092 0.258	0.102 0.273	0.053 0.198	0.403 0.622	0.092 0.258	0.102 0.273
% Error	*17.52%*	*16.47%*	*26.69%*	***58.24%***	*20.32%*	*14.58%*	*12.63%*	***38.13%***

Large Stature Segment

	Random Parameter SEP Veh. 1	Veh. 2	Veh. 3	Veh. 4	Random Parameter INT Veh. 1	Veh. 2	Veh. 3	Veh. 4
Predicted	**0.182**	**0.310**	**0.150**	**0.357**	**0.203**	**0.266**	**0.157**	**0.373**
Observed	0.203	0.230	0.122	0.446	0.203	0.230	0.122	0.446
95% C. I.	0.111 0.294	0.134 0.326	0.047 0.196	0.333 0.559	0.111 0.294	0.134 0.326	0.047 0.196	0.333 0.559
% Error	*10.01%*	*35.05%*	*23.36%*	*19.85%*	*0.05%*	*15.98%*	*29.36%*	*16.26%*

Table 6.10: Age Market Segment Test

Fixed Parameter SEP / Fixed Parameter INT

Low Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.177**	**0.354**	**0.164**	**0.306**	**0.186**	**0.341**	**0.185**	**0.288**
Observed	0.174	0.419	0.163	0.244	0.174	0.419	0.163	0.244
95% C. I.	0.094 0.255	0.314 0.523	0.085 0.241	0.153 0.335	0.094 0.255	0.314 0.523	0.085 0.241	0.153 0.335
% Error	*1.38%*	*15.53%*	*0.43%*	***25.35%***	*6.54%*	*18.54%*	*13.51%*	***18.10%***

Medium Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.093**	**0.333**	**0.190**	**0.384**	**0.083**	**0.350**	**0.215**	**0.352**
Observed	0.100	0.314	0.200	0.386	0.100	0.314	0.200	0.386
95% C. I.	0.030 0.170	0.206 0.423	0.106 0.294	0.272 0.500	0.030 0.170	0.206 0.423	0.106 0.294	0.272 0.500
% Error	*7.50%*	*6.05%*	*5.00%*	*0.41%*	*16.80%*	*11.20%*	*7.50%*	*8.66%*

High Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.146**	**0.385**	**0.175**	**0.294**	**0.131**	**0.414**	**0.191**	**0.264**
Observed	0.141	0.397	0.180	0.282	0.141	0.397	0.180	0.282
95% C. I.	0.064 0.218	0.289 0.506	0.094 0.265	0.182 0.382	0.064 0.218	0.289 0.506	0.094 0.265	0.182 0.382
% Error	*3.40%*	*3.22%*	*2.40%*	*4.36%*	*6.88%*	*4.08%*	*6.30%*	*6.35%*

Random Parameter SEP / Random Parameter INT

Low Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.170**	**0.361**	**0.159**	**0.310**	**0.178**	**0.369**	**0.163**	**0.290**
Observed	0.174	0.419	0.163	0.244	0.174	0.419	0.163	0.244
95% C. I.	0.094 0.255	0.314 0.523	0.085 0.241	0.153 0.335	0.094 0.255	0.314 0.523	0.085 0.241	0.153 0.335
% Error	*2.47%*	*13.69%*	*2.64%*	***26.95%***	*1.83%*	*11.80%*	*0.12%*	***18.84%***

Medium Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.086**	**0.367**	**0.186**	**0.361**	**0.092**	**0.343**	**0.209**	**0.357**
Observed	0.100	0.314	0.200	0.386	0.100	0.314	0.200	0.386
95% C. I.	0.030 0.170	0.206 0.423	0.106 0.294	0.272 0.500	0.030 0.170	0.206 0.423	0.106 0.294	0.272 0.500
% Error	*14.10%*	*16.67%*	*6.85%*	*6.40%*	*8.50%*	*9.07%*	*4.50%*	*7.52%*

High Age Segment

	SEP *Veh. 1*	SEP *Veh. 2*	SEP *Veh. 3*	SEP *Veh. 4*	INT *Veh. 1*	INT *Veh. 2*	INT *Veh. 3*	INT *Veh. 4*
Predicted	**0.141**	**0.402**	**0.176**	**0.281**	**0.145**	**0.402**	**0.181**	**0.272**
Observed	0.141	0.397	0.180	0.282	0.141	0.397	0.180	0.282
95% C. I.	0.064 0.218	0.289 0.506	0.094 0.265	0.182 0.382	0.064 0.218	0.289 0.506	0.094 0.265	0.182 0.382
% Error	*0.07%*	*1.08%*	*1.89%*	*0.35%*	*3.12%*	*1.03%*	*0.95%*	*3.62%*

In general, the choice shares vary more substantially by Stature market segment than by Age segment. In the Stature study, only 2 choice share predictions are outside the 95% confidence intervals; one prediction is from the Fixed Parameter SEP model, while the other is from the Random Parameter SEP model. No observations are outside the 95% confidence intervals in the Age study.

Modeling Heterogeneity: The non-linear choice vs. utility curve implies that the model with the best representation of heterogeneity and the least restrictions on the choice probabilities (i.e. I.I.A.) will provide the best estimate of the choice share for a design change or the introduction of a new design. To demonstrate this concept, a comparison among five DCA models (i.e. M_1 level only models) estimated for the set of four competing vehicles in $\boldsymbol{DS}_1$ and $\boldsymbol{DS}_2$ is provided. The five choice models estimated are as follows:

- *Model 0*: Aggregate Logit model estimated using average vehicle ratings.
- *Model 1*: MNL model with no demographics (no **S**).
- *Model 2*: MNL model with demographics included (**S**).
- *Model 3*: MXL model with no demographics (no **S**).
- *Model 4*: MXL model with demographics included (**S**).

The five models are estimated, resulting in initial choice shares for each of the four vehicles of [0.141, 0.380, 0.180, 0.299]. Case studies are conducted in which various hypothetical changes to the design of each of the vehicles are made individually, which are assumed to increase the respondent ratings by two points (e.g. rating 4 increases to 6) for each change. The effects of the design changes upon the choice share of the changed vehicle are shown in Table 6.11 for each of the four cases investigated. For example, the Ingress (*Ing*) and Exterior Styling (*Ext*) ratings are increased (+) for Vehicle 1 (*Case 1*) in the first case study, with the predicted choice share (*C.S.*) estimated using each model (*Model 0–Model 4*) for the improved design shown in the respective row.

Table 6.11. Effect of Design Changes on Choice Share using Different Models

	Case 1 C.S.	*Case 2 C.S.*	*Case 3 C.S.*	*Case 4 C.S.*	
Initial C.S.	*0.141*	*0.380*	*0.180*	*0.299*	
Attribute Change	*+ Ing + Ext*	*+ Room + Ext*	*+ Room + Ext*	*+ Q Mat. + W to P*	ρ_0^2
Model 0	0.418	0.830	0.814	0.489	*0.056*
Model 1	**0.270**	**0.588**	**0.383**	**0.416**	*0.350*
Model 2	**0.260**	**0.595**	**0.395**	**0.407**	*0.381*
Model 3	**0.318**	**0.592**	**0.411**	**0.466**	*0.678*
Model 4	**0.313**	**0.600**	**0.429**	**0.472**	*0.806*

The differences among the five models are characterized by the goodness-of-fit measure, ρ_0^2, which ranges between 0-1, with higher values indicating better model fit. The ρ_0^2 metric indicates that the MXL with **S** model has the best fit, while the aggregate (i.e. average rating) model has the worst fit. Considering only the aggregate logit model in comparison to the four other models for each case, it is seen that the aggregate model always overestimates the effect of a design change. In the first case study using Vehicle 1, the models without **S** tend to overestimate the effect of the change for a given model type (i.e. MNL and MXL), while the MNL models underestimate the effect of the change versus the respective MXL models (i.e. with **S** and without **S**). Other patterns of over- or under-estimation are seen in the case studies for Vehicles 2-4.

6.5 Discussion and Summary

The Integrated Bayesian Hierarchical Choice Modeling framework proposed in this work utilizes multiple model levels to create a link between qualitative attributes considered by consumers when selecting a product and quantitative attributes used for engineering design. This new framework addresses the shortcomings of previous methods while providing a highly flexible modeling framework to address the needs of complex system design, such as the vehicle design problem considered in this work. In the proposed framework, both systematic and random consumer heterogeneity is explicitly modeled at all levels of the model hierarchy. The importance of including a complete representation of consumer heterogeneity in the model framework is clearly demonstrated using the vehicle design example. The ability to combine multiple sources of data for model estimation and updating is significantly expanded over previous methods. A comprehensive method to mitigate error propagated throughout the model hierarchy is developed and its effectiveness demonstrated. The new modeling framework is demonstrated for the vehicle occupant package design, in which optimal vehicle package dimensions are identified. The modeling approach is validated using several metrics and techniques, demonstrating the ability of the new approach to better capture heterogeneous consumer preferences and mitigate error propagated.

The integrated Bayesian hierarchical choice model, formulated for model updating and model fusion, can be incorporated into the overall economic benefits equation of Figure 6.2 (i.e. $V = QP - C$), forming the selection criterion used in the enterprise utility function. The role of the Bayesian hierarchical choice model is illustrated in Figure 6.12. Initially, at time t=0, the prior information and the evidence are combined to estimate the $\boldsymbol{\beta}$ parameters of the choice model (including M_1, M_2, and M_3 levels), which together with an estimate of the total market size, $D(t)$, enables estimation of Q. To complete the profit function, costs and relevant uncertainties are quantified as was demonstrated in the sensor case study in Section 3.5. With a profit function available and the risk attitude of the enterprise, the enterprise utility function can be formulated. As illustrated in Figure 6.12, the demand model can be updated in future time periods (e.g., $t = t_1, t_2, \ldots, t_j$) as more information becomes available, as preferences change, or a combination of both information and preference change.

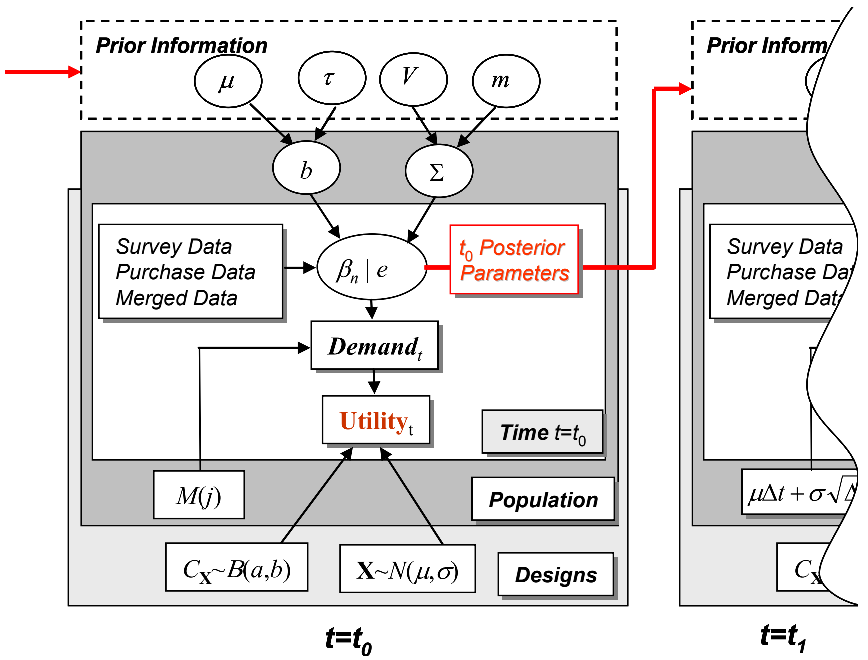


Figure 6.12: Bayesian Choice Modeling Framework

The hierarchical choice modeling approach presented in this chapter will make possible the realization of a comprehensive Decision-Based Design framework for complex systems, in which a hierarchy of systems and sub-systems exist, as well as multiple sources of survey data over different time periods. Using this method, detailed design decisions can be made on a single or multiple sub-systems, or for the entire system.

Chapter 7

CONCLUSIONS AND INTELLECTUAL MERIT

7.1 CONTRIBUTIONS OF THE DISSERTATION

The primary research contribution of this work is the development of the comprehensive **Enterprise-Driven Design approach** to support design decision-making for configuring complex engineering systems, considering the heterogeneity of the consumers for these systems. The specific tools and methodologies which comprise the system configuration approach are necessary for the design of complex systems in which there is a hierarchy of sub-systems and product attributes, consumer heterogeneity results in different experiences and preferences for the system design, and there is a need to combine multiple sources of data, update models, and quantify uncertainty for decision making. This framework is built upon the principles of the Decision-based Design (DBD) paradigm, providing a tool to implement the DBD method in application, a method to conduct the surveys required to support preference modeling, and a comprehensive choice modeling approach to support engineering design. The proposed approach provides a rigorous design approach which is suitable for use on a wide variety of engineering systems in a wide variety of markets.

The specific research contribution of each of the new tools and methodologies comprising the enterprise-drive design approach is detailed as follows. The **Product Attribute Function Deployment** (PAFD) method is developed to offer a mathematically rigorous, decision-theoretic process tool for use during the product planning phase of a product development program. Such a method is needed based on an investigation of the flaws of current methods, such as Quality

Function Deployment, which could lead to a faulty design decision process. The PAFD method extends the QFD mapping matrix concept to qualitatively identify relationships and interactions among product design attributes while employing the DBD principles to provide rigorous quantitative assessments for design decisions. The PAFD method can be implemented for a real design problem, with a team composed of marketing science, engineering, and manufacturing experts.

The **Optimal Design of Human Appraisal** method can be widely applied to assess consumer preference for any system in which an interaction between the design and a human user exists. In this work, the attributes of the consumer which were believed to influence the consumer's experience of the system where human attributes, such as the consumer's height, weight, or gender; however, the demographics used in the design of the experiment can be much broader, such as usage context or skill level of the user of the system. The optimal design of human appraisal method is a necessary development to complement the development of flexible prototype hardware, such as the Programmable Vehicle Model, which can assume a wide variety of configurations for evaluation.

The methods for **Statistical Data Analysis of Consumer Heterogeneity** can be applied to a wide variety of human appraisal data and the respective predictive models estimated using the data. These methods are developed specifically for separating the effect of the different types of respondent heterogeneity from the influence of the product design attributes. Application of the methods will result in better predictive models for forecasting the impact of new designs or design improvements on consumer opinion, and ultimately enterprise profitability.

The proposed **Integrated Bayesian Hierarchical Choice Modeling** approach provides the necessary comprehensive choice modeling methodology to guide the design of a complex system, characterized by a hierarchy of component, sub-system, and system design activities. This approach explicitly captures the effect of consumer heterogeneity in the choice process, and is formulated to address the challenges of complex system design, such as qualitative choice attributes, multiple data sets, and the need to quantify uncertainty for decision making. This methodology could find wide spread use in setting target levels of performance for complex systems, such as automobiles, airplanes, medical devices, or power tools in which the system design must meet the needs of a diverse consumer population. The approach could also find application in other disciplines, such as in the design of services, which are characterized by multiple data sets, preference heterogeneity, and mapping from qualitative choice attributes (e.g. convenience of a service) to quantitative measures (e.g. hours service is offered).

The methods developed in this research can be applied to several trends within industry today. One such trend is the development and management of ***incremental innovation***. Much focus has been directed to *breakthrough* innovation, in which a new technological breakthrough creates a brand new market, with no immediate competitors and potentially high profits. These types of breakthroughs are rare, however (Pine, 1993; Otto and Wood, 2001); it has been noted most recently in the *Harvard Business Review* (Kanter, 2006) that attention must also be paid to incremental innovations, which are capable of creating competitive advantages for a firm in existing markets, to enable incremental improvements in profitability and/or market share. These incremental innovations must be implemented in product configurations to ensure consumer acceptance and profitability. Without solid methods for decision-making to manage innovation, enterprises must overly rely upon mimicking (benchmarking) successful competitors, creating superficial cosmetic changes to existing products to generate interest, or introducing a wide variety of disparate products to mitigate the uncertainties of the market place. The methods provided in this dissertation can be used to guide the design process for configuring systems to include incremental innovations.

Another issue to address is the increasingly rapid obsolescence of product designs. For example, in the cell phone industry, product cycles are short and consumers demand new products at an ever increasing rate. As noted recently in *Business Week* (Crockett, 2007), companies are looking toward updating popular products to maintain interest throughout the product life, rather than waiting for introduction of entirely new products. These changes are intended to improve the base design, as well as to optimize the features to correspond with current consumer preferences. The Integrated Bayesian Hierarchical Choice Model provides a method to update consumer preference models at any time throughout the product design life cycle, and make decisions upon feature improvements.

7.2 RECOMMENDATIONS FOR FUTURE WORK

PAFD Method: The recommended future work for the PAFD method is to study the use of the method for the design of a complex system, such as the vehicle problem. When the PAFD tool is used in the conceptual design phase, it is not reasonable to expect that the detailed survey data used in the hierarchical choice modeling approach would be available. Also, the knowledge needed to conduct the PAFD analysis may be spread over several disciplines, requiring the use of multidisciplinary optimization techniques. Therefore, research is required to determine how to implement the method considering the unique issues present in the selection of a preferred complex system concept. Another area of future work is extension of the PAFD method for use in the detailed design phase, such as the target setting process for the vehicle problem. While it is intended that the PAFD method be applied throughout the design process, from the conceptual design phase through the detailed design phase, it has been demonstrated in this work for use in conceptual design. The principles of mapping qualitative customer-desired attributes to quantitative engineering attributes are applicable to the design of a human appraisal experiment presented in Chapter 4 and to guide the modeling described in Chapters 5 and 6.

Optimal Design of Human Appraisal Experiments: The primary research need for the Design of Human Appraisal Experiment method is an improved search method. Currently several tries are used to identify the optimal design, with no methodology available for determining the number of tries needed to indentify an acceptable design. Improvements to the efficiency of the search algorithm should be investigated using memetic or stochastic evolutionary algorithms, which would eliminate the need for multiple tries and potentially lead to more repeatable results using the algorithm. Another future research area is to adapt the method to design choice experiments to support building the discrete choice model at the top level of the model hierarchy. This could be accomplished using the basic framework in place for the design of human appraisal experiments, but replacing the information matrix for the random effects ordered logit model with the information matrix for the MNL or MXL model.

Analysis of Human Appraisal Experiments: Methods have been developed for preprocessing the data collected in a human appraisal experiment, specifically for response reduction, understanding the factor-response relationship, and general methods to guide the random effects ordered logit modeling process. The use of machine learning methods from data mining, specifically a decision tree and a Bayesian network, were also investigated as methods to better understand the data. The primary future work in this area is further exploration and use of the machine learning methods to support the modeling process. The machine learning methods have a primary advantage in that they can identify the important factors which influence the classification process, such as selection of a rating or a choice made from among a choice set.

Further, a better understanding of the advantages and disadvantages of the parametric modeling methods, such as ordered logit and discrete choice analysis, versus the machine learning classification methods would be beneficial to the Decision-based design paradigm. For example, design situations in which a well structured data set with defined choice sets does not exist may be better served with the use of a Bayesian network or Decision Tree.

Integrated Bayesian Hierarchical Choice Model: Several areas for future work remain in for the Bayesian hierarchical choice modeling approach. Convergence of the hierarchical model, as measured by the Gelman-Rubin statistic, is a challenge for such models. In this work, a simplified set of models at the M_1, M_2, and M_3 levels was used to demonstrate the method; however, it is desired to utilize more descriptive models, such as the complete random-effects ordered logit models of Chapter 5, in the hierarchy. The Bayesian framework has been developed with the ability to both combine multiple data sources to estimate the choice model, as well the ability to update the model over time. An example of combining data at the choice model level, and updating at the ratings level was provided; however, an example in which data from several product segments is combined to estimate the models, or in which the entire set of models is updated with a complete set of data from another model year is used should be investigated. Also, the use of other types of data sets, for example actual purchase data (i.e. revealed preference) such as data collected by J. D. Power and Associates, should be investigated.

References

Allenby, G., G. Fennell, J. Huber, T. Eagle, T. Gilbride, D. Horsky, J. Kim, P. Lenk, R. Johnson, and E. Ofek. "Adjusting Choice Models to Better Predict Market Behavior." *Marketing Letters* 16, no. 3 (2005): 197-208.

Altshuller, G. S., and A. Williams. *Creativity as an Exact Science: The Theory of the Solution of Inventive Problems*. New York: Gordon and Breach Science Publishers, 1984.

Armacost, R. L., P. J. Componation, M. A. Mullens, and W. W. Swart. "An AHP Framework for Prioritizing Customer Requirements in QFD: An Industrialized Housing Application." *IIE Transactions* 26, no. 4 (Jul. 1994): 72-79.

Atkinson, A. C., and A. N. Donev. *Optimum Experimental Designs*. Oxford: Oxford University Press, 1992.

Aungst, S., R. Barton, and D. Wilson. "The Virtual Integrated Design Method." *Quality Engineering* 15, no. 4 (Oct. 2003): 565-79.

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

Besharati, B., S. Azarm, and A. Farhang-Mehr. "A Customer-Based Expected Utility Metric for Product Design Selection." *Proceedings of ASME 2002 IDETC Conference*. Montreal, Canada, September, 2002.

Box, G. E. P., J. S. Hunter, and W. G. Hunter. *Statistics for Experimenters: Design, Innovation, and Discovery*. New York: John Wiley & Son, 2005.

Brackin, P., and J. Colton. "A Strategy for Extending the House of Quality to Obtain Preliminary Design Specifications." *Proc. 1999 ASME Design Engineering Technical Conf.*, Las Vegas NV, September, 1999.

Bradley, S. R., and A. M. Agogino. "Intelligent Real Time Design Methodology for Component Selection: An Approach to Managing Uncertainty." *Journal of Mechanical Design* 116, no. 4 (Dec. 1994): 980-88.

Braun, R. D. "Collaborative Optimization: An Architecture for Large-Scale Distributed Design." PhD Thesis, Stanford University, 1996.

Brooks, S. P., and G. O. Roberts. "Assessing Convergence of Markov Chain Monte Carlo Algorithms." *Statistics and Computing* 8, no. 4 (1998): 319-35.

Brownstone, D., D. S. Bunch, and K. Train. "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles." *Transportation Research Part B* 34, no. 5 (2000): 315-38.

Brownstone, D., and K. Train. "Forecasting New Product Penetration with Flexible Substitution Patterns." *Journal of Econometrics* 89, no. 1-2 (1998): 109-29.

Chan, L. K., H. P. Kao, and M. L. Wu. "Rating the Importance of Customer Needs in Quality Function Deployment by Fuzzy and Entropy Methods." *International Journal of Production Research* 37, no. 11 (Jul. 1999): 2499-518.

Chan, L. K., and M. L. Wu. "Quality Function Deployment: A Literature Review." *European Journal of Operational Research* 143, no. 3 (Dec. 2002): 463-97.

Chen, W., R. Jin, and A. Sudjianto. "Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design Under Uncertainty." *Journal of Mechanical Design* 127, no. 5 (Sept. 2005): 875-86.

Chipman, H. A., and W. J. Welch. "D-Optimal Design for Generalized Linear Models." *Unpublished* (1996).

Clausing, D., and J. Hauser. "The House of Quality." *Harvard Business Review* 66, no. 3 (May 1988): 63-73.

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

Cortina, J. M. "What Is Coefficient Alpha? An Examination of Theory and Applications." *Journal of Applied Psychology* 78, no. 1 (1993): 98-104.

Cox III, E. P. "The Optimal Number of Response Alternatives for a Scale: A Review." *Journal of Marketing Research* 17, no. 4 (1980): 407-22.

Crockett, R. O. "Honing the Razr Edge." *Business Week*, May 28, 2007.

de Weck, O.L., and E.S. Suh. "Flexible Product Platforms: Framework and Case Study." *Proceedings of the 2003 ASME Design Engineering Technical Conferences*. Philadelphia, PA, Sept. 10-13, 2006.

Dong, H., and W. Wood. "Integrating Computational Synthesis and Decision-Based Conceptual Design." *Proc. 2004 ASME Design Engineering Technical Conf.*, Salt Lake City, UT, September, 2004.

Edmunds Inc. www.edmunds.com. Accessed: Sept. 21, 2007, 1995-2007.

Erdem, T., and M. P. Keane. "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets." *Marketing Science* 15, no. 1 (1996): 1-20.

Frey, D. D., F. Engelhardt, and E. M. Greitzer. "A Role for "One-Factor-at-a-Time" Experimentation in Parameter Design." *Research in Engineering Design* 14, no. 2 (2003): 65-74.

Garneau, C. J., and M. B. Parkinson. "Including Preference in Anthropometry-Driven Models for Design." *2007 ASME Design Engineering Technical Conference (DETC)*. Las Vegas, NV, Sept. 4-7, 2007.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Vol. 25, Texts in Statistical Science. Boca Raton, FL: Chapman & Hall/CRC, 2004.

Gershenson, J. K., and L. A. Stauffer. "A Taxonomy for Design Requirements from Corporate Customers." *Research in Engineering Design* 11, no. 2 (Aug. 1999): 103-15.

Goldberg, D. E. *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Berlin: Kluwer Academic Publishers, 2002.

Goos, P. *The Optimal Design of Blocked and Split-Plot Experiments*. New York: Springer-Verlag, 2002.

Goos, P., and M. Vandebroek. "Outperforming Completely Randomized Designs." *Journal of Quality Technology* 36, no. 1 (2004): 12-26.

Green, P. E., and A. M. Krieger. "Alternative Approaches to Cluster-based Market Segmentation." *Journal of the Market Research Society* 37, no. 3 (1995): 221-39.

Green, P. E., A. M. Krieger, and M. K. Agarwal. "Adaptive Conjoint Analysis: Some Caveats and Suggestions." *Journal of Marketing Research* 28, no. 2 (1991): 215-22.

Green, P. E., and V. R. Rao. "Rating Scales and Information Recovery. How Many Scales and Response Categories to Use?" *Journal of Marketing* 34, no. 3 (1970): 33-39.

Green, P. E., and V. Srinivasan. "Conjoint Analysis in Consumer Research: Issues and Outlook." *The Journal of Consumer Research* 5, no. 2 (1978): 103-23.

———. "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice." *Journal of Marketing* 54, no. 4 (1990): 3-19.

Greene, W. H. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall, 2002.

Greenleaf, E. A. "Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles." *Journal of Marketing Research* 29, no. 2 (1992): 176-88.

Gu, X., J. E. Renaud, L. M. Ashe, S. M. Batill, A. S. Budhiraja, and L. J. Krajewski. "Decision-Based Collaborative Optimization." *Journal of Mechanical Design* 124, no. 1 (Mar. 2002): 1-13.

Hamza, K., I. Hossoy, J. F. Reyes-Luna, and P. Y. Papalambros. "Combined Maximisation of Interior Comfort and Frontal Crashworthiness in Preliminary Vehicle Design." *International Journal of Vehicle Design* 35, no. 3 (2004): 167-85.

Hauptmann, P. *Sensors: Principles and Applications*. Upper Saddle River, NJ: Prentice Hall, 1993.

Hazelrigg, G. A. "A Framework for Decision-Based Engineering Design." *Journal of Mechanical Design* 120, no. 4 (Dec. 1998): 653-58.

———. "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design." *Journal of Mechanical Design* 118, no. 2 (Jun. 1996): 161-64.

———. "Validation of Engineering Design Alternative Selection Methods." *Engineering Optimization* 35, no. 2 (Apr. 2003): 103-20.

Hedeker, D., and R. D. Gibbons. "A Random-Effects Ordinal Regression Model for Multilevel Analysis." *Biometrics* 50, no. 4 (1994): 933-44.

Heise, M. A., and R. H. Myers. "Optimal Designs for Bivariate Logistic Regression." *Biometrics* 52, no. 2 (1996): 613-24.

Herrmann, J. W., and L. C. Schmidt. "Viewing Product Development as a Decision Production System." *Proc. 2002 ASME Design Engineering Technical Conf.*, Montreal, Canada, September, 2002.

Hines, Rjoh. "Analysis of Clustered Polytomous Data Using Generalized Estimating Equations and Working Covariance Structures." *Biometrics* 53, no. 4 (1997): 1552-56.

———. "Comparison of Two Covariance Structures in the Analysis of Clustered Polytomous Data using Generalized Estimating Equations." *Biometrics* 54, no. 1 (1998): 312-16.

Hoyle, C., D. Kumar, and W. Chen. "Product Attribute Function Deployment (PAFD) for Decision-Based Conceptual Design." *Proc. of the 2006 ASME Design Engineering Technical Conf.*, Philadelphia, PA, September, 2006.

Ihaka, R., and R. Gentleman. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5 (1996): 299-314.

Jin, R., W. Chen, and T. W. Simpson. "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria." *Structural and Multidisciplinary Optimization* 23, no. 1 (2001): 1-13.

Johnson, R. A., and D. W. Wichern. *Applied Multivariate Statistical Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall, 2002.

Johnson, V. E., and J. H. Albert. *Ordinal Data Modeling*. New York: Springer, 1999.

Kahraman, C., T. Ertay, and G. Büyüközkan. "A Fuzzy Optimization Model for QFD Planning Process using Analytic Network Approach." *European Journal of Operational Research* 171, no. 2 (Jun. 2006): 390-411.

Kanter, R. M. "Innovation: The Classic Traps." *Harvard Business Review* 84, no. 11 (2006): 72-83.

Keeney, R. L., and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Cambridge University Press, 1993.

Kessels, R., P. Goos, and M. Vandebroek. "A Comparison of Criteria to Design Efficient Choice Experiments." *Journal of Marketing Research* 43, no. 3 (2006): 409–19.

———. "Optimal Designs for Rating-based Conjoint Experiments." *Computational Statistics & Data Analysis* 52, no. 5 (2008): 2369-87.

Kim, H. M. "Target Cascading in Optimal System Design." PhD Dissertation, University of Michigan, 2001.

Kim, H. M., M. Kokkolaras, L. S. Louca, G. J. Delagrammatikas, N. F. Michelena, Z. S. Filipi, P. Y. Papalambros, J. L. Stein, and D. N. Assanis. "Target Cascading in Vehicle Redesign: A Class VI Truck Study." *International Journal of Vehicle Design* 29, no. 3 (2002): 199-225.

Kim, H. M., N. F. Michelena, P. Y. Papalambros, and T. Jiang. "Target Cascading in Optimal System Design." *Journal of Mechanical Design (Transactions of the ASME)* 125, no. 3 (2003): 474-80.

Kim, H. M., D. G. Rideout, P. Y. Papalambros, and J. L. Stein. "Analytical Target Cascading in Automotive Vehicle Design." *Journal of Mechanical Design (Transactions of the ASME)* 125, no. 3 (2003): 481-89.

Kim, K. J., H. Moskowitz, A. Dhingra, and G. Evans. "Fuzzy Multicriteria Models for Quality Function Deployment." *European Journal of Operational Research* 121, no. 3 (2000): 504-18.

Koppelman, F., C. Bhat, V. Sethi, and B. Williams. "A Self-Instructing Manual on Discrete Choice Modeling." *Report Submitted to the US Department of Transportation* (2005).

Krishnan, V., and K. T. Ulrich. "Product Development Decisions: A Review of the Literature." *Management Science* 47, no. 1 (Jan. 2001): 1-21.

Kuhfeld, W. F., R. D. Tobias, and M. Garratt. "Efficient Experimental Design with Marketing Research Applications." *Journal of Marketing Research* 31, no. 4 (1994): 545-57.

Kumar, D. "Demand Modeling for Enterprise-Driven Product Design." PhD Dissertation, Northwestern University, 2007.

Kumar, D., C. Hoyle, W. Chen, N. Wang, G. Gomez-Levi, and F. Koppelman. "A Hierarchical Choice Modeling Approach for Incorporating Customer Preferences and Market Trends in Engineering Design." *Proceedings of 2007 ASME DETC/CIE Conference*. Las Vegas NV, Sept. 4-7, 2007.

———. "A Hierarchical Choice Modelling Approach for Incorporating Customer Preferences in Vehicle Package Design." *International Journal of Product Development* 8, no. 3 (2009): 228-51.

Kumar, D., H. Kim, and W. Chen. "Multi-level Optimization for Enterprise Driven Decision-based Product Design." In *Decision Making in Engineering Design*, edited by K. Lewis, W. Chen and L. Schmidt. New York: ASME Press, 2006.

Lancaster, T. *An Introduction to Modern Bayesian Econometrics*. Oxford, UK: Blackwell Publishing, 2004.

Lewis, K., W. Chen, and L Schmidt, Eds. *Decision Making in Engineering Design*. Edited by Eds. New York: ASME Press, 2006.

Li, H., and S. Azarm. "Product Design Selection under Uncertainty and with Competitive Advantage." *Transactions of ASME: Journal of Mechanical Design* 122, no. 4 (2000): 411-18.

Liang, Kyee, and S. L. Zeger. "Longitudinal Data Analysis using Generalized Linear Models." *Biometrika* 73, no. 1 (1986): 13-22.

Locascio, A., and D. L. Thurston. "Transforming the House of Quality to a Multiobjective Optimization Formulation." *Structural and Multidisciplinary Optimization* 16, no. 2 (Oct. 1998): 136-46.

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

Lu, M. "Determinants of Residential Satisfaction: Ordered Logit vs. Regression Models." *Growth and Change* 30, no. 2 (1999): 264-87.

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

MacDonald, E., Gonzalez, R. and Papalambros, P.Y. "Preference Inconsistency in Multidisciplinary Design Decision Making." *Proceedings of the 2007 International Design Engineering Technical Conferences*. Las Vegas, NV, Sept. 4-8, 2007.

Marston, M., J. K. Allen, and F. Mistree. "The Decision Support Problem Technique: Integrating Descriptive and Normative Approaches in Decision Based Design." *Engineering Valuation and Cost Analysis* 3 (Nov. 2000): 107-29.

McAdams, D. A., R. B. Stone, and K. L. Wood. "Functional Interdependence and Product Similarity Based on Customer Needs." *Research in Engineering Design* 11, no. 1 (Apr. 1999): 1-19.

McCullagh, P. "Regression Models for Ordinal Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 42, no. 2 (1980): 109-42.

McCutcheon, A. L. *Latent Class Analysis*. Beverly Hills, CA: Sage Publications, 1987.

McKelvey, R. D., and W. Zavoina. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4, no. 1 (1975): 103-20.

Michalek, J. J. "Preference Coordination in Engineering Design Decision-Making." PhD. Dissertation, University of Michigan, 2005.

Michalek, J. J., F. M. Feinberg, and P. Y. Papalambros. "Linking Marketing and Engineering Product Design Decisions via Analytical Target Cascading." *Journal of Product Innovation Management* 22, no. 1 (2005): 42-62.

Michelena, N., H. M. Kim, and P. Y. Papalambros. "A System Partitioning and Optimization Approach to Target Cascading." *Proceedings of the 12th International Conference on Engineering Design*. Munich, Germany, 1999.

Montgomery, D. C. *Design and Analysis of Experiments*. New York: John Wiley and Sons, Inc., 2005.

Myers, R. H., and D. C. Montgomery. *Response Surface Methodology: Process and Product in Optimization Using Designed Experiments*. New York, NY: John Wiley & Sons, Inc., 2002.

Neelamegham, R., and P. Chintagunta. "A Bayesian Model to Forecast New Product Performance in Domestic and International Markets." *Marketing Science* 18, no. 2 (1999): 115-36.

Noui-Mehidi, Ali. *Applying Constraints to Vehicle Packaging*. Customer Papers, ILOG Corporation, 1997.

Olewnik, A. T., and K. Lewis. "On Validating Engineering Design Decision Support Tools." *Concurrent Engineering* 13, no. 2 (Jun. 2005): 111-22.

Otto, K. N., and K. L. Wood. *Product Design: Techniques in Reverse Engineering and New Product Development*. New York: Prentice Hall, 2001.

Parkinson, Mathew P., and Mathew P. Reed. "Optimizing Vehicle Occupant Packaging." *SAE World Congress* Detroit, Michigan, April 3-6, 2006.

Perevozskaya, I., W. F. Rosenberger, and L. M. Haines. "Optimal Design for the Proportional Odds Model." *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* 31, no. 2 (2003): 225-35.

Petiot, J. F., and B. Yannou. "Measuring Consumer Perceptions for a Better Comprehension, Specification and Assessment of Product Semantics." *International Journal of Industrial Ergonomics* 33, no. 6 (2004): 507-25.

Phadke, M. S. *Quality Engineering Using Robust Design*. Upper Saddle River, NJ: Prentice Hall, 1995.

Pine, B. J. *Mass Customization: The New Frontier in Business Competition*. Boston: Harvard Business School Press, 1993.

Prasad, B. "A Concurrent Function Deployment Technique for a Workgroup-Based Engineering Design Process." *Journal of Engineering Design* 11, no. 2 (Jun. 2000): 103-19.

Reed, M. P., M. A. Manary, C. A. Flannagan, and L. W. Schneider. "Effects of Vehicle interior Geometry and Anthropometric Variables on Automobile Driving Posture." *Human Factors* 42, no. 4 (2000): 541-52.

Reed, Mathew P., Miriam A. Manary, Carol A.C. Flannagan, and Lawrence W. Schneider. "A Statistical Method for Predicting Automobile Driving Posture." *The Journal of the Human Factors and Ergonomics Society* 44, no. 4 (Winter 2002): 557-68.

Rossi, P. E., and G. M. Allenby. "Bayesian Statistics and Marketing." *Marketing Science* 22, no. 3 (2003): 304-28.

Rossi, P. E., G. M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. Hoboken, NJ: John Wiley and Sons, Ltd., 2005.

Rossi, P. E., Z. Gilula, and G. M. Allenby. "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach." *Journal of the American Statistical Association* 96, no. 453 (2001): 20-31.

Sandor, Z., and M. Wedel. "Designing Conjoint Choice Experiments Using Managers' Prior Beliefs." *Journal of Marketing Research* 38, no. 4 (2001): 430-44.

Sawtooth Software. "The CBC/HB Module for Hierarchical Bayes." *at* www.sawtoothsoftware.com (1999).

Shah, J. J., and P. K. Wright. "Developing Theoretical Foundations of DfM." *Proceedings of 2000 ASME Design Engineering Technical Conf.*, Baltimore, MD, September, 2000.

Siddall, J. N. *Optimal Engineering Design: Principles and Applications*. New York: Marcel Dekker, 1982.

Simpson, T. W., J. D. Poplinski, P. N. Koch, and J. K. Allen. "Metamodels for Computer-Based Engineering Design: Survey and Recommendations." *Engineering with Computers* 17, no. 2 (2001): 129-50.

Small, K. "Fundamentals of Economic Demand Modeling: Lessons from Travel Demand Analysis." In *Decision Making in Engineering Design*, edited by K. Lewis, W. Chen and L Schmidt. New York: ASME Press, 2006.

Society of Automotive Engineers. *Surface Vehicle Recommended Practice-Motor Vehicle Dimensions*. Warrendale, Pennsylvania, 2002.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. *WinBUGS Version 1.4. MRC Biostatistics Unit, Cambridge, UK*. 2003.

Stata Corporation. *StataSE 9.2*. College Station, Texas 1996-2008.

Steckel, J. H., W. S. DeSarbo, and V. Mahajan. "On the Creation of Acceptable Conjoint Analysis Experimental Designs." *Decision Sciences* 22, no. 2 (1991): 435-42.

Stevens, S. S. *Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects*. New Brunswick, N.J.: Transaction Publishers, 1986.

Stone, R. B., and K. L. Wood. "Development of a Functional Basis for Design." *Journal of Mechanical Design* 122, no. 4 (Dec. 2000): 359-70.

Suh, N. P. *The Principles of Design*. New York: Oxford University Press, 1990.

Tamhane, A. C., B. E. Ankenman, and Y. Yang. "The Beta Distribution as a Latent Response Model for Ordinal Data (I): Estimation of Location and Dispersion Parameters." *Journal of Statistical Computation and Simulation* 72, no. 6 (2002): 473-94.

Tamhane, A. C., and D. D. Dunlop. *Statistics and Data Analysis: From Elementary to Intermediate*. Upper Saddle River, NJ: Prentice Hall, 2000.

Terninko, J. *Step-by-Step QFD: Customer-Driven Product Design*. Boca Raton, FL: CRC Press, 1997.

Train, K. E. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press, 2003.

Train, K., and D. McFadden. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* 15, no. 5 (2000): 447-70.

Ullman, D. G. *The Mechanical Design Process*. Boston, MA: McGraw-Hill, 2002.

van de Poel, I. "Methodological Problems in QFD and Directions for Future Development." *Research in Engineering Design* 18, no. 1 (May 2007): 21-36.

Varian, H. R. "The Information Economy." *Scientific American* 273, no. 3 (1995): 200-02.

Wang, N., V. Kiridena, G. Gomez-Levi, and J. Wan. "Design and Verification of a New Computer Controlled Seating Buck." *Proceedings of the 2006 ASME IDETC/CIE*. Philadelphia, Pennsylvania, September 10-13, 2006.

Wassenaar, H. J., and W. Chen. "An Approach to Decision-Based Design with Discrete Choice Analysis for Demand Modeling." *Transactions of the ASME: Journal of Mechanical Design* 125, no. 3 (2003): 490-97.

Wassenaar, H. J., W. Chen, J. Cheng, and A. Sudjianto. "Enhancing Discrete Choice Demand Modeling for Decision-Based Design." *Journal of Mechanical Design* 127, no. 4 (2005): 514-23.

———. "An Integrated Latent Variable Choice Modeling Approach to Enhancing Product Demand Modeling." *Proceedings of the 2004 ASME DETC/CIE*. Salt Lake City, Utah, 2004.

Wassenaar, H. J., D. Kumar, and W. Chen. "Discrete Choice Demand Modeling for Decision-Based Design." In *Decision Making in Engineering Design*, edited by K. Lewis, W. Chen and L. Schmidt. New York: ASME Press, 2006.

Williamson, J. M., K. Kim, and S. R. Lipsitz. "Analyzing Bivariate Ordinal Data Using a Global Odds Ratio." *Journal of the American Statistical Association* 90, no. 432 (1995): 1432 - 37.

Witten, I. H., and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA: Morgan Kaufmann, 2005.

Wood, S. N. *mgcv: GAMs with GCV Smoothness Estimation and GAMMs by REML/PQL.* Vienna, Austria: R Foundation for Statistical Computing, 2004.

Zocchi, S. S., and A. C. Atkinson. "Optimum Experimental Designs for Multinomial Logistic Models." *Biometrics* 55, no. 2 (1999): 437-44.

Zorn, C. J. W. "Generalized Estimating Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45, no. 2 (2001): 470-90.

Appendix A: Choice Set and Analytic Relationships for PAFD Example

Table A.1: Sample of Choice Set Used for Estimation of DCA Model

	Demographic S		
Customer	*Region*	*Market Segment*	**Purchase**
Customer 1	N. America	$30,000.00	***Our***
Customer 2	N. America	$29,000.00	***Our***
Customer 3	Asia	$22,000.00	***Sensor A***
Customer 4	N. America	$24,000.00	***Sensor B***
Customer 5	N. America.	$24,500.00	***Sensor B***
Customer 6	Asia	$34,000.00	***Sensor C***

Table A.2: Analytical Relationships between E and X

Engineering Attribute E	***Concept 1*** **E as a function of X^1**	***Concept 2*** **E as a function of X^2**
Sense element accuracy	ε(calibration) + ε(A/D)	ε(calibration) + ε(A/D)
Full scale span	$k*\Delta l/l$	$\varepsilon_0 \varepsilon_r A/\Delta d$
Temperature range	*Min*[T_{max}(IC), T_g(Housing)]	*Min*[T_{max}(IC), T_g(Housing)]
Housing footprint	Housing *width*length*	Housing *width*length*
Natural frequency	$Cn_j\sqrt{EI/\rho AL^4}$	$Cn_j\sqrt{EI/\rho AL^4}$
Connector mating force	25, 35 , 40	25, 35 , 40

Appendix B: Information Matrix Computation for Algorithmic Implementation

This appendix provides a method for expressing the information matrix, **M**, and estimating the prediction variance of a given extended design point, **f(x)**, for use in the optimization algorithm. The ordinal data GLM information matrix of Eq. (4.9) can be written in analogous fashion to the GLS formulation of Eq. (4.7) (Johnson and Albert, 1999). An $\mathbf{H}_n$ matrix is defined as a matrix of derivatives of the logistic CDF as $\mathbf{H}_n = \mathrm{diag}\left(f_{n1}, f_{n2}, \ldots f_{n(P-1)}\right)$. The extended design point $\mathbf{f}(\mathbf{x}_{in})$ for a given respondent and given configuration is defined as:

$$\mathbf{f}\left(\mathbf{x}_{in}\right) = \begin{bmatrix} 1 & 0 & \cdots & 0 & -\mathbf{x}_{in} \\ 0 & 1 & \cdots & 0 & -\mathbf{x}_{in} \\ \vdots & \vdots & \ddots & \vdots & -\mathbf{x}_{in} \\ 0 & 0 & 0 & 1 & -\mathbf{x}_{in} \end{bmatrix}. \tag{B.1}$$

A $\mathbf{C}_n$ matrix defined as:

$$\mathbf{C}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & -1 \end{bmatrix}. \tag{B.2}$$

With $\mathbf{C}_n$, $\mathbf{H}_n$ and **f(x)** defined, the information matrix can be written as (Johnson and Albert, 1999):

$$\mathbf{M} = \sum_{n=1}^{N} \mathbf{F}_n' \mathbf{W}_n^{-1} \mathbf{F}_n \tag{B.3}$$

where $\mathbf{W}_n^{-1} = \mathbf{H}_n \mathbf{C}_n' \mathbf{V}_n^{-1} \mathbf{C}_n \mathbf{H}_n$, and **F** is the extended design matrix composed of the **f(x)**.

Appendix C: Example of 3 Experimental Configuration Blocks

Config	L38	W35	H61	Gender	Stature	Resp.
1	-1	-1	-1	-1	-0.33	1
2	-1	-1	1	-1	-0.33	1
3	-1	0	-0.33	-1	-0.33	1
4	-1	0	1	-1	-0.33	1
5	-1	1	-1	-1	-0.33	1
6	-1	1	0.33	-1	-0.33	1
7	-1	1	1	-1	-0.33	1
8	0	-1	-1	-1	-0.33	1
9	0	-1	1	-1	-0.33	1
10	0	0	1	-1	-0.33	1
11	0	1	-1	-1	-0.33	1
12	1	-1	-1	-1	-0.33	1
13	1	-1	-1	-1	-0.33	1
14	1	-1	0.33	-1	-0.33	1
15	1	0	1	-1	-0.33	1
16	1	1	-1	-1	-0.33	1
17	1	1	0.33	-1	-0.33	1
18	1	1	1	-1	-0.33	1
1	-1	-1	-1	1	0.33	2
2	-1	-1	-1	1	0.33	2
3	-1	-1	1	1	0.33	2
4	-1	0	1	1	0.33	2
5	-1	1	-0.33	1	0.33	2
6	-1	1	-0.33	1	0.33	2
7	-1	1	1	1	0.33	2
8	0	-1	-1	1	0.33	2
9	0	0	-1	1	0.33	2
10	0	0	-0.33	1	0.33	2
11	0	1	1	1	0.33	2
12	1	-1	-0.33	1	0.33	2
13	1	-1	1	1	0.33	2
14	1	-1	1	1	0.33	2
15	1	0	1	1	0.33	2
16	1	1	-1	1	0.33	2
17	1	1	-1	1	0.33	2
18	1	1	1	1	0.33	2
1	-1	-1	-1	1	1	3
2	-1	-1	-1	1	1	3
3	-1	-1	1	1	1	3
4	-1	-1	1	1	1	3
5	-1	1	-1	1	1	3
6	-1	1	-1	1	1	3
7	-1	1	1	1	1	3
8	-1	1	1	1	1	3
9	0	1	-0.33	1	1	3
10	0	1	1	1	1	3
11	1	-1	-1	1	1	3
12	1	-1	-1	1	1	3
13	1	-1	1	1	1	3
14	1	-1	1	1	1	3
15	1	0	-1	1	1	3
16	1	1	-1	1	1	3
17	1	1	0.33	1	1	3
18	1	1	1	1	1	3

Appendix D: PVM 4 Block Human Appraisal Designs

Table D.1: Full 1-Part 4 Block Experiment

x_1 SgRP to Hinge (PtoH)	x_2 SgRP to Rocker Y (ROK_Y)	x_3 SgRP to Heel Z (HEL_Z)	x_4 SgRP to Ground Z (GRD_Z)	x_5 Sill to Heel (StoH)	x_6 SgRP to Roof Z (HR_Z)	x_7 SgRP to Frt Hdr X (HR_X)	x_8 SgRP to Side Rail Y (HR_Y)	Block
800	380	288	450	70	777	241	122	1
700	520	400	625	140	777	241	122	1
800	520	175	625	140	977	241	122	1
725	380	175	800	70	777	366	122	1
800	380	400	800	0	977	366	122	1
800	520	175	625	0	777	491	122	1
800	380	400	625	140	877	491	122	1
700	380	175	800	0	977	491	122	1
700	520	175	625	0	777	241	197	1
700	380	175	625	140	777	241	272	1
725	520	175	625	70	877	241	272	1
700	520	400	625	0	977	241	272	1
725	380	400	625	140	977	241	272	1
800	520	400	625	140	777	366	272	1
700	380	400	625	0	777	491	272	1
800	380	175	800	0	777	491	272	1
800	380	175	450	70	977	491	272	1
700	520	288	450	140	977	491	272	1
800	380	175	625	0	777	241	122	2
800	380	400	625	140	777	241	122	2
700	380	175	450	140	977	241	122	2
700	520	400	800	0	877	366	122	2
800	450	288	450	70	877	366	122	2
700	380	175	450	140	777	491	122	2
725	380	400	625	0	977	491	122	2
800	520	400	625	70	977	491	122	2
800	380	175	800	70	977	491	122	2
725	450	288	625	140	977	491	122	2
700	380	175	800	0	777	241	272	2
700	520	288	450	70	777	241	272	2
800	380	175	450	140	777	241	272	2
800	520	175	625	140	777	241	272	2
800	520	288	800	0	977	241	272	2
800	520	400	800	0	777	491	272	2
700	520	175	625	0	977	491	272	2
700	380	400	625	140	977	491	272	2

700	520	400	625	0	777	241	122	3
800	520	175	450	140	777	241	122	3
800	380	175	800	0	877	241	122	3
800	380	288	450	70	977	241	122	3
700	380	400	625	140	777	366	122	3
800	520	400	625	140	977	366	122	3
800	380	400	625	0	777	491	122	3
700	520	175	450	70	977	491	122	3
700	520	400	800	70	777	491	197	3
800	380	400	800	70	777	241	272	3
700	380	400	625	0	977	241	272	3
700	520	175	625	140	977	241	272	3
800	520	175	625	0	777	366	272	3
700	450	175	450	140	777	491	272	3
700	380	288	625	0	877	491	272	3
800	450	400	625	0	977	491	272	3
700	380	175	800	70	977	491	272	3
800	520	175	625	140	977	491	272	3
700	380	175	450	70	777	241	122	4
800	520	288	800	70	777	241	122	4
725	520	175	625	0	977	241	122	4
700	380	400	800	70	977	241	122	4
700	520	400	625	0	777	491	122	4
700	380	175	800	0	777	491	122	4
700	520	175	625	140	877	491	122	4
800	380	175	450	140	977	491	122	4
800	520	400	625	0	877	241	197	4
700	450	288	625	70	977	366	197	4
700	520	400	800	0	777	241	272	4
800	380	400	625	0	977	241	272	4
800	380	175	800	70	977	241	272	4
800	520	175	450	140	977	241	272	4
800	520	175	450	70	777	491	272	4
725	380	400	625	140	777	491	272	4
700	380	288	450	70	977	491	272	4
800	520	400	800	70	977	491	272	4

Table D.2: Blocks 3 and 4 to be Augmented in 2-Part Experiment

x_1 SgRP to Hinge (PtoH)	x_2 SgRP to Rocker Y (ROK_Y)	x_3 SgRP to Heel Z (HEL_Z)	x_4 SgRP to Ground Z (GRD_Z)	x_5 Sill to Heel (StoH)	x_6 SgRP to Roof Z (HR_Z)	x_7 SgRP to Frt Hdr X (HR_X)	x_8 SgRP to Side Rail Y (HR_Y)	**Block**
700	520	400	625	0	777	241	122	3
800	520	288	800	70	777	241	122	3
700	520	400	625	140	777	241	122	3
800	380	175	800	0	877	241	122	3
800	380	288	450	70	977	241	122	3
700	380	175	450	140	977	241	122	3
800	520	175	625	0	777	491	122	3
700	380	175	450	140	777	491	122	3
700	380	175	800	0	977	491	122	3
800	520	400	625	70	977	491	122	3
700	520	288	450	70	777	241	272	3
700	380	400	625	0	977	241	272	3
800	520	288	800	0	977	241	272	3
725	380	400	625	140	977	241	272	3
700	380	400	625	0	777	491	272	3
800	380	175	800	0	777	491	272	3
700	380	288	450	70	977	491	272	3
800	520	175	625	140	977	491	272	3
700	380	175	450	70	777	241	122	4
800	520	175	450	140	777	241	122	4
725	520	175	625	0	977	241	122	4
700	380	400	800	70	977	241	122	4
725	380	175	800	70	777	366	122	4
800	380	400	625	0	777	491	122	4
700	520	175	625	140	877	491	122	4
725	380	400	625	0	977	491	122	4
800	380	175	450	140	977	491	122	4
700	450	288	625	70	977	366	197	4
700	520	400	800	0	777	241	272	4
800	380	175	450	140	777	241	272	4
800	520	175	625	140	777	241	272	4
800	380	400	625	0	977	241	272	4
700	520	175	625	140	977	241	272	4
725	380	400	625	140	777	491	272	4
700	380	175	800	70	977	491	272	4
800	520	400	800	70	977	491	272	4

Appendix E: PVM Investigation Questionnaire

PVM Roominess and Ingress/Egress Experimental Protocol
Northwestern URP
March 2008

Setup (for the test administrator):

At the beginning of each day of experiments:

1. Setup the video camera to capture motion.
2. Determine the set of configurations and the configuration order for each respondent.
3. Ensure all platforms needed for the respondents to be tested are available and close to the PVM.
4. Install A-pillar, B-pillar, armrest
5. Have hinge pillar and sill inserts handy
6. Ensure scale, chair and meter are ready to get anthropometric dimensions from subjects
7. Ensure seat motion is turned on

For each respondent:

1. Place the PVM in configuration 1 for the given respondent.
2. Place the driver's seat in a "neutral" position (lowest and rearmost)
3. Ensure the armrest is in the correct position for configuration 1
4. Read the "**Description of Experiments**" to the respondent.
5. Record the respondent's demographic attributes: gender, age, and current vehicle ownership.
6. Measure and record the respondent's anthropomorphic dimensions: height, seated height, weight, shoe size, heel height.
7. Respondent enters the PVM as a practice for the experiments.
8. Once the respondent is in the vehicle, read the complete Ingress questions to get the respondent familiar with the questions
9. The respondent adjusts the driver's seat position in the **vertical** (**z**) and **frontal** (**x**) positions and the seat back **angle** (α) to attain a comfortable driving position. Record the seat position.
10. Close the door and read the complete Roominess questions to get the respondent familiar with the questions
11. Open the door and let the respondent exit the PVM. Read the complete Egress questions to get the respondent familiar with the questions

Description of Experiments (read to the respondent):

You will be evaluating 18 different vehicle configurations for ingress, egress and roominess of the driver's compartment.

Before you begin the experiments, I will record the age group you belong to, the current vehicle you drive and will measure your height, seated height, weight, shoe size and heel height.

You will then be asked to enter the PVM to practice the procedure, as well as to adjust the driver's seat to your preferred position in the vertical and frontal directions, and the seat back angle. The seat adjustment will be done this one time only for the rest of the experiment. Once you enter the vehicle for the practice run I will read the experiment questions regarding Ingress so that you get familiar with them. You can then adjust the driver's seat and I will record the adjustment. I will then close the door and read the questions regarding Roominess. I will then open the door for you and once you exit the vehicle, I will read the questions regarding Egress. All your responses will be given as ratings which are posted in front of the vehicle. In each case I will indicate which of the three scales we are going to be using.

The experiment procedure is as follows.

1. Enter the vehicle on the driver's side when instructed by the test administrator, but do not close the door.
2. Evaluate the ease of entering the vehicle.
3. Wait for the test administrator to close the door.
4. Evaluate the driver's compartment amount of head room, the room to the left of the driver seating area, the knee room in the driver's compartment and the overall roominess.
5. I will open the door for you and you may exit the vehicle.
6. Evaluate the ease of exiting the vehicle.
7. We may repeat these steps as necessary to ensure the evaluation is accurate.

In few cases, I will need to make adjustments to the vehicle dimensions before evaluating roominess, in which case you will need to evaluate ingress and egress first and then reenter the vehicle to evaluate roominess only. I will let you know when this is the case.

You may now enter the vehicle. Please adjust the seat to a comfortable position.

[wait for subject to adjust seat and record adjustment]

These will be the ingress questions:

1. How acceptable is this vehicle configuration for ingress? This is rated on a 1 to 4 scale with the following definition for each rating as you can see posted in front of the vehicle: 1 is "very unacceptable", 2 is "somewhat unacceptable", 3 is "somewhat acceptable" and 4 is "very acceptable".

Very unacceptable	**Somewhat unacceptable**	**Somewhat acceptable**	**Very acceptable**
1	**2**	**3**	**4**

2. What is the overall ease of ingress, for the vehicle? This includes evaluation of stepping up and passing through the door opening. This question is rated on a 1 to 5 scale, again as you can see posted in front of the vehicle, with the following definition for each rating: 1 is “very strong effort”, 2 is “strong effort”, 3 is “moderate effort”, 4 is “weak effort”, and 5 is “no effort at all”.

Very strong effort	Strong effort	Moderate effort	Weak effort	No effort at all
1	**2**	**3**	**4**	**5**

3. How would you rate the space available for ingress? This includes evaluation of the step-up height and the size of the door opening. This is rated on a 1 to 5 scale, with the following definition for each rating: 1 is “very insufficient”, 2 is “insufficient”, 3 is “barely sufficient”, 4 is “sufficient”, and 5 is “excellent”.

Very insufficient	Insufficient	Barely sufficient	Sufficient	Excellent
1	**2**	**3**	**4**	**5**

I will now close the door as will be done when you evaluate roominess.

[Evaluator closes the door]

The questions you will be answering to evaluate the roominess are as follows. All questions are rated on a 1 to 5 scale, with the following definition for each rating: 1 is “very insufficient”, 2 is “insufficient”, 3 is “barely sufficient”, 4 is “sufficient”, and 5 is “excellent”, as posted in front of the vehicle. The questions are:

1. How do you rate the amount of headroom? This includes space above, to the left side, and in front of your head.
2. How do you rate the overall room to the left of the driver seating area? This includes the space between your shoulder, upper arm, hips, elbows, and the left side of the vehicle.
3. How do you rate the amount of knee room? This includes the space to the left, and in front of your knees.
4. How do you rate the overall roominess of the driver’s compartment? This includes to the left, in front, and above you.

Very insufficient	Insufficient	Barely sufficient	Sufficient	Excellent
1	**2**	**3**	**4**	**5**

[Open the door]

You may now exit the vehicle.

These will be the egress questions:

1. How acceptable is this vehicle configuration for egress? This is rated on a 1 to 4 scale with the following definition for each rating as you can see posted in front of the vehicle: 1 is "very unacceptable", 2 is "somewhat unacceptable", 3 is "somewhat acceptable" and 4 is "very acceptable".

Very unacceptable	Somewhat unacceptable	Somewhat acceptable	Very acceptable
1	2	3	4

2. What is the overall ease of egress for the vehicle? This includes evaluation of stepping up and passing through the door opening. This question is rated on a 1 to 5 scale, again as you can see posted in front of the vehicle, with the following definition for each rating: 1 is "very strong effort", 2 is "strong effort", 3 is "moderate effort", 4 is "weak effort", and 5 is "no effort at all".

Very strong effort	Strong effort	Moderate effort	Weak effort	No effort at all
1	2	3	4	5

3. How would you rate the space available for or egress? This includes evaluation of the step-up height and the size of the door opening. This is rated on a 1 to 5 scale, with the following definition for each rating: 1 is "very insufficient", 2 is "insufficient", 3 is "barely sufficient", 4 is "sufficient", and 5 is "excellent".

Very insufficient	Insufficient	Barely sufficient	Sufficient	Excellent
1	2	3	4	5

Now that you have completed the practice run, we will begin the experiments.

Appendix F: Sample Respondent Data

Name: S34

Recorded Anthropomorphic and Demographic Information

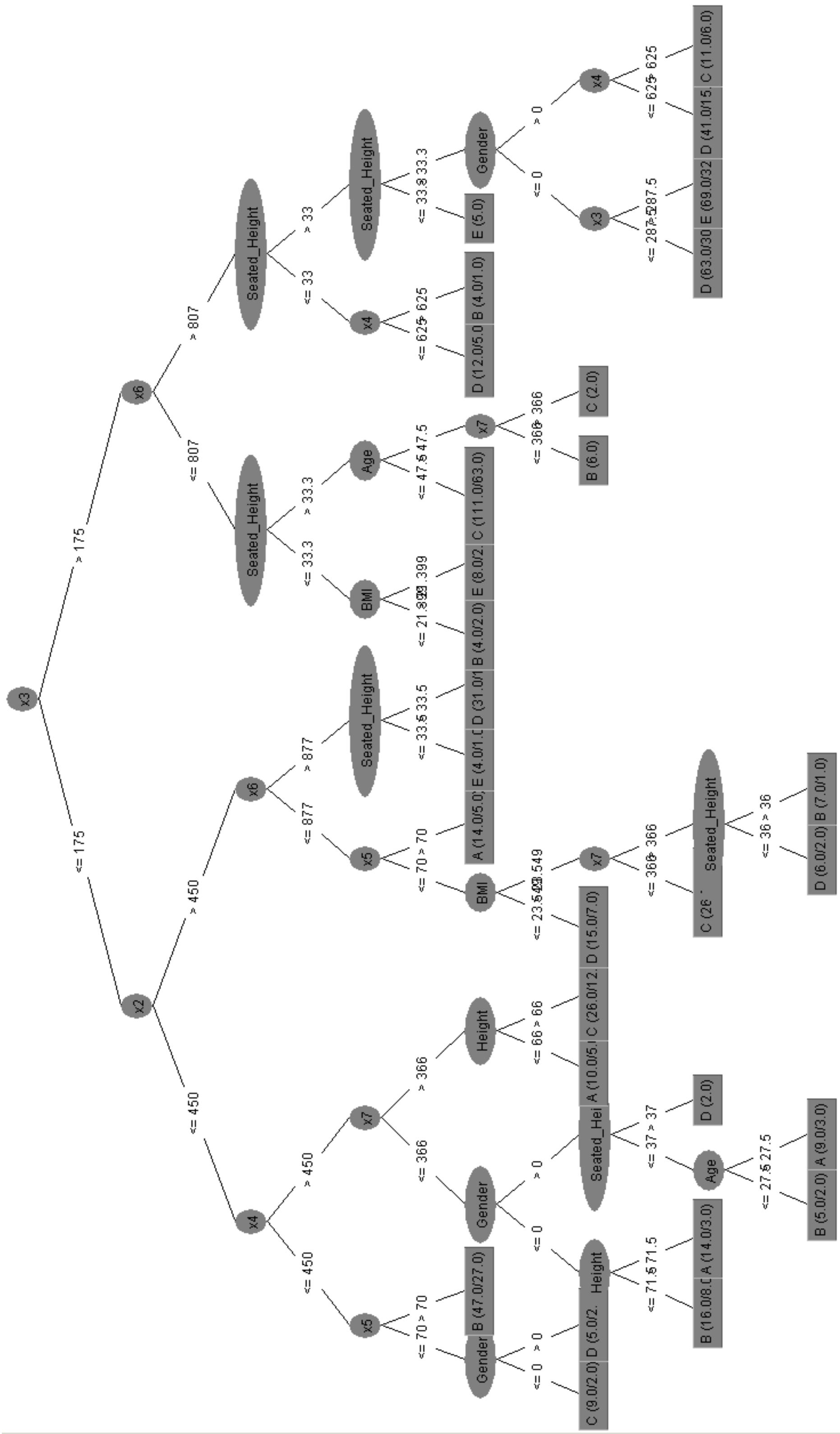
Gender	Height	Seated Height	Weight	Shoe Size	Heel Height	Age*	Current Vehicle	BMI
	in.	in.	lb	in.	in.	Category		
F	66.5	35.8	121	10.25	2.0	3	Taurus X	19.21

Seat adjustment
U/D
F/A 70
Angle 22

Recorded Ratings

#		Ingress			Interior				Egress				
		Acceptable	Ease/effort	Space	Headroom	Left Room	Knee Room	Roominess	Acceptable	Ease/effort	Space	Platform	Inserts
1	6000	1	1	1	1	4	4	2	1	1	1	175	**Move belt 105mm down for Roominess test** Armrest Mark: Top 22mm Sill
2	1504	1	1	1	1	4	4	1	2	2	2	175	Armrest Mark: Top
3	5317	1	2	1	1	5	5	2	1	2	1	175	Armrest Mark: Top
4	1512	2	2	2	1	5	5	2	2	3	2	175	Armrest Mark: Bottom 95mm Front Clip
5	1909	3	3	3	3	4	4	3	3	4	4	175	Armrest Mark: Top
6	2482	2	3	2	2	5	5	3	3	3	3	175	Armrest Mark: Top
7	651	2	2	2	5	4	4	4	1	1	1	175	**Move belt 105mm down for Roominess test** Armrest Mark: Top 22mm Sill
8	1429	3	4	4	5	5	5	5	2	3	3	175	Armrest Mark: Top
9	6528	4	4	4	5	4	4	4	4	4	4	175	Armrest Mark: Top
10	6400	1	2	1	4	3	3	2	2	2	2	175	Armrest Mark: Bottom
11	6498	3	4	4	5	5	5	5	2	2	3	200	Armrest Mark: Bottom
12	642	2	3	2	5	4	4	4	2	2	1	0	Armrest Mark: Bottom 95mm Front Clip
13	4930	1	1	1	4	4	4	4	1	1	1	0	Armrest Mark: Top
14	4528	1	1	1	1	3	3	1	1	1	1	0	Armrest Mark: Top
15	4431	1	1	1	1	4	4	1	1	1	1	0	Armrest Mark: Bottom
16	28	1	1	1	1	4	4	1	2	2	2	175	Armrest Mark: Bottom
17	1899	1	1	1	3	4	4	3	1	1	1	175	**Move belt 105mm down for Roominess test** Armrest Mark: Top 22mm Sill
18	2147	3	4	4	5	4	4	4	2	3	2	175	Armrest Mark: Bottom

Appendix G: C4.5 Decision Tree for Ingress Response



Appendix H: Full M2 and M3 Models with Cut points

	Fixed SEP		Fixed INT		Random SEP		Random INT	
M_3: Headroom								
	coef.	*t-value*	coef.	*t-value*	coef.	*t-value*	coef.	*t-value*
ROK_Y	**1.53**	*8.26*	**1.50**	*8.10*	**1.55**	*8.53*	**1.60**	*6.74*
GRD_Z	**-0.19**	*-0.97*	**-0.18**	*-0.89*	**0.11**	*0.54*	**0.09**	*0.46*
HR_Z	**4.02**	*18.80*	**3.90**	*16.81*	**4.64**	*19.91*	**4.67**	*16.05*
HR_Y	**1.36**	*9.01*	**1.29**	*8.50*	**1.67**	*10.43*	**1.67**	*10.97*
Stat	**-2.16**	*-6.84*	**-2.20**	*-7.23*	**-3.27**	*-8.00*	**-3.13**	*-7.92*
BMI	**1.46**	*4.84*	**1.53**	*5.23*	**1.76**	*4.56*	**1.62**	*4.43*
Age	**-0.84**	*-3.47*	**-0.87**	*-2.71*	**-1.18**	*-3.72*	**-1.07**	*-3.67*
cut 1	-4.97	-7.01	-5.03	-6.78	-6.73	-8.55	-6.61	-9.13
cut 2	-3.99	-8.19	-4.08	-7.12	-5.75	-9.68	-5.63	-10.04
cut 3	-2.64	-7.86	-2.73	-5.94	-4.33	-9.38	-4.23	-9.48
cut 4	-1.97	-6.53	-2.06	-4.80	-3.56	-8.32	-3.45	-8.24
cut 5	-1.08	-3.84	-1.17	-2.88	-2.43	-6.00	-2.30	-5.87
cut 6	-0.05	-0.17	-0.14	-0.36	-0.97	-2.45	-0.85	-2.24
cut 7	1.18	4.17	1.07	2.68	0.79	1.99	0.91	2.33
cut 8	2.38	8.30	2.26	5.59	2.55	6.13	2.71	6.59
M_3: Leftroom								
ROK_Y	**3.48**	*16.83*	**3.45**	*18.27*	**3.54**	*17.16*	**3.51**	*18.16*
GRD_Z	**-1.53**	*-8.24*	**-1.64**	*-9.77*	**-1.44**	*-7.67*	**-1.37**	*-8.00*
StoH	**0.28**	*1.63*	**0.25**	*1.59*	**0.27**	*1.55*	**0.29**	*1.51*
HR_Z	**1.72**	*10.30*	**1.63**	*9.65*	**1.71**	*10.59*	**1.69**	*9.87*
Stat	**-2.17**	*-6.88*	**-2.13**	*-6.15*	**-3.17**	*-8.07*	**-3.07**	*-8.62*
BMI	**-1.21**	*-4.33*	**-1.08**	*-3.90*	**-1.87**	*-5.29*	**-1.78**	*-5.07*
Age	**-0.36**	*-1.44*	**-0.32**	*-1.22*	**-0.63**	*-2.03*	**-0.30**	*-0.92*
cut 1	-6.19	-8.71	-6.32	-9.02	-9.07	-11.59	-8.74	-10.69
cut 2	-5.40	-9.96	-5.54	-9.56	-8.26	-12.89	-7.96	-11.60
cut 3	-4.40	-10.77	-4.48	-9.97	-7.15	-13.52	-6.86	-12.02
cut 4	-3.40	-9.99	-3.45	-8.95	-6.02	-12.80	-5.70	-11.42
cut 5	-2.08	-6.78	-2.12	-5.98	-4.25	-9.90	-3.91	-8.60
cut 6	-0.92	-3.08	-0.96	-2.79	-2.37	-5.76	-2.02	-4.85
cut 7	0.41	1.39	0.38	1.13	-0.07	-0.18	0.30	0.74
cut 8	1.42	4.71	1.38	4.11	1.70	3.94	2.10	5.27
M_3: Kneeroom								
ROK_Y	**1.38**	*9.62*	**1.44**	*8.88*	**1.36**	*8.98*	**1.36**	*8.55*
HEL_Z	**1.01**	*6.70*	**1.03**	*6.62*	**1.05**	*6.93*	**1.04**	*6.80*
StoH	**-0.36**	*-2.22*	**-0.34**	*-2.32*	**-0.34**	*-2.06*	**-0.38**	*-2.24*
HR_Z	**0.81**	*5.04*	**0.86**	*5.51*	**0.84**	*5.06*	**0.88**	*4.82*
Stat	**-1.40**	*-4.52*	**-1.38**	*-4.41*	**-1.02**	*-2.37*	**-1.67**	*-4.25*
BMI	**-1.01**	*-3.55*	**-1.06**	*-4.19*	**-1.42**	*-4.13*	**-1.42**	*-4.25*
Age	**1.02**	*4.13*	**0.92**	*3.56*	**1.70**	*5.15*	**1.08**	*3.50*

cut 1	-4.95	-7.29	-4.97	-7.34	-5.99	-7.72	-6.86	-9.00
cut 2	-3.63	-8.30	-3.62	-8.15	-4.59	-8.30	-5.44	-10.21
cut 3	-2.69	-7.83	-2.68	-7.60	-3.56	-7.54	-4.41	-9.35
cut 4	-1.74	-5.96	-1.74	-5.88	-2.46	-5.66	-3.29	-7.29
cut 5	-0.67	-2.45	-0.66	-2.38	-1.05	-2.51	-1.87	-4.41
cut 6	0.47	1.77	0.48	1.82	0.71	1.72	-0.11	-0.26
cut 7	1.67	6.17	1.68	6.42	2.67	6.39	1.86	4.38
cut 8	2.58	9.19	2.60	9.41	4.22	9.75	3.44	7.62

M_2: Roominess

Head	**0.38**	*6.93*	**0.27**	*2.90*	**0.52**	*9.53*	**0.21**	*2.02*
Left	**0.48**	*6.50*	**0.39**	*2.50*	**0.66**	*7.63*	**0.42**	*1.32*
Knee	**0.59**	*8.00*	**-0.18**	*-1.15*	**0.58**	*7.47*	**-0.53**	*-1.23*
cut 1	2.03	2.46	1.46	-5.66	2.84	1.90	-9.30	-7.05
cut 2	4.26	5.77	5.77	-8.92	5.35	6.15	-7.23	-10.34
cut 3	5.23	8.58	8.58	-9.72	6.45	8.44	-6.32	-11.17
cut 4	6.92	12.37	12.37	-10.27	8.46	11.89	-4.87	-10.37
cut 5	8.76	15.45	15.45	-7.10	10.70	14.69	-3.10	-7.26
cut 6	10.73	17.55	17.55	-1.97	13.08	16.81	-1.16	-2.78
cut 7	12.93	19.33	19.33	4.10	15.81	18.58	1.12	2.60
cut 8	14.68	20.73	20.73	8.68	17.97	19.72	3.15	6.87

M_2: Ingress/Egress

ROK_Y	**0.83**	*5.39*	**0.86**	*5.55*	**0.80**	*5.26*	**0.81**	*5.77*
HEL_Z	**1.86**	*10.31*	**1.93**	*12.81*	**1.87**	*10.56*	**1.85**	*10.91*
GRD_Z	**-2.03**	*-10.83*	**-2.09**	*-11.53*	**-2.26**	*-11.57*	**-2.23**	*-12.67*
StoH	**-2.79**	*-14.97*	**-2.79**	*-15.32*	**-2.83**	*-15.30*	**-2.84**	*-16.53*
HR_Z	**1.87**	*10.23*	**1.81**	*11.46*	**1.86**	*9.95*	**1.88**	*10.90*
Stat	**-2.28**	*-6.82*	**-2.03**	*-5.98*	**-2.92**	*-7.41*	**-2.87**	*-8.06*
Age	**-1.04**	*-3.53*	**-1.01**	*-3.88*	**-1.75**	*-4.83*	**-1.66**	*-6.68*
Gend	**-0.09**	*-0.60*	**-0.01**	*-0.04*	**-0.31**	*-1.11*	**0.02**	*0.10*
cut 1	-7.48	-12.81	-12.81	-12.73	-10.39	-15.65	-9.98	-14.94
cut 2	-6.36	-14.73	-14.73	-14.82	-9.03	-17.27	-8.70	-16.61
cut 3	-5.24	-14.25	-14.25	-15.07	-7.60	-17.46	-7.33	-17.52
cut 4	-4.48	-12.68	-12.68	-13.60	-6.56	-16.08	-6.32	-16.72
cut 5	-3.78	-10.83	-10.83	-11.75	-5.58	-14.27	-5.37	-14.84
cut 6	-2.47	-7.18	-7.18	-7.53	-3.69	-9.65	-3.49	-10.33
cut 7	-1.22	-3.58	-3.58	-3.39	-1.80	-4.70	-1.59	-4.71
cut 8	-0.02	-0.06	-0.06	0.50	-0.01	-0.01	0.27	0.77

Appendix I: Distribution of Beta Parameters M1, M2, M3 Models

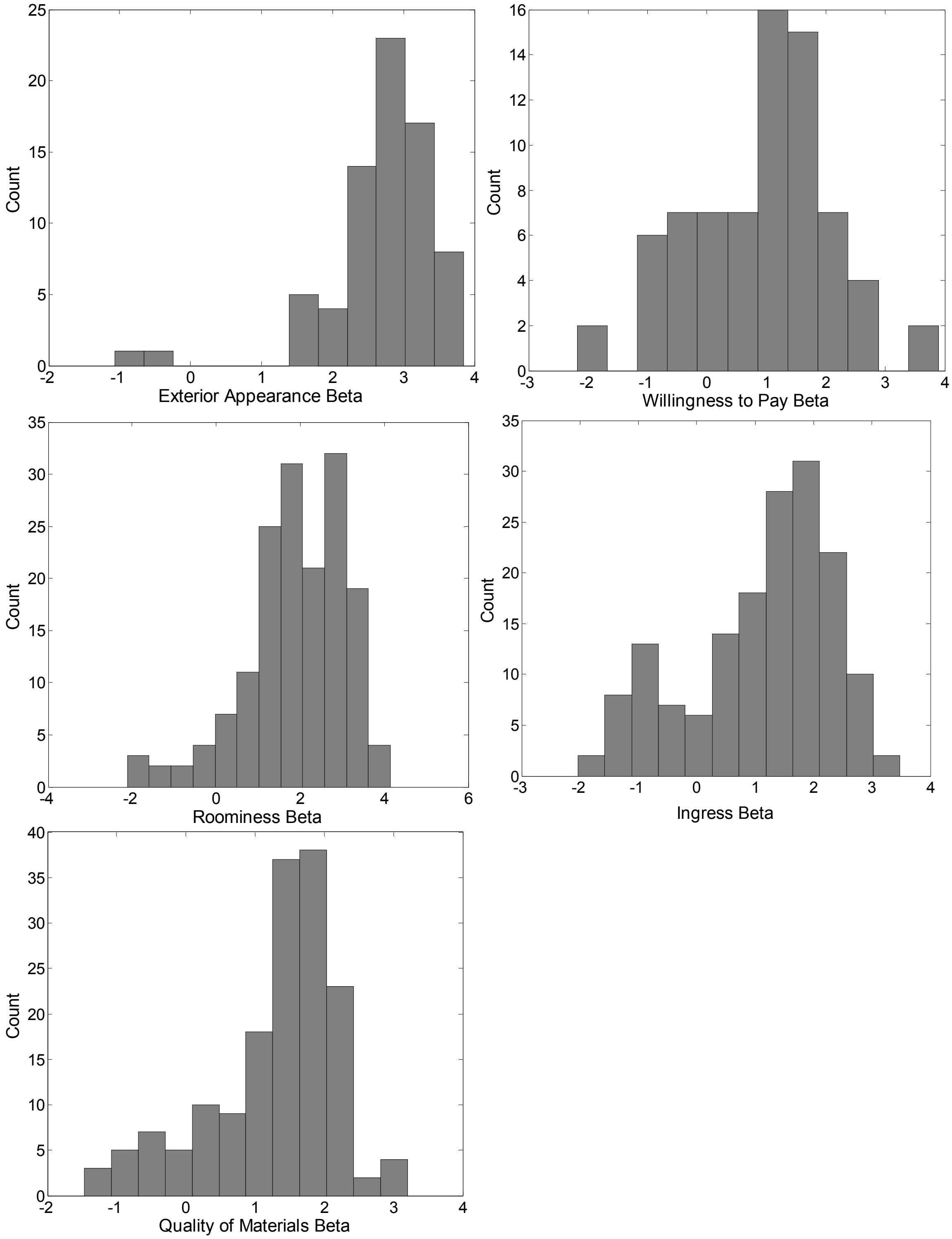


Figure I.1: M_1 Level (Choice) Beta Distributions

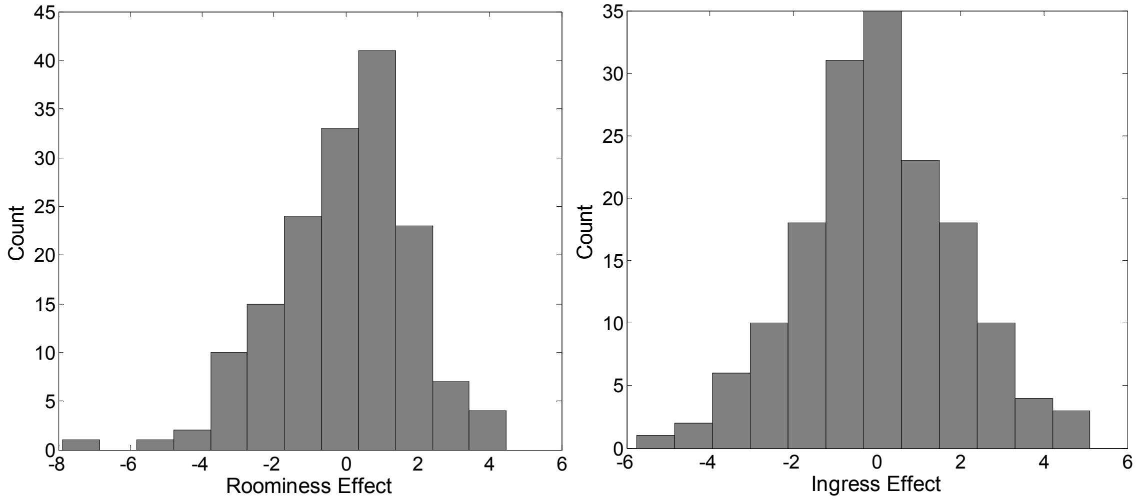


Figure I.2: M_2 Level (Ratings) Random Respondent Effect Distribution

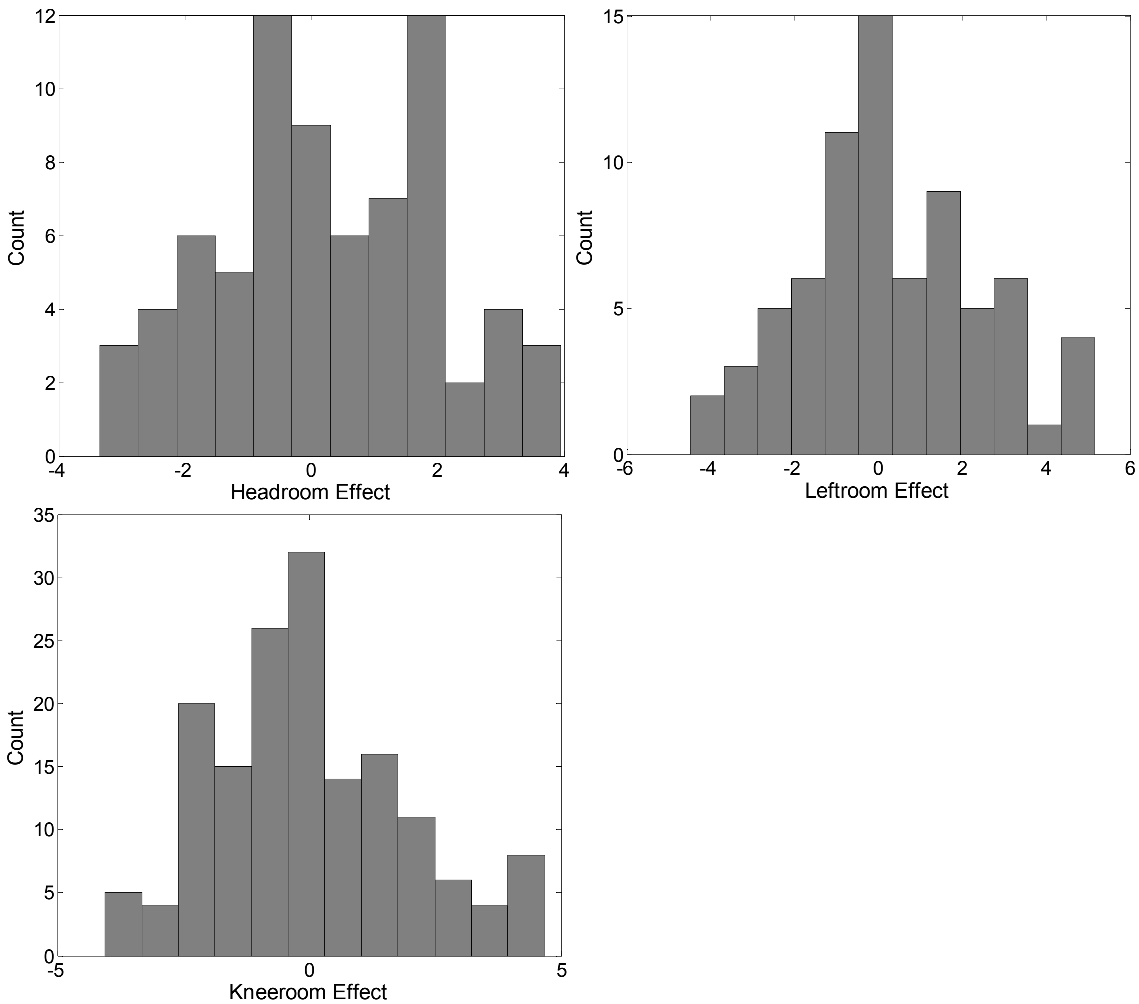


Figure I.3: M_3 Level (Ratings) Random Respondent Effect Distribution

Vita

NAME Christopher Hoyle

EDUCATION

PhD in Mechanical Engineering, Northwestern University, Evanston, IL, Dec 2009.

MS Mechanical Engineering, Purdue University, West Lafayette, IN, May 1994.

BS General Engineering, University of Illinois, Urbana-Champaign, IL, June 1988.

EMPLOYMENT

NASA Ames Research Center, Mountain View, CA. **6/2006 − 9/2006**
Summer Research Intern

Motorola, Inc., Deer Park, IL. **5/1994 – 3/2004**
Mechanical Engineering Manager

ITW Deltar, Frankfort, IL. **9/1989 – 12/1991**
Project Engineer

TEACHING

- Teaching Assistant for ISEN *Introduction to Energy Systems for the 21st Century*, Spring 2009.
- CAD Instructor for *Engineering Design and Innovation M.S. Boot Camp,* Fall 2008.
- Teaching Assistant for ME 341 *Computational Methods for Engineers,* Fall 2007, 2008.
- Teaching Assistant for ME 398 *Engineering Design*, Winter 2006, 2007, 2008, 2009.
- Teaching Assistant for ME 359 *Reliability Engineering*, Spring 2007.

HONORS

- *Presidential Fellowship Nominee*: 2007.
- *Walter P. Murphy Fellowship*: 2005-2006.
- *Altair Corporation Fellowship*: 2008.

PUBLICATIONS

Journal Publications:

Hoyle, C., Chen, W., Ankenman, B., Wang, N., "Optimal Experimental Design of Human Appraisals for Modeling Consumer Preferences in Engineering Design", In Press: (*ASME*) *Journal of Mechanical Design,* 2009.

Hoyle, C., Mehr, A., Tumer, I., Chen, W., "Health Management Allocation during Conceptual System Design", *Journal of Computing & Information Science in Engineering*, Vol. 9, No. 2, 2009.

Hoyle, C. and Chen, W., "Product Attribute Function Deployment (PAFD) for Decision–Based Conceptual Design", *IEEE Transactions on Engineering Management*, Vol. 56, No. 2, 2009.

Kumar, D., **Hoyle, C**., Chen, W., Wang, N., Gomez-Levi, G., Koppelman, F., "A Hierarchical Choice Modeling Approach for Incorporating Customer Preferences and Market Trends in Engineering Design", *International Journal of Product Development*, Vol. 8, No. 3, 2009.

Ramani, K. and **Hoyle, C**., "Processing of Thermoplastic Composites Using a Powder Slurry Technique. I. Impregnation and Preheating*," Materials and Manufacturing Processes*, Vol. 10, No. 6, pp. 1169-1182, 1995.

Ramani, K. and **Hoyle, C**., "Processing of Thermoplastic Composites Using a Powder Slurry Technique. II. Coating and Consolidation," *Materials and Manufacturing Processes*, Vol. 10, No. 6, pp. 1183-1200, 1995.

Ramani, K., Borgoankar, H., **Hoyle, C**., "Experiments on Compression Molding and Pultrusion of Thermoplastic Powder Impregnated Towpregs," *Composites Manufacturing*, Vol. 6, No. 1, pp. 35-43, 1995.

Conference Publications (Peer Reviewed)

Hoyle, C., Chen, W., Wang, N., and Koppelman, F., "Bayesian Hierarchical Choice Modeling Framework for Capturing Heterogeneous Preferences in Engineering System Design", *2009 ASME Design Engineering Technical Conference* (*IDETC/CIE*), September 2009.

Yannou, B., Wang, J., Rianantsoa, N., **Hoyle, C.**, Drayer, M., Chen, W., et al., "Usage Coverage Model for Choice Modeling: Principles and Taxonomy", *2009 ASME Design Engineering Technical Conference* (*IDETC/CIE*), September 2009.

Tucker, C., **Hoyle, C.**, Kim, H., Chen, W., "A Comparative Study of Data-Intensive Demand Modeling Techniques in Relation to Product Portfolio Design", *2009 ASME Design Engineering Technical Conference* (*IDETC/CIE*), September 2009.

Hoyle, C., Chen, W., and Wang, N., "Understanding Heterogeneity of Human Preferences for Engineering Design", *International Conference on Engineering Design* (*ICED*), August 2009.

Hoyle, C. Chen, W., Ankenman, B., Wang, N., "Optimal Experimental Design of Human Appraisals for Modeling Consumer Preferences in Engineering Design", *2008 ASME Design Engineering Technical Conference*, August 2008.

Hoyle, C., Mehr, A., Tumer, I., Chen, W., "Cost-Benefit Quantification of ISHM in Aerospace Systems", *2007 ASME Design Engineering Technical Conference* (*IDETC/CIE*), Sept. 2007.

Kumar, D., **Hoyle, C**., Chen, W., Wang, N., Gomez-Levi, G., "Incorporating Customer Preferences and Market Trends in Vehicle Package Design", *2007 ASME Design Engineering Technical Conference* (*IDETC/CIE*), September 2007.

Hoyle, C. and Chen, W., "Next Generation QFD: Decision-Based Product Attribute Function Deployment", *International Conference on Engineering Design* (*ICED*), August 2007.

Liu, H., **Hoyle, C**., Yin, X., Chen, W., "Setting Performance Targets Based on Subsystem Pareto Frontiers in Multilevel Optimization", *ASME International Mechanical Engineering Congress*, November, 2006.

Hoyle, C., Kumar, D., Chen, W, "Product Attribute Function Deployment (PAFD) for Decision–Based Conceptual Design", *2006 ASME Design Engineering Technical Conference* (*IDETC/CIE*), September 2006.

Ramani, K., Borgaonkar, H., **Hoyle, C**., "Experiments on Compression Molding and Pultrusion of Thermoplastic Powder Impregnated Towpregs," *ASME International Mechanical Engineering Conference Symposium on Processing, Design and Performance of Composite Materials*, Vol. 52, pp. 183-204, Nov, 1994.

Ramani, K., **Hoyle, C**., and Parasnis, N., "Flexible Thermoplastic Powder Impregnated Tows for Net-Shape Manufacturing," *ASME Winter Annual Meeting Symposium on Use of Plastic and Plastic Composites: Materials and Mechanics Issues*, New Orleans, LA, Vol. 46, pp. 633-657, November, 1993.

Ramani, K., Tryfonidis, M., **Hoyle, C.**, and Gentry, J., "Thermoplastic Powder Composite Manufacturing Using a Wet Slurry Method," *ASME Winter Annual Meeting Symposium on Processing Fabrication and Manufacturing of Composite Materials*, Anaheim, CA, Vol. 35, pp. 115-130, November, 1992.

Conference Publications:

Hoyle, C. and Chen, W., "Using HyperWorks and iSIGHT for Teaching Computational Methods in Engineering Design", *PACE Annual Forum*, Darmstadt Germany, July 2007.

Hoyle, C., Mehr, A., Tumer, I., Chen, W., "On Quantifying the Cost-Benefit of ISHM in Aerospace Systems"**,** 2007 IEEE Aerospace Conference, Big Sky, MN 2007.

Hoyle, C. et al., "Optimal Stamping Binder Design Methodology", *7th World Congress on Computational Mechanics*, Los Angeles, CA, July 2006.

PATENTS

- Repplinger, S., Slaby, J., Hoyle, C., Lau, B., "Low Inductance Termination for Electronic Components", U.S. Patent 6,545,855, April 8, 2003.
- Chen, C., Hoyle, C., Kosberg, R., Meny, K., Poglitsch, L., Nowicki Jr., J. "Electrical Device Having Atmospheric Isolation", U.S. Patent 6,053,049, April 25, 2000.
- Hoyle, C. and Peek, B., "Pin and Grommet", U.S. Patent 5,193,961, March 16, 1993.
- Hoyle, C. and Peek, B., "Sliding Grommet", U.S. Patent 5,129,768, July 14, 1992.

MEMBERSHIPS

- American Society of Mechanical Engineers (ASME)
- American Institute of Aeronautics and Astronautics (AIAA)
- Society of French Automotive Engineers (SIA)