NORTHWESTERN UNIVERSITY

Usage and Social Context-based Choice Modeling for Engineering Design

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Mechanical Engineering

By

Lin He

EVANSTON, ILLINOIS

August 2012

UMI Number: 3527571

Abstract

Usage and Social Context-based Choice Modeling for Engineering Design

Lin He

This dissertation is motivated by the need to develop analytical techniques that integrate engineering, marketing, and social science domains to incorporate heterogeneity in consumer preferences into product design, considering the influence of usage context and impact of social network on consumers' choices. The research primarily uses a vehicle case study as a motivating example, to both illustrate the challenges in the product design and demonstrate the features of the proposed choice modeling approach. The research can be divided into four primary contributions.

A new procedure called *Integrated Mixed Logit Model* (IMLM) is introduced to incorporate consumer perceptions, such as rating data, as well as quantitative attributes into the decision-theoretic choice modeling process. The IMLM method is built upon established Decision-Based Design approach in engineering to address the challenges in modeling subjective ratings collected in market surveys.

To capture the critical role usage context plays in consumer choices, a framework of *Usage Context-based Choice Modeling* (UCBCM) is developed to quantify the influence of usage context on both product performances and consumer preferences. The advantages of this approach over existing approaches for considering distinctive usage patterns and improving

model accuracy are clearly demonstrated through the jigsaw case study with stated preference data and the vehicle case study with revealed preference data.

The adoption of hybrid electric vehicle in the past decade has gained wide attention because its potential in reducing greenhouse gas emission and utilizing renewable energy sources. With two survey datasets from different sources, multivariate statistical techniques are utilized to understand the relationship between consumer profile and usage context attributes, identifying key characteristics of *hybrid electric vehicle adopters*, and estimating choice models to capture the heterogeneity in consumer preferences towards hybrid electric vehicles.

The *Agent-based Choice Modeling* (ABCM) framework provides a unified choice modeling approach for considering social impact through interpersonal network on new product adoption. It utilizes agent-based simulation and discrete choice modeling to create the social network structure and evaluate the social impact, which is later integrated into the choice utility function to capture the dynamic nature of consumers' attitude towards new product or new technology changing over time.

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 very thankful to my committee members: Professor Frank Koppelman, and Professor Uri Wilensky for their suggestions and contributions to my research. Collaborating with them on my research projects were valuable experiences. I would like to acknowledge Professor Christopher Hoyle of Oregon State University, formerly of Northwestern University for his guidance and collaboration on the research projects in and outside of this dissertation. I also owe special gratitude to Professor Bernard Yannou and Jiliang Wang in the Industrial Engineering Department at Ecole Centrale Paris for their contribution in the collaboration project, as well as their thorough review of the findings in this work. My gratitude also goes to Professor Gary Wang and Zhila Pirmoradi at Simon Fraser University for their assistance with the PSAT simulation tools, which made Chapter 5 of this dissertation possible. I would also like to acknowledge research and staff members of Palo Alto Research Center, particularly Dr. Tolga Kurtoglu, who hosted me as a summer intern in 2011 and collaborated with me on research projects outside of this dissertation.

I am grateful to my great colleagues at the Integrated Design Automation Laboratory (IDEAL). Many thanks go to current members Paul Arendt, Steven Greene, Peilin Tian, Hongyi Xu, Zhen Jiang, Shuangcheng Yu, Mingxian Wang, and Katherine Rorschach, as well as previous members Sanghoon Lee, Ying Xiong, Xiaolei Yin, Shikui Chen, Fenfen Xiong, and Yu Liu for their valuable discussions, advices, and help getting me started in the graduate school. I really enjoyed the friendly and collaborative environment in IDEAL.

Finally, I would like to thank my husband Yuchen Wu, my parents, other family members and friends for their continued support during my PhD study. 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 DUE-0920047) and Initiative for Sustainability and Energy at Northwestern (ISEN). 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

Consumer-desired Product 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.

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 described 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.

Usage Context Attributes (E): Characteristics or attributes used to describe the usage context, which includes all aspects related to the use of a product excluding consumer profile (demographic attributes) and product attributes.

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

Social Influence Attributes (N): Attributes describing social influence process, in which consumers may modify their behavior to bring them more closely into alignment with behavior of their friends.

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

Rating (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.

Consumer Profile Attributes (S): Attributes of the consumer including socio-economic (e.g. income), anthropomorphic (e.g. height), and purchase history (e.g. Ford Focus).

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

Product Performance (Y): Attributes of a product or system used in engineering analysis and decision making, such as horse power, occupant package headroom, or fuel economy of an automobile.

Model Attributes (Z): The set of all consumer-desired product attributes **A** or product performances **Y**, consumer profile attributes S, usage context attributes **E**, and social influence attributes **N** included in the choice or rating model, including interactions among the model terms and high order terms.

List of Abbreviations

ABM	Agent-based Modeling
ABCM	Agent-based Choice Modeling
AER	All Electric Range
AFV	Alternative Fuel Vehicle
ANOVA	Analysis of Variance
ASC	Alternative Specific Constants
ASV	Alternative Specific Variables
AV	Advanced Vehicle
BSC	Brand Specific Constants
CD	Conventional Vehicle Driver
CD-mode	Charge-deleting Mode
CSI	Consumer Satisfaction Index
CSS	Consumer Satisfaction Survey
CS-mode	Charge-sustaining Mode
CV	Conventional Vehicle
DBD	Decision-Based Design
DCA	Discrete Choice Analysis
DOE	Design of Experiments
EV	Electric Vehicle
GenCos	Power Generation Companies
GHG	Greenhouse Gas
HBE	Hierarchical Bayes Estimation

HCI	Human-Computer Interface
HD	Hybrid Electric Vehicle Driver
HEV	Hybrid Electric Vehicle
IMLM	Integrated Mixed Logit Model
LR	Linear Regression
MAV	Multi-activity Vehicle
MLE	Maximum Likelihood Estimation
MNL	Multinomial Logit
MSRP	Manufacturer Suggested Retail Price
MXL	Mixed Logit
NHTS	National Household Travel Survey
NL	Nested Logit
OL	Ordered Logit
PAFD	Product Attribute Function Deployment
PHEV	Plug-in Hybrid Electric Vehicle
PSAT	Powertrain System Analysis Toolkit
RP	Revealed Preference
SOC	State of Charge
SP	Stated Preference
SUV	Sport Utility Vehicle
UCBD	Usage Context-based Design
UCBCM	Usage Context-based Choice Modeling
VQS	Vehicle Quality Survey

Nomenclature

A	Consumer-desired product attributes
α	Rewiring probability
β	Coefficients in consumer's choice utility function
C	Cost
d_{ij}	Social distance from consumer i to consumer j
E	Usage context attributes
$\mathbf{E_W}$	Preference-related usage context attributes
$\mathbf{E_Y}$	Performance-related usage context attributes
ε_{ik}	Random disturbance of consumer choice utility of product k by consumer i
F	Usage importance index
J_n	Choice set of consumer n
L_{ij}	Social link from consumer i to consumer j
l_{ij}	Strength of social connection from consumer i to consumer j
M	Exogenous variables
M	Choice share
$\mathbf{N}_{i,t}$	Social impact on consumer j's attitude towards product adoption at time t
N	Market population
P	Price
$P_{i,n}$	Choice probability
Q	Demand
γ	Coefficients in social impact function
R	Rating

S	Consumer profile /demographic attributes
$\mathbf{S_W}$	Preference-related consumer profile attributes
$\mathbf{S_Y}$	Performance-related consumer profile attributes
t	Time
u	Consumer choice utility
U	Enterprise utility function
U	Usage context scenario
V	Selection criterion
$W_{ik,t}$	Observed (deterministic) part of the consumer choice utility of product *k* by consumer *i* at time *t*
X	Engineering design options or variables
x_i^m	Value of consumer *j* in the *m*th social dimension
Y	Engineering performances
$\mathbf{Y_r}$	Performance of service results
$\mathbf{Y_t}$	Performance of service delivery / transformation
y	Consumer behavior
Z	Explanatory variables
Π	Profit

Table of Contents

List of Tables

List of Figures

Chapter 1

PROBLEM DESCRIPTION AND RESEARCH OBJECTIVE

This dissertation is motivated by the need to develop analytical techniques that integrate engineering, marketing, and social science domains to incorporate consumer preferences into product design. Designing a product consumers like requires not only a good knowledge of how engineering performance are linked to consumer desired product attributes, but also a deep understanding of the context of consumers' choices and how they make their choices. To make rigorous decisions for product design, it will be demonstrated in this dissertation that it is necessary to consider factors beyond the traditional engineering domain to include consumers' perception, usage context, and social influence. As shown in Figure 1.1, the interest in this work is to explore the intersection and interaction of consumer, product, and context in design research by integrating fundamental principles from multiple disciplines across fields like engineering, marketing, and social science.

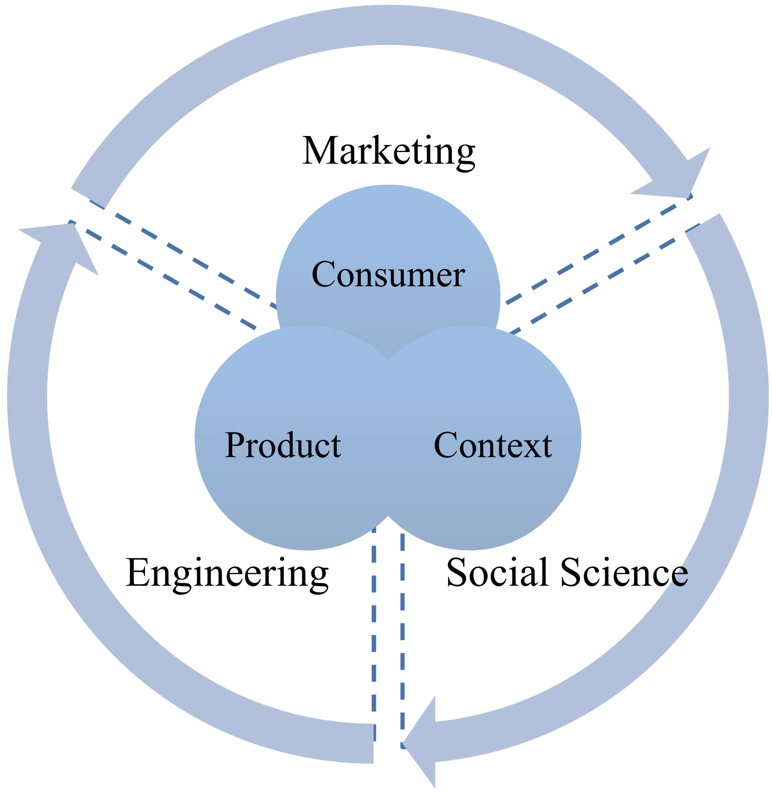


Figure 1.1: Connections Among Engineering, Marketing and Social Science Domains

Traditional engineering design is conducted primarily with a product-centric viewpoint, in which the objective is to achieve the best engineering performance subject to the cost restrictions. As noted in a variety of contexts (Clausing and Hauser, 1988, Krishnan and Ulrich, 2001, Ullman, 2002), each of the major functional domains within a firm (or enterprise), such as engineering, marketing, production, and management seeks to optimize domain-specific objectives with limited communication with other functional domains. While previous works (Kumar et al., 2009b, Hoyle et al., 2010) attempt to bridge the gap between the engineering and marketing domains by mapping the relationship between engineering design attributes and consumers' choices, the domain of social science that contains rich information of usage and social context is often overlooked. In the social science domain, it has been long recognized that usage context influences consumers' behavior (Robertson and Ward, 1973, Lavidge, 1966, Engel et al., 1969). Empirical studies also show that social context, such as neighbor effects impact consumers' choice behavior (Case, 1992). The question is how should the qualitative principles in social science and marketing research be integrated with the quantitative methods in engineering design and decision making?

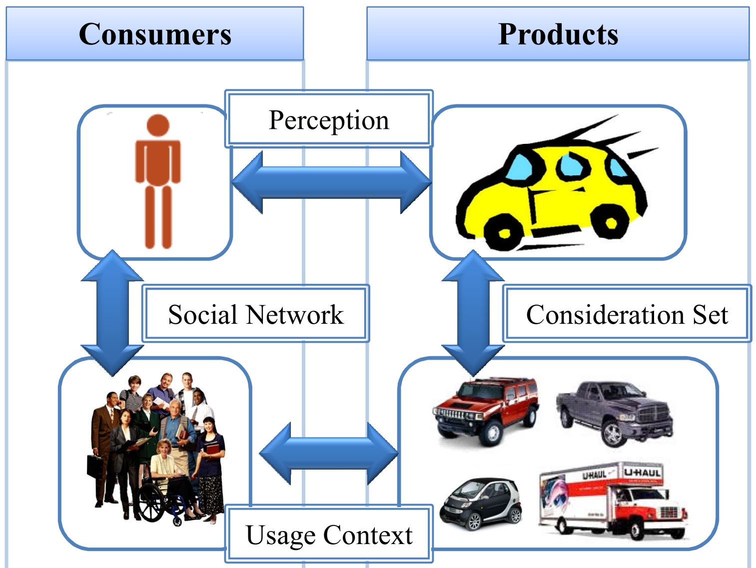


Figure 1.2: Interactions between Consumers and Products

As illustrated in Figure 1.2, consumers, products, and the contexts in which the products are used, are seldom isolated. Numerous connections among individual consumers, consumer groups, individual products, and product families play a key role in consumer's choice behavior. Consumers' **perceptions** of products' performance are formed through human and product interactions, and are normally expressed in the form of ratings in market survey, which contribute to the design utility underlying consumer' choice decision. **Consideration set** refers to the set of competing products consumers consider during purchase; the set is constructed based on competing products that fall into the same market segmentation. The **usage context** information emerges from the interaction between consumers and a series of products belonging to either the same or different product families, as common usage scenarios among groups of consumers can be identified. The **social network** provides interpersonal links, through which consumers communicate with each other, exchange their opinions, and influence each others' attitudes toward new products.

Take the design of Plug-in Hybrid Vehicle (PHEV) as an example, as the vehicle is new to the market, to plan for product design development efforts, it is critically important for the vehicle manufacturers to forecast the market potential of such product. As a significant advancement to the existing literature, new choice modeling techniques are needed to consider not only the technical performance of the vehicles and the socio-demographic background of consumers, but also the intended usages for the vehicle (e.g., commute length, local vs. highway driving), consumers' uncertain attitudes toward new technology, the "green attitudes", and the influences of social networking (Axsen and Kurani, 2009) as well as regional differences and infrastructure (e.g., dealer network, road types). From a broader system point of view, such choice models should be integrated into a multi-agent energy market simulation framework to study the impact of consumer vehicle choices on future electric generation needs. As the energy market continues to evolve, there is a growing need for advanced consumer preference modeling approaches and a multi-agent simulation framework to study the dynamic effect of engineering design decisions, considering consumer profiles, their perceptions, usage context, social network, and consideration set.

1.1 RESEARCH MOTIVATION AND CHALLENGES

The need for integrating marketing research into engineering design has been widely recognized over the last few decades (Krishnan and Ulrich, 2001). To better understand consumers' choice behavior and analytically predict demand (the choice share of a design among competing products), recent years have seen a growing interest in developing quantitative choice modeling approaches to predict consumers' choice as a function of engineering designs (Cook, 1997, Li and Azarm, 2000). **Discrete Choice Analysis** (DCA) (Ben-Akiva and Lerman, 1985) is a probabilistic choice modeling approach that estimates **choice probability** for a given design alternative among a set of discrete competing designs over a sample population and aggregates the choice probability for a given design alternative to estimate its **market share,** and ultimately its demand. DCA is a flexible approach which models choice using a utility function composed of observed *product* and *consumer* level attributes as inputs, and its model coefficients can be estimated using survey or actual choice data, or a combination of both. Wassenaar et al. (2003, 2005, 2006) utilized multinomial logit to model vehicle demand based on the revealed vehicle purchase data. Kumar et al (2009b) proposed a nested logit modeling (Koppelman and Sethi, 2000) approach to estimate a system-level demand model for vehicle package design by pooling data from multiple component/subsystem-specific surveys. Further, Kumar et al (Kumar et al., 2009a) integrated the nested logit within a design optimization formulation to optimize platform-based product family design. Using the mixed logit models, Michalek (2005) considered random consumer preference heterogeneity in choice modeling by including the distributions of attribute coefficients, but the method did not include consumer demographic attributes as exploratory variables (termed systematic heterogeneity). In the most recent development, Hoyle (2009, 2010) developed the hierarchical choice modeling framework and used the mixed logit modeling technique to model both systematic and random heterogeneity across component, subsystem, and system levels in vehicle package design.

While the previous work has laid the foundation for incorporating demand modeling in engineering design, there are several issues which must be addressed to enable a comprehensive understanding of consumer preference in support of engineering design. The issues are summarized as follows:

- A systematic approach does not exist for incorporating consumers' perception such as ratings into choice modeling. Consumer satisfaction survey data collected after product purchases provides rich information of consumers' perception of product performance in forms of ratings. However, the question remains whether individual ratings can be directly used as a measure for qualitative product attributes in consumer choice modeling. Previous work (Hoyle et al., 2009) has developed a choice modeling structure for modeling qualitative attributes, in which rating are collected during human appraisal survey, but issues related to rating data from satisfaction survey after product purchases, such as ownership bias, differences in rating style, and missing choice alternatives' attributes, are not addressed.
- While usage context is a critical factor in forming consumer preferences, it has not been formally addressed in any existing choice modeling working in engineering design. Similar to consumer profile and market offerings, it is widely recognized that usage context exerts a large impact on consumers' choice (Berkowitz et al., 1977). Green (2005, 2004, 2006) introduced usage context into the product design process, but in a rather qualitative way. In order to fully capture the impact of usage contexts upon consumer choices, taxonomy for understanding usage context in engineering design and a comprehensive choice modeling framework are needed.
- Existing choice modeling approaches in engineering design are static in nature in that consumers are independent and isolated from each other, while in reality, people interact, and

influence each other by all means on a daily basis. A simulation framework is needed to study the social influence on product choice by modeling interactions among consumers; further, such simulation needs to be integrated with conventional choice modeling under a unified framework to consider social influence together with the influences of other factors, such as engineering performance, consumer demographic attributes, and usage context.

Given the aforementioned issues in incorporating consumer preferences, the usage, and social context in engineering design, the objectives of this dissertation are 1) to create a rigorous modeling approach for effectively incorporating consumers' perception of qualitative performance, such as rating data collected in consumer satisfaction survey after purchase, into the demand modeling framework; 2) to develop the taxonomy and a choice modeling framework for considering the role of usage context in consumers' preferences; 3) to establish a data analysis process to understand consumer preference of new products and identify key characteristics of the early adopters with respect to usage and social contexts; and 4) to develop an agent-based choice modeling and simulation approach to capture the impact of social network upon consumers' choices.

1.2 Usage and Social Context-based Choice Modeling Approach for Engineering Design

To realize the usage and social context-based choice modeling approach for engineering design, research is required in four core areas: the Choice Modeling Incorporating Rating Data, the Usage Context-based Choice Modeling, the Statistical Data Analysis of Early Adopters, and the Agent-based Choice Modeling Considering Social Impact. Research in these four core elements forms the focus of this dissertation; each research task is described in more detail in the following paragraphs.

Tasks	Rating Data	Usage Context	Early Adopters	Social Context
Case Study	Conventional Vehicles (CV)	Jigsaw	Hybrid Electric Vehicles (HEV) Plug-in Hybrid Electric Vehicles (PHEV)	
Research Question	***How to incorporate rating data into choice modeling?***	***How to model usage context's influence upon choice?***	***What are the key characteristics of early adopters?***	***How does social impact evolve throughout time?***
Topics	✓ Examination of rating data in consumer satisfaction survey ✓ Integrated mixed logit modeling procedure	✓ Taxonomy of usage context-based design ✓ Usage context-based choice modeling	✓ Multivariate analysis ✓ Factor analysis ✓ Choice modeling with pooled data	✓ Social network construction ✓ Social impact evaluation ✓ Choice model estimation

Figure 1.3: Overview of Research Tasks

Research Task 1---- Choice Modeling Incorporating Rating Data

As noted, product design decisions require consideration of engineering performance and cost, as well as market acceptance to ensure the resulting design will be profitable and benefit the enterprise. However, challenges remain for directly using the individual ratings in consumer satisfaction survey for consumer choice modeling to guide engineering design. A close examination of consumer satisfaction survey data with respect to its applicability to consumer choice modeling will be first provided under this task. The key issues will be identified. To alleviate the limitations, a systematic mixed logit based choice modeling procedure will be developed to incorporate the use of both quantitative and subjective rating measures in the model utility function, together with the consumer demographic attributes. A case study using the real Vehicle Quality Survey data acquired from J.D. Power and Associates will be used to demonstrate many of the key findings under this task.

Research Task 2----Usage Context-based Choice Modeling

Usage context attributes play a critical role in consumers' choice by influencing both product performance and consumer preference. Previous works recognized the importance of usage context and demonstrated ways of qualitatively identifying usage context in the product design. In this task, a taxonomy for the Usage Context-based Design (UCBD) is first defined by following the established usage context terminology in the market research domain and the needs associated with choice modeling. A framework and a step-by-step procedure for implementing consumer choice modeling in UCBD will be developed. To implement the proposed approach, methods for common usage identification, data collection, linking performance with usage context, and choice model estimation will be developed. The developed approach will be illustrated with case studies of the jigsaw and hybrid electric vehicles (HEVs).

Research Task 3---- Statistical Data Analysis of Early Adopters

Forecasting the demand of new product is challenging in that limited market data is available to support the modeling efforts. For example, the use of advanced vehicles (AVs), such as the PHEVs and/or all-electric vehicles (EVs) in the near future is expected to grow significantly. Most recently, several major vehicle manufacturers launched their first PHEV models in the market. For the adoption of PHEVs, both usage and social context play a critical role in influencing consumers' preference and attitudes toward the new technology. While the ultimate goal of this research is the development of choice modeling to forecast new product demand and support product design, the data used in the modeling process must be analyzed to identify the key characteristics of early adopters and understand their preferences of new products. Specific methods are developed to analyze consumer profile and usage context attributes, which present a unique set of challenges compared to industrial or scientific experiments due to the effects of consumer heterogeneity and human behavior. The research topics under this task will cover the multivariate statistical methods such as analysis of variance and factor analysis, and choice model estimation using multiple data sources.

Research Task 4----Agent-based Choice Modeling considering Social Context

Consumer choices are social – people are influenced by friends and other contacts in their social network. While works have been published on social network studies (Deffuant et al., 2005, Delre et al., 2007), no attempt has been made to quantify the social impact at the individual level as part of the utility function in a choice model. To account for these interpersonal interactions, an agent-based modeling approach that links the discrete choice model and the agent-based simulation will be developed in this work. Research under this task will focus on developing methods to support the major stages in agent-based modeling considering the influences of social context, including social network construction, social influence evaluation, and choice model estimation. The proposed approach will be demonstrated with HEV market example.

1.3 Alternative Fuel Vehicle Example

In this dissertation, alternative fuel vehicle example has been used for case studies throughout the following chapters. Alternative fuel vehicles (AFV) have drawn increasing attention in the past few years, because of their potential to reduce greenhouse-gas emissions and utilize renewable energy sources (Ehsani et al., 2009, Axsen et al., 2008, Shiau et al., 2009a). Forecasting the adoptions (first-time purchases) of AFVs is a critical but challenging task. This topic is not only important to marketing and managerial division of the enterprise, but also of great interest to the engineering design team, as the market potential determines the success of a new product. For example, with new PHEV design entering the market in the near future, understanding the tradeoffs among engineering design attributes related to PHEVs such as battery weight, charging time, driving range, etc., is critically important for setting up appropriate targets for designing new battery systems. However, understanding consumer choices of alternative fuel vehicles is challenging because their preference construction process involves many aspects beyond traditional engineering considerations, which calls for a comprehensive modeling framework to incorporate social impact into engineering design. Specifically, consumers' attitudes towards green technology are strongly influenced by social network (Axsen and Kurani, 2009) - social network plays an important role in consumers' choice behavior by means of neighbor effects or work-of-mouth effects. Large amount of product reviews and recommendation features made available by the rapid growing online shopping websites and social networking sites further accelerate the social impact, which needs to be considered in modeling consumer preferences.

Because of its potential impact on energy and environment, there are a handful of research works on forecasting the PHEVs' market potential. Rousseau et al (2007) developed a process to define the requirements of energy storage systems for plug-in applications and describe the impact of All Electric Range (AER), drive cycle, and control strategy on battery requirements. Shiau et al (2009a) studied PHEV batteries' effect on vehicle cost, weight, and performance, and concluded that the focus should be placed on adoption of small-capacity PHEVs by urban drivers who can charge frequently. Axsen et al. (2008) reviewed the research on designing PHEV batteries including nickel-metal hydride (NiMH) and lithium-ion (Li-Ion) and discussed the inherent trade-offs among power, energy, longevity, safety, and cost, in battery design. In the area of studying consumer preferences, a few pilot projects have been conducted to better understand consumers' knowledge and awareness of PHEV, how they evaluate PHEV during the demonstration program, and what they expect from a PHEV (Axsen and Kurani, 2008). While these works have explored the PHEV potentials and provided insights into future market of PHEV, their focus is either on cost-performance engineering models or qualitative consumer study. Linking engineering design to consumer preferences is necessary to understand its market potential and therefore to improve the PHEV design.

The design of alternative fuel vehicle represents a complex system with multi-players, i.e., vehicle manufacturers, consumers, and power generation companies (GenCos), as illustrated in Figure 1.4. Because alternative fuel vehicles, such as PHEVs, will increase load on the electrical grid due to battery charging, the demand and power needs for such vehicles will have a large impact on the future energy market and must be forecasted accurately to plan for the reconstruction and expansion of the power industry. In contrast, the future electric generation needs depend on the types of alternative fuel vehicles offered by vehicle manufacturers, the energy capacity (and hence energy cost) provided by GenCos, and the vehicle choices made by individual consumers. Consumers are heterogeneous, differing in their preferences for vehicles based on their socio-demographics (e.g., income, age, region), the influence of social networks (e.g., neighbors, family), and vehicle usage (e.g., miles-driven, transportation mode). As the energy market continues to evolve, there is an increasing need for advanced consumer preference modeling to study the dynamic effect of vehicle manufacturer offerings, consumer preferences, and the trend of green product adoption.

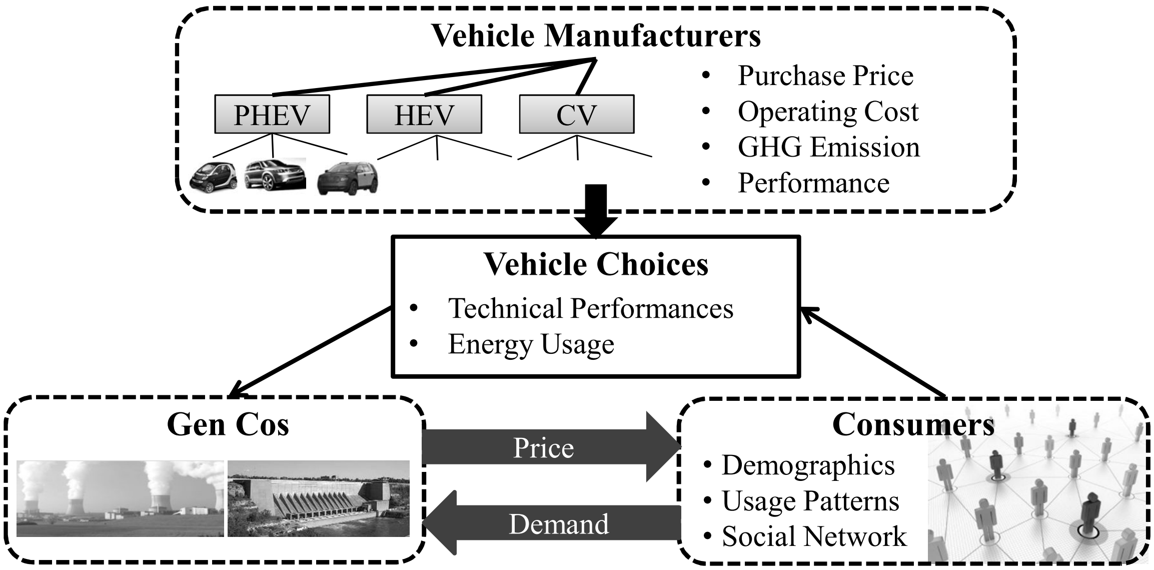


Figure 1.4: Interactions Among Multi-Agents in Designing HEVs & PHEVs

1.4 Organization of the Dissertation

The organization of the dissertation is as follows. 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 Integrated Mixed Logit Modeling (IMLM) framework for modeling rating data to address Research Task 1. While the method is general and can be used for different types of rating data in various product categories, the method is demonstrated with consumer satisfaction survey data in a vehicle case study. Chapter 4 presents the choice modeling framework for UCBD to address Research Task 2. This method provides the means to collect usage context data, link product performance to usage context, and model consumer preference for single or multiple usage context, which is innovative in quantitatively building preference models and understanding consumer heterogeneity in terms of product usage. Chapter 5 provides a methodology to statistically analyze data from multiple sources to understand heterogeneous preferences of early adopters as well as to preprocess the data to create efficient preference models. The data analysis results are used in the vehicle design selection study. The methods presented address Research Task 3. Chapter 6 presents an Agent-based Choice Modeling (ABCM) framework considering social impact which provides a comprehensive choice modeling approach for modeling impact of social network upon consumers' choice of new products, in particular green products (Research Task 4). The model is estimated using both market data from multiple years and social network simulation. Chapter 7 details the contribution of this research as well as areas for future work.

Chapter 2

TECHNICAL BACKGROUND

2.1 INTRODUCTION

The research work in this dissertation is rooted in the discrete choice method for modeling product demand, as a part of the 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 DBD framework is introduced, a brief tutorial on discrete choice analysis for demand modeling is provided; the hierarchical choice modeling approach and the multilevel formulation of the DBD framework is presented; social network theories and their integration into the choice modeling approach is summarized; agent-based model and simulation is described, and method for statistical analysis and preprocessing of data are provided.

2.2 DECISION-BASED DESIGN (DBD) FRAMEWORK

2.2.1 DBD Motivation and Overview

Within the engineering research community, there is a growing recognition that decisions are the fundamental construct in engineering design (2006b, 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 through a hierarchical model linkage in which design concepts and variables (***engineering attributes*** **X**) are linked to demand, Q, through engineering analysis and attribute mapping between ***engineering performances*** **Y** and ***consumer-desired product attributes*** **A**. Also key is the inclusion of ***demographic attributes*** **S**, in addition to consumer-desired product attributes **A,** in the estimation of demand, to capture the heterogeneity of consumer preferences. In the DBD method, 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.

2.2.2 Enterprise-Driven Design Formulation

The DBD approach takes an enterprise view in formulating a design. In our 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. Profit is expressed as a function of product demand Q, price P, and cost C, where demand Q, is expressed as a function of consumer-desired attributes **A**, consumers' demographic attributes **S**, price P, and time t:

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

Consumer-desired attributes **A** are product characteristics that a consumer typically considers when purchasing the product. To enable engineering decision-making, consumer-desired attributes **A** must be expressed as a function of engineering attributes **X** in the demand modeling phase. This functional relationship can consist of a *hierarchy* of models mapping **A** to **X** to establish the relationships necessary for decision-making. Cost, C, is a function of the engineering attributes, **X**, exogenous variables **M** (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. Based upon these functional relationships, the selection criterion can be expressed as:

$$V = \Pi = Q\left(\mathbf{A}\left(\mathbf{X}\right), \mathbf{S}, P, t\right) \cdot P - C\left(\mathbf{X}, \mathbf{M}, Q, t\right) . \qquad (2.2)$$

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:

$$\mathbf{max}: \quad E\left(U\right) = \int_V U\left(V\right) pdf\left(V\right) \, dV , \qquad (2.3)$$

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

Hence, decision-making regarding a preferred design concept, as well as optimal levels (targets) of engineering design attributes **X** can be performed using optimization to maximize the expected enterprise utility $E(U)$, subject to appropriate constraints.

2.3 Discrete Choice Analysis (DCA) for Demand Modeling

Discrete Choice Analysis (DCA) refers to a statistical technique of building probabilistic choice models (Ben-Akiva and Lerman, 1985, Koppelman and Bhat, 2006), which originated in mathematical psychology (Luce, 1959, Thurstone, 1994) and found wide application in transportation (Wen and Koppelman, 2001), marketing research (Ben-Akiva and Boccara, 1995) and econometrics (Greene, 2003). It is used to model product demand by capturing *individual* consumers' 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 consumers could be either individual consumers or industrial consumers. DCA is based upon the assumption that individuals seek to maximize their personal ***consumer choice utility***, *u*, (not to be confused with enterprise utility, *U*) when selecting a product from a choice set.

2.3.1 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):

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

As formulated in our previous work (Wassenaar and Chen, 2003), observed utility W_{in} is expressed as a function of consumer-desired attributes **A** and consumer demographic attributes **S**.

$$W_{in} = f(\beta : \mathbf{A}_i, \mathbf{S}_n) \,. \tag{2.5}$$

where $\mathbf{A}_i$ denotes the consumer-desired attributes of alternative i, and $\mathbf{S}_n$ denotes the consumer attributes of respondent *n*. In this formulation, *f* indicates that W_{in} is a function of **A** and **S** as well as the β coefficients, which are estimated by observing choices respondents make. Note that W_{in} does not have to be linear in $\mathbf{A}_i$ and $\mathbf{S}_n$ but rather, $f(\beta, \mathbf{A}_i, \mathbf{S}_n)$ could take any arbitrary function of $\mathbf{A}_i$ and $\mathbf{S}_n$ (e.g. interaction terms, quadratic terms).

By far, the most basic and most widely used discrete choice model is the multinomial logit (MNL) model, in which each ε_{ni} is assumed to be independently, identically distributed extreme value, also called Gumbel or Type I Extreme Value. In multinomial logit, the choice probability for product *i* and person *n* can be calculated in the following closed form expression:

$$P_{in} = \frac{\exp(W_{in})}{\sum_j \exp(W_{jn})}. \tag{2.6}$$

Recently, another advanced, highly flexible discrete choice model, the mixed logit model (MXL) (McFadden and Train, 2000), has gained wide popularity because, unlike the MNL model, mixed logit models allow for random taste variation, i.e. the parameters β vary over respondents. Therefore, the mixed logit probabilities are integrals of the MNL probabilities over a density of parameters, as expressed in the form:

$$P_{in} = \int \left(\frac{\exp(W_{in}(\beta, \mathbf{A}_i, \mathbf{S}_n))}{\sum_j \exp(W_{jn}(\beta, \mathbf{A}_j, \mathbf{S}_n))} \right) f(\beta) d\beta, \tag{2.7}$$

where $f(\beta)$ is the probability density function of model parameter β s. One of the most important advantages of the mixed logit model is that heterogeneity in consumer 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 Eqn. (2.7). Therefore in practical applications, the mixed logit choice probability is approximated using numerical simulations by taking a finite number of draws $k = 1, 2, 3, ..., K$ from the distribution:

$$\breve{P}_{in} = \frac{1}{K}\sum_{k=1}^{K} P_{ink} = \frac{1}{K}\sum_{k=1}^{K}\frac{\exp(W_{in}(\beta_k))}{\sum_{j}\exp(W_{jn}(\beta_k))}, \tag{2.8}$$

where K is the number of random draws, P_{ink} is the probability of respondent n choosing product i in the kth draw, and β_k is the corresponding simulated random coefficients. In order to reduce the computational burden raised from multivariable sampling when solved using Maximum Likelihood Estimation, hierarchical Bayes models were developed by utilizing Markov Chain Monte Carlo Methods with a Gibbs sampler to estimate mixed logit model (Rossi et al., 2005, Allenby et al., 2005).

2.3.2 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:

$$L\left(y_n \mid \beta\right) = \prod_{n=1}^{N}\prod_{i=1}^{J}\left(\Pr_n(i)\right)^{y_{ni}}, \tag{2.9}$$

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, HBE 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^*:

$$pdf^*\left(\beta, b, \Sigma \mid Y\right) \propto \prod_{n=1}^{N} L\left(y_n \mid \beta_n\right) pdf\left(\beta_n \mid b^0, \Sigma^0\right) pdf\left(b^0, \Sigma^0\right). \tag{2.10}$$

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 Eqn. (2.10) 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.

2.3.3 Demand Forecasting using Discrete Choice Analysis

Estimation of the consumer choice utility function (W) allows the ***choice share***, M, for a choice alternative i to be determined by summing over the market population, N, all probabilities, P_{in}, of a sampled individual, n, choosing alternative i from a set of J competitive choice alternatives:

$$M(i) = \sum_{n}^{N} P_{in} \quad (2.11)$$

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

$$Q(i)_t = M(i) \cdot D(t) \quad (2.12)$$

2.4 Hierarchical Choice Modeling Framework

2.4.1 Challenges and Previous Work

A key challenge in choice modeling of engineering products is the modeling of the ***heterogeneous consumer preferences***. For the design of a complex engineering product like an automobile, it is important to model the diversity in consumer-preferences in a more complete way. In general, capturing consumer heterogeneity is a necessary component in understanding the perception of a design for a given population segment. 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) employ 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; to capturing consumers' perception 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.

To fully consider the impact of ***consumer preferences for individual product features***, a **Hierarchical Choice Modeling** strategy has been proposed (Kumar et al., 2009b) as shown in Figure 2.1, in which the top system level choice model only contains a reasonable set of system-level consumer-desired attributes **A** (including price *P*), while the lower level models establish the relationships between qualitative consumer perceptual attributes **A** as functions of quantitative product performance attributes **Y** and demographic attributes **S,** i.e., **A**=f**(Y, S).** In the automobile market, for instance, consumers 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. Attributes **A** considered by consumers in a choice situation may be qualitative, and require mapping to physical, measureable design attributes **Y** at the subsystem and component levels. The hierarchical choice modeling approach uses consumer **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 2.1, may be linked to engineering attributes through a series of prediction models of

ordered logit models for the subsystems, labeled $\mathbf{M_2}$ and $\mathbf{M_3}$ in the figure. Most recently, an Integrated Hierarchical Bayesian Choice Modeling framework is developed by Hoyle et al (2009, 2010), which utilizes an all-in-one estimation process to mitigate the error propagation in previous Hierarchical Choice Models.

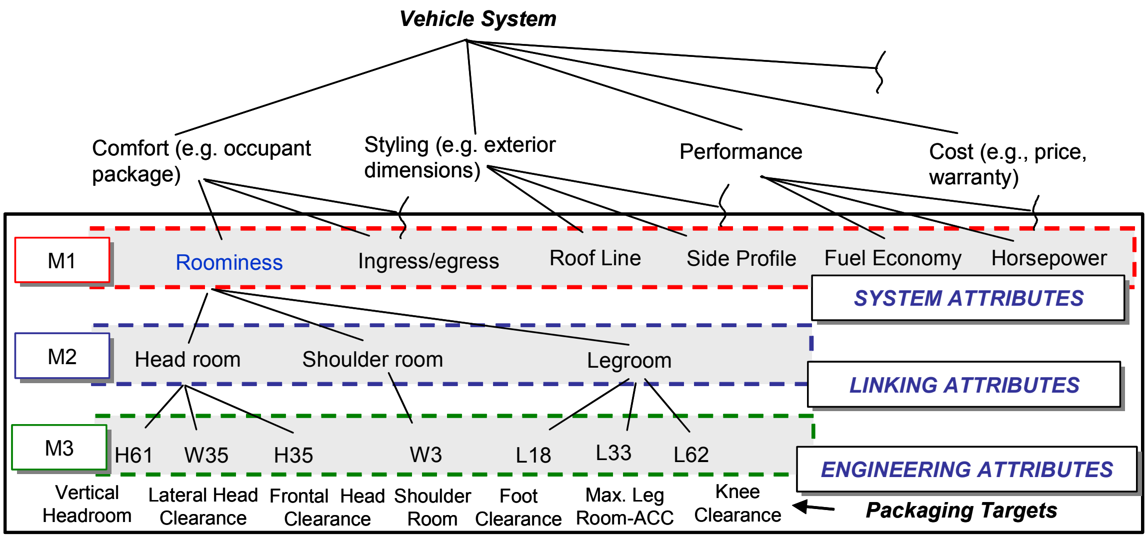


Figure 2.1: Example of Hierarchical Choice Model Approach to Vehicle Packaging (Hoyle et al., 2010)

2.4.2 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 2.2 (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).

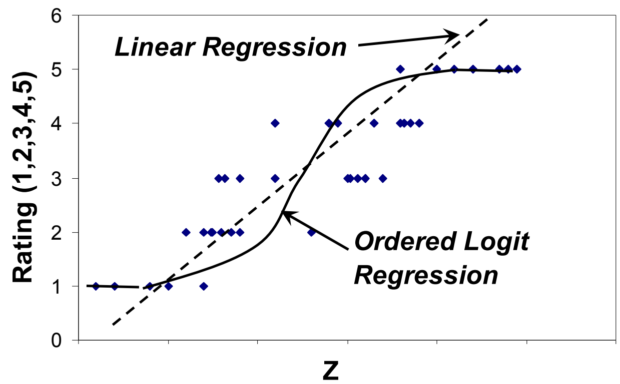


Figure 2.2: Illustration of the Variation of Ratings vs. Explanatory Variables Z (McKelvey and Zavoina, 1975)

For this reason, the **ordered logit model** is used in this work to estimate models for ordinal consumer 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 can be written as a function 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 (OL) 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 Eqn. (2.4) and (2.5). Also in the OL approach, it is assumed that the error variance is the smallest at maximum or minimum values of **Z** and the 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:

$$\begin{aligned} \Pr\left[R = R_p\right] &= \Pr\left[k_{p-1} < u_{in} < k_p\right] \qquad \left(p = 1,2,\ldots,P\right) \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \frac{e^{k_p - \boldsymbol{\beta}'\mathbf{Z}}}{1+e^{k_p - \boldsymbol{\beta}'\mathbf{Z}}} - \frac{e^{k_{p-1} - \boldsymbol{\beta}'\mathbf{Z}}}{1+e^{k_{p-1} - \boldsymbol{\beta}'\mathbf{Z}}} \qquad \left(k_0 = -\infty, k_p = +\infty\right) \end{aligned} \quad . \qquad (2.13)$$

The model parameters, $\boldsymbol{\beta}$, and cut points, **k**, are determined using the MLE or the Bayesian estimation. A **random-effects** version of the model is used in this work in which a random intercept term is used to capture the 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 consumer preferences. Ratings represent relative, or *ordinal*, preferences for an attribute, as opposed to absolute, or *cardinal*, preferences and thus require special consideration in modeling.

2.5 SOCIAL NETWORK THEORIES

With the growing public awareness of the complex "connectedness" of modern society, the idea of *social network*, in which a group of people are connected to some or all of the others following a random or particular pattern in graph, has been gaining more attention (Easley and Kleinberg, 2010, Faust and Wasserman, 1994). For example, the leading online social networking site Facebook has so far attracted more than 800 million active users (Facebook), demonstrating the power of interpersonal connections in our daily lives. There are two key elements of a social network which includes ***nodes***, representing members of the network, i.e. consumers in the context of product design, and ***links***, illustrating the connections between members, i.e. linked consumers. Depending on the specific network structure, distinctive influences through social network are observed, modeled, and researched in numerous domains including social science, humanities, etc. The meaning of "**connectedness**" encompasses two related issues in social network modeling and simulation: one is the network structure – the ***media*** of social impact; the other is behavioral interactions– the ***mechanism*** of social impact. The fundamental principals in social network theories relevant to this research are detailed in the following sections.

2.5.1 **Basic Definitions in Social Network Theories**

Nodes and Edges in Graphs. In the field of network theories, graphs are often used to represent how things are physically or logically linked to one another in a network structure, because they serve as mathematical models of network structures. A graph is a way of specifying relationships among a collection of items and consists of a set of objects, called *nodes*, with certain pairs of these objects connected by edges called *links*. In social networks, nodes are people or groups of people, and links represent some kind of social interactions. For example, the graph in Figure 2.3 consists of 4 nodes labeled A, B, C, and D, with B connected to each of the other three nodes by links, and C and D are connected by a link as well. We say that two nodes are ***friends*** if they are connected by a link. Figure 2.3 shows the typical way one draws a graph – with circles representing the nodes, and a line connecting each pair of nodes as the link. Depending on specific applications, links can be symmetric: the link simply connects two nodes to each other; or asymmetric – the link is directed from A to B, but not vice versa, as shown in Figure 2.3(b). Graphs with these two types of links are defined as undirected graph and directed graph, respectively. In general the graphs in this work will be undirected unless noted otherwise.

The Clustering Coefficient. The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other (Watts and Strogatz, 1998). In other words, it is the fraction of pairs of A's friends that are connected to each other by links. For example in Figure 2.3(a), because there exists only one single C-D link among the three pairs of friends, A-C, A-D, and C-D, the clustering coefficient of node B is 1/3. In general, the clustering coefficient of a node ranges from 0 (when none of the node's friends are friends with each other) to 1 (when all of the node's friends are friends with each other). According to the principle of "*triadic closure*", if two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future (Rapoport, 1953). If nodes A and C have a friend B in common, as shown in Figure 2.3(a), then the formation of a link between B and C produces a situation in which all three nodes A, B, and C have links connecting each other – a structure we refer to as a triangle in the network. As a result, the more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be.

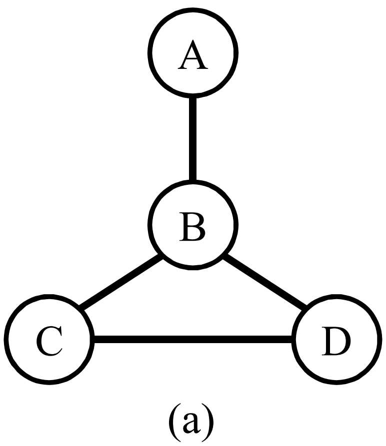


Figure 2.3: Two Graphs: (a) An Undirected Graph, and (b) A Directed Graph

Paths. One of the common elements in the use of graphs across different areas is the idea that things often travel across the links of a graph, moving from node to node in sequence – this could be a passenger taking a sequence of airline flights, a piece of information or a type of virus being passed from person to person in a social network, or a computer user or piece of software visiting a sequence of web pages by following links. This idea motivated the definition of a *path* in a graph: a path is simply a sequence of nodes with the property that each consecutive pair in the sequence is connected by a link. Sometimes it is also useful to think of the path as containing not just the nodes but also the sequence of links connecting these nodes. For example, the sequence of nodes A, B, C, is a path in the graph from Figure 2.3(a), as is the sequence of A, B, D, or A, B, C, D.

Average Path Length. Related to the definition of path is the calculation of average path length. Consider a graph G with a set of M nodes n_1, n_2, …, n_M. Let $d(n_1,n_2)$ denotes the shortest distance between n_1 and n_2. Assume that $d(n_1,n_2)=0$ if $n_1=n_2$ or n_2 cannot be reached from n_1. Then the average path length l_G is defined in Eqn. (2.14).

$$l_G = \frac{1}{M(M-1)} \sum_{i,j}^{M} d(n_i, n_j) \quad (2.14)$$

where M is the total number of nodes in the graph.

Homophily. One of the most basic notions governing the structure of social networks is homophily – the principle that we tend to be similar to our friends. Typically, your friends don't look like random sample of the general population: viewed collectively, your friends are generally similar to you along racial and ethnic dimensions; they are similar in age, the places they live, their occupations, their levels of affluence, and their interests, belief, and opinions. Clearly most of us have specific friendships that cross all these boundaries; but in aggregate, the pervasive fact is that links in a social network tend to connect people who are similar to one another. This observation has a long history; as noted by McPherson, Smith-Lovin, and Cook (2001) in their extensive review of research on homophily, the underlying idea can be found in writings of Plato ("similarity begets friendship") and Aristotle (people "love those who are like themselves"), as well as in proverbs such as "birds of a feather flock together".

Selection. In the case of immutable characteristics such as race or ethnicity, the tendency of people to form friendships with others who are like them is often termed selection, in that people are selecting friends with similar characteristics. Selection may operate at several different scales, and with different levels of intentionality. In a small group, when people choose friends who are most similar from among a clearly delineated pool of potential contacts there is clearly active choice going on. In other cases, and at more global levels, selection can be more implicit. For example, when people live in neighborhoods, attend schools, or work for companies that are relatively homogeneous compared to the population at large, the social environment is already favoring opportunities to form friendships with other like oneself. The process of selection is the basis for the stage of social network construction in Section 6.3.1.

Social Influence. When we consider how immutable characteristics interact with network formation, the order of events is clear: a person's attributes are determined at birth, and they play a role in how this person's connections are formed over the course of his or her life. With characteristics that are more mutable, on the other hand – behaviors, activities, interests, beliefs, and opinions – the feedback effects between people's individual characteristics and their links in the social network become significantly more complex. The process of selection still operates, with individual characteristics affecting the connections that are formed. But now another process comes into play as well: people may modify their behaviors to bring them more closely into alignment with the behaviors of their friends. This process has been variously described as socialization (Kandel, 1978) and social influence (Friedkin, 2006), since the existing social connections in a network are influencing the individual characteristics of the nodes. Social influence can be viewed as the reverse of selection: with selection, the individual characteristics drive the formation of links, while with social influence, the existing links in the network serve to shape people's (mutable) characteristics, such as behavior, attitudes, etc. In the context of product design, consumers are often influenced by their friends' choices when they purchase a new product. The process of social influence is the basis for the stage of social impact evaluation in Section 6.3.2.

2.5.2 Small-World Phenomenon As Media

As one type of social impact media, i.e. network structure, Small-World network has gained much attention in the social network studies. In 1998, Duncan Watts and Steve Strogatz (1998) argued that model with Small-World phenomenon follows naturally from a combination of two basic social-network ideas: ***homophily*** (the principle that we connect to others who are like ourselves) and ***weak ties*** (the links to acquaintances that connect us to parts of the network that would otherwise be far away). Homophily creates many triangles, while the weak ties still produce the kind of widely branching structure that reaches many nodes in a few steps.

Watts and Strogatz made this proposal concrete in a very simple model that generates random networks with the desired properties. Consider their original ring lattice example in Figure 2.4, we suppose that everyone lives on the ring - we can imagine the ring as a model of geographic proximity, or potentially some more abstract kind of social proximity, but in any case a notion of similarity that guides the formation of links. Figure 2.4 shows the set of nodes arranged on the ring; we say that two nodes are one step apart if they are directly adjacent to each other in either the clockwise or counterclockwise direction.

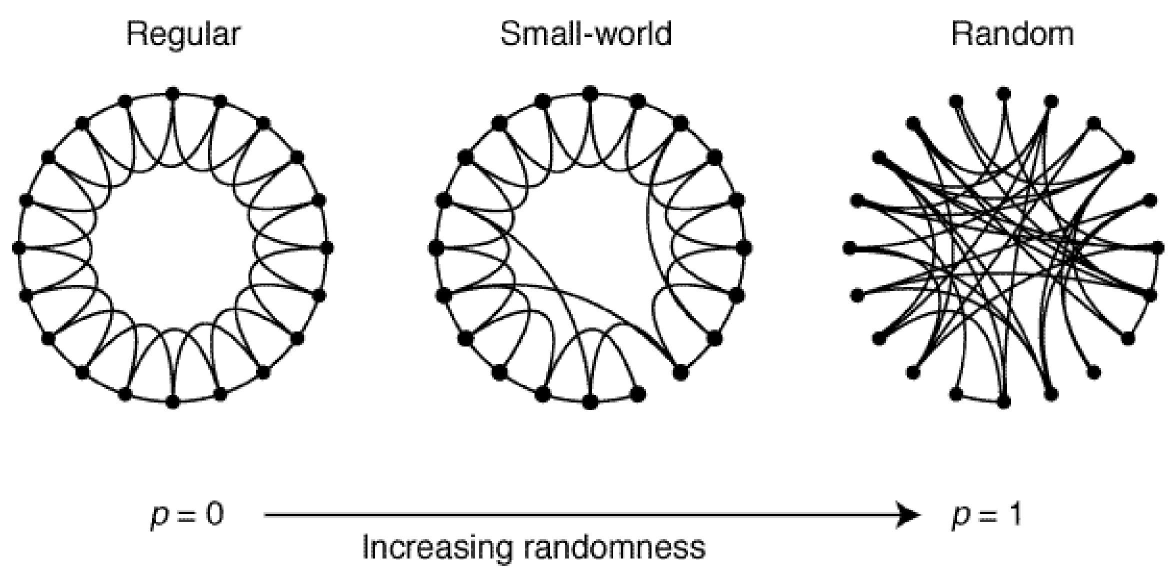


Figure 2.4: Ring Lattice With Increasing Randomness:
(a) Regular, (b) Small-world, (c) Random (Watts and Strogatz, 1998)

We now create a network by giving each node two kinds of links: those explainable purely by homophily, and those that constitute weak ties. Homophily is captured by having each node form a link to all other nodes that lie within a radius of up to *r* steps away, for some constant value of *r* (*r*=2 in Figure 2.4(a)): these are the links you form to people because you are similar to them. Then, for some other constant value *p*, each node also forms a link to *k* other nodes selected uniformly at random from the ring – these correspond to weak ties, connecting nodes who lie very far apart on the ring.

Figure 2.4(b) gives a schematic picture of the resulting network – a hybrid structure consisting of a small amount of randomness (the weak ties) sprinkled onto an underlying structured pattern (the homophilous links). Watts and Strogatz observe first that the network has many triangles: any two neighboring nodes (or nearby nodes) will have many common friends, where their neighborhoods of radius *r* overlap, and this produces many triangles. But they also find that there are – with high probability – very short paths connecting every pair of nodes in the network. A mathematically precise version of this argument was carried out by Bollobas and Chung (1988), who determined the typical lengths of paths that it implies. Moreover, Easley and Kleinberg (2010) pointed out that a surprisingly small amount of randomness is needed to achieve the same qualitative effect, which is the crux of the Watts-Strogatz model: introducing a tiny amount of randomness – in the form of long-range weak ties – is enough to make the world "small", with short paths between every pair of nodes.

Going back to the product design, this small-world phenomenon implies that people not only consider the choices of close friends, but are also influenced by remote contacts such as online reviews from people outside the regular social proximity. As literature has shown that many empirical networks exhibit the small-world phenomenon, the Watts-Strogatz method is used in the case study to simulate the network structure, in lieu of the social network data.

2.5.3 Contagion Theories As Mechanism

In the past few decades, many research works have been done to explain the emergence of social networks based on individuals' perception about other people and the relations among those individuals, under which the contagions theories are most relevant to our interest in product design. Seeking to explain networks as media for "infectious" attitudes and behavior, *contagion* theories are based on the assumption that the opportunities for contact provided by communication networks serve as a mechanism that exposes people, groups, and organization to information, attitudinal messages, and the behavior of others (Burt, 1987, Contractor and Eisenberg, 1990). This exposure increases the likelihood that network members will develop beliefs, assumptions, and attitudes that are similar to those of others in their network (Carley, 1991). Contagion mechanisms have been used to explain network members' attitudes as well as behavior on the basis of information, attitudes and behavior of others in the network to whom they are linked. Cohesion and structural equivalence models offer alternative, and in some cases complementary, explanations of the contagion process. Contagion by cohesion implies that the attitudes and behaviors of the others to whom they are directly connected influence network members. Contagion by structural equivalence implies that people who have similar structural patterns of relationships within the network influence each other.

Contagion mechanism seeks to explain a focal person's behavior, based on the behavior of other persons in the network and the relations through which these other individuals' behavior "infects" the ones of the focal consumer. Assume that the focal person is *i* and each other person is *j*. The persons' behavior y may be their attitudes, choices, or other practices. A contagion mechanism would propose that the value of a focal person's behavior, y_i, is contagiously influenced by the values of the attribute, y_j, of other people in the network. Further, the extent to which the focal person is influenced by each other person's behavior is determined by the strength of the focal person *i*'s relation, l_{ij}, with each of other persons. In other words, the contagion theories would posit that the value of the focal person *i*'s behavior y_i is a function of the combined influence of each other person *j*'s behavior and the relation, l_{ij}, between *i* and *j*. For example, the primitive formulation is presented in Eqn. (2.15), where the behavior y_i for person *i* is contagiously influenced by the sum of the behavior y_j for each other person *j* weighted by person *i*'s relation, l_{ij}, with that person *j*.

$$y_i = f(\sum_j l_{ij} y_j) \tag{2.15}$$

Several researchers have examined the extent to which contagion explains network members' attitudes toward technologies. Fulk, Schmitz, and Ryu (1995) found that organizational members' perceptions and use of an electronic mail system were significantly influenced by the attitudes and use of the members' supervisors and five closest co-workers. Later, Fulk (1993) found that social influences was even more pronounced in more cohesive groups. The attitudes and use of other members in their communication networks significantly influenced individuals' attitudes and use of an electronic mail system. Rice, Schmitz, and Torobin (1990) found that individuals' use of e-mail in a decentralized federal agency was predicted by the use of technology by others in their communication network. Further, groups of individuals who communicated more strongly with one another were more likely to share similar e-mail usage patterns. Using longitudinal data from a federal government agency, Burkhardt (1994) found that individuals' general attitudes and behaviors of a recently implemented distributed data processing computer network itself were more influenced by the attitudes and behaviors of those with whom they shared similar communication patterns, that is, contagion by structural equivalence.

In the realm of product design, new product adoption can be contagious under many circumstances. For instance, seeing a friend driving an alternative fuel vehicle will likely increase the probability that a person purchase an alternative fuel vehicle as well. While contagion theories are widely applied in many cases to capture interpersonal influences on attitudes toward technology, no work exists to integrate the social network influences through contagion into the traditional choice modeling framework, that is, social impact is modeled as an explicit term in the utility function which guides consumers' rational choice behavior.

2.6 Agent-based Model and Simulation

Agent-based modeling (ABM) is a powerful modeling and simulation technique that have been applied to a number of applications, including natural, social, and engineered systems. In ABM, a system is modeled as a collection of autonomous decision-making entities called agents, each of which individually assesses its situation and makes decisions on the basis of a set of rules. ABM features complex interactions between agents which relies on the power of computer simulations to explore dynamics out of the reach of pure mathematical methods (Epstein and Axtell, 1996, Axelrod, 1997).

To study the interpersonal influence through social network upon consumer behavior, ABM is applied to cases where consumers are influenced by their social context, that is, what others around them do. Farrell (2000) developed a synthetic world populated by virtual agents to predict how and when hit movies happen. Said et al. (2002) created a consumer behavioral model based on a set of behavioral primitives such as imitation, conditioning and innovativeness, which are founded on the new concept of behavioral attitude. Using ABM, Delre et al. (2007) investigated the role of promotional strategies upon the diffusion of a new product in the early stage of its life cycle. Deffuant et al. (2005) proposed an individual-based model of innovation diffusion with mixing social values and individual benefits.

In recent years, ABM gained popularity in modeling the vehicle market. Sullivan et al. (2005) portrayed the market interaction among the major three players, consumer, manufacturer and government. Their study suggested that consumers make purchasing decisions based on their own personal attributes as well as vehicle attributes. Later on, Sullivan et al. (2009) developed an agent simulation for modeling market penetration of plug-in hybrid electric vehicles' (PHEVs). While these approaches provide a foundation for ABM in studying consumers' choices under social context, their focuses are mainly on the marketing attributes, such as price. The linkage between product design and agents' behavior in the market is missing, and an enhanced agent-based modeling is needed to bridge the gap between the domains of engineering, marketing, and social science. As will be demonstrated in the Chapter 6, ABM is used to simulate the social network structure in lieu of the social network data at the individual level.

2.7 Statistical Data Analysis and Processing of Survey Data

Market data from multiple sources require analysis prior to creating discrete choice models due to the correlation of multiple attributes, 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 usage and social-context based choice modeling approach, must be examined.

Chapter 3

CHOICE MODELING INCORPORATING RATING DATA

3.1 INTRODUCTION

The use of ratings as a subjective measure of consumer preference is quite prevalent in market surveys, such as consumer satisfaction surveys (CSS), due to their simplicity and straightforwardness. As opposed to quantitative measures of performance, such as horsepower, fuel economy, etc., ratings function as a subjective measure for a consumer to describe his/her *perception* of performance. Ratings have been prevalently used in many other fields. For example, in market research, the increasing availability of consumers' preference ratings for various products provides the possibility of exploring consumer product ratings for use in a customized marketing strategy (Cheung et al., 2003). In psychological experiments, ratings have been prevalently applied as a psychological attitude measurement (Ajzen, 2005). In the engineering design domain, human appraisal experiments, where respondents were asked to rate overall and individual performance of product attributes, have been studied extensively in support of building preference models for engineering design optimization (Hoyle et al., 2009). However, rating data are subject to measurement error due to their subjective and discrete nature (Bound et al., 2001). Survey research practitioners have long commented that respondents vary in their usage of the scale and sometime exhibit inconsistency in rating. Another issue with using rating data in choice modeling lies in the unique nature of the choice model structure. In the choice model context, we assume that each consumer gives a rating for each alternative in the choice set. However, this is seldom the case in a CSS setting, as respondents in such surveys are often asked to rate the product they chose rather than all products they considered. Hence, method is needed to address this issue of limited rating data for the products not purchased.

The objective of this research is two-fold: first, a close examination of CSS data with respect to its applicability to consumer choice modeling is provided; and second, a systematic mixed logit based choice modeling procedure is developed to incorporate the use of both quantitative engineering attributes and subjective rating measures in the choice utility function to capture consumers' perception.

3.2 State of the Art in Rating Study

Survey questions that use a discrete rating scale to measure qualitative attributes are commonplace in supporting product design. Examples in the engineering domain include a willingness-to-pay study in a vehicle purchase survey (Kumar et al., 2009b), and human appraisal experiments in which respondents are asked to rate the interior roominess of certain vehicle configurations (Hoyle et al., 2009). While ratings have the benefit of being straightforward and simple, they are a subjective measurement in nature. Self-reports of consumer satisfaction invariably possess distributions that exhibit a positivity bias, i.e. self reports generally result in high ratings (Peterson and Wilson, 1992). Thus, measurement errors (Bound et al., 2001), bias from heterogeneous respondents, as well as different rating scale usage style (Rossi et al., 2001) are common issues in rating studies.

3.3 Characteristics of Rating Data in the Consumer Satisfaction Survey of Vehicle Purchase

In order to fully explore the potential of incorporating rating data in choice modeling, we must understand the characteristics of rating data. To facilitate this discussion, the Vehicle Quality

Survey (VQS) collected by J.D. Power and Associates is used; however, the methods proposed here apply equally well to any CSS data to be used in building a choice model (for example cell phone CSS or transportation modes evaluation).

In 2007, over 90,000 respondents participated in the VQS, rating more than three hundred different vehicles. Hundreds of questions are asked in terms of the quality, reliability and problems of the vehicle. All ratings are made on a scale of 1 to 10 in this survey. The list of 10 rating questions is shown in Table 3.1. In order to understand the characteristics of this data set, the ratings of 9 subjective measures and the overall vehicle rating are studied, and the following common issues with the CSS data are identified.

Table 3.1: List of 10 Rating Questions in VQS

	Rating Questions in VQS
1	Overall rating of attractiveness of your vehicle's exterior
2	Overall rating of attractiveness of your vehicle's interior
3	Overall rating of vehicle's storage and space usage
4	Overall rating of audio/entertainment/navigation system
5	Overall rating of vehicle's seats
6	Overall rating of heating, ventilation and air conditioning
7	Overall rating of vehicle's driving dynamics
8	Overall rating of vehicle's engine/transmission
9	Overall rating of visibility and driving safety
10	Taking into consideration all aspects of your new vehicle, please rate your new vehicle overall

3.3.1 **Ownership Bias**

Due to the subjective nature of rating measurements, we first look at a histogram of rating scores to check whether there exists a potential bias in the VQS rating data. As shown in Figure 3.1, there is an increasing trend from 1 to 10 in the percentage of a rating score, which reaches a maximum of 25% at a score of 10. Rating scores of 9 and 10 together added up to 45%, or approximately one half of the total ratings, while the lower half of the 10-point discrete rating scale (i.e. ratings 1 to 5) represents only 10% in total. There are various explanations for this. One would be the observation that asking respondents to rate their own recently purchased vehicle leads to ownership bias. It is reasonable that respondents love their own vehicles, or at least, would say good things about their vehicles in a quality survey because of the significant financial outlay in purchasing a vehicle. From this point of view, potential ownership bias in rating data could influence the coefficients estimation results when used in choice modeling.

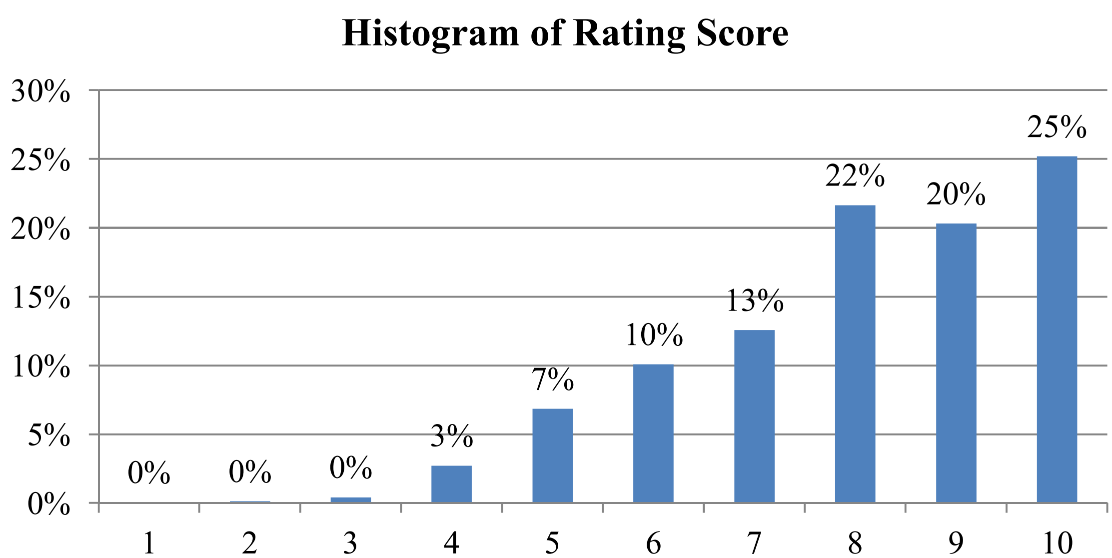


Figure 3.1: Histogram of Rating Scores

3.3.2 Differences in Rating Style

Researchers have long observed that respondents vary in their usage of the rating scale (Rossi et al., 2001). Common patterns include using only the middle of the rating scale or using the upper or lower end. The differences in scale usage can impart biases to correlation analysis. Hence, the range of responses for each of the over two thousand respondents versus the median of the 10 measures mentioned above is plotted in Figure 3.2. In the range versus median plot, a point in the lower right corner (high median and small range) indicates the respondent uses the upper end of the rating scale, while a point in the lower left corner (low median and small range) means the respondent uses the lower end of the rating scale. Whereas all responses are integer, the points are "jittered" slightly so that the number of respondents at any given combination of range (0-9) and median (1-10) can be gauged. Figure 3.2 shows considerable evidence of differences in rating style, also called *scale usage heterogeneity*. A number of respondents are using only the top end of the scale, which is represented by points in the lower right hand corner of the figure. In fact, a reasonably large number give only the top rating response (10) to all questions. On the other hand, there are very few respondents who use the lower end of the scale (lower left hand corner) and a large number who use much of the scale (points in the middle of the figure). These differences in rating styles must be accounted for in the modeling process, since these differences result in bias when fitting a choice model or calculating a variance-covariance matrix.

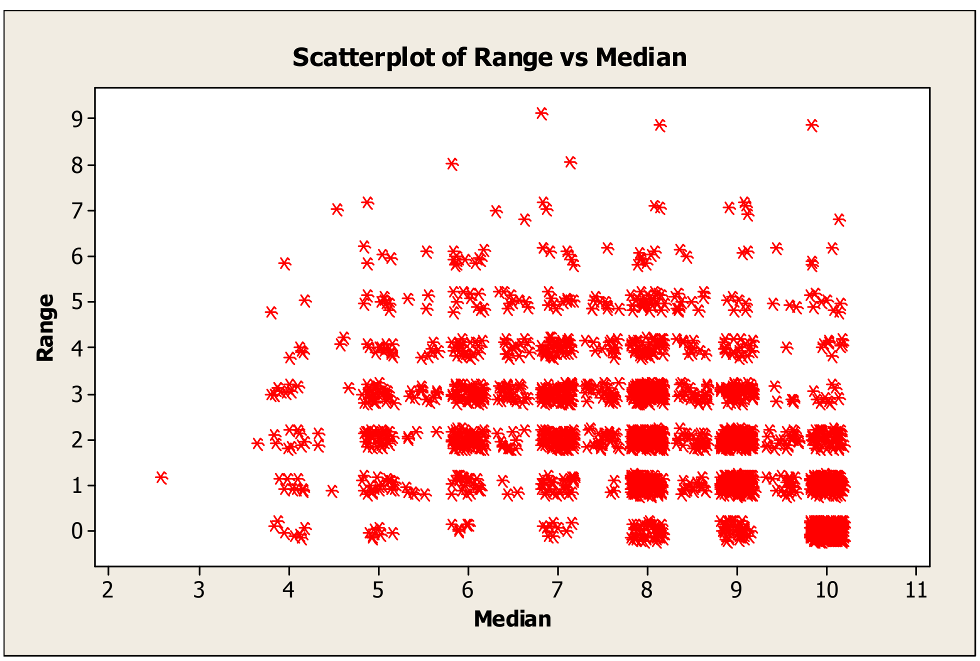


Figure 3.2: Rating Range Versus Rating Median (Jittered Values)

3.3.3 High Correlation of Rating Responses

Since high correlation will introduce problems in model estimation, correlations among different VQS rating scores are examined. Table 3.2 provides correlation among the different ratings in the data. The correlations among the different ratings are uniformly positive and in the [0.68, 0.83] range. As we would expect, all ten ratings are highly correlated since the questions are related. High correlations may result in redundancy and suppression of estimators in the choice model and problems in achieving model convergence. Therefore, including multiple ratings in a choice model is not advisable.

Table 3.2: Correlation Matrix of Ratings

Correlation Matrix									
1	0.81	0.72	0.72	0.72	0.68	0.74	0.70	0.72	0.80
2	1	0.83	0.80	0.82	0.76	0.82	0.74	0.81	0.83
3		1	0.74	0.78	0.73	0.77	0.70	0.80	0.78
4			1	0.76	0.74	0.77	0.72	0.77	0.76
5				1	0.79	0.81	0.74	0.81	0.79
6					1	0.81	0.77	0.78	0.77
7						1	0.82	0.82	0.82
8							1	0.77	0.78
9								1	0.81
10									1

3.3.4 Lack of Correlation between Individual Ratings and Choices

Using the data, we also examine whether the ranking of individual rating scores corresponds to the ranking of choice share. To keep the rating data size to a manageable size, we have selected the 7 most popular midsize sedans in the U.S. market for demonstration. Table 3.3 shows the correlation between rating scores of the selected 7 vehicles and their corresponding choice share. As seen in the first data column, the correlations between *individual rating* scores, i.e. the ratings given by each of respondents, and choice share for ten rating questions listed in Table 3.3, are all under 0.08. The lack of correlation between individual ratings and choices made will lead to an insignificant, or even negative, relationship between ratings and choices in the choice model. In a second analysis shown in the second data column of Table 3.3, a correlation analysis between vehicle *average ratings* and choice share is conducted. The average rating for each question is the average of all ratings given by the entire survey population. It is found that the average rating scores of the same ten questions exhibit higher correlation with the choice shares, i.e. between 0.09 and 0.64. Since all correlations are positive, it indicates a reasonable relationship in which increasing a vehicle rating for a given measure will increase its choice share. This finding lays the foundation for the proposed Consumer Satisfaction Index (CSI), as detailed in Section 4, for mitigating the effects of individual-level rating styles.

Table 3.3: Correlation between Ratings and Choice Share

Rating Questions	Individual Rating	Average Rating
Style / Exterior	0.0113	0.5731
Interior	0.0277	0.3922
Storage & Space	0.0137	0.1007
Audio / Entertainment / Navigation	0.0759	0.6229
Seats	0.0710	0.6391
HVAC	0.0068	0.0923
Driving Dynamics	0.0248	0.3233
Engine / Transmission	0.0565	0.3856
Visibility and Driving Safety	0.0527	0.4061
Vehicle overall	0.0411	0.4703

3.3.5 Lack of Ratings for Choice Set Alternatives

The biggest difficulty encountered in incorporating rating data in a choice model is the lack of ratings for choice set alternatives. As presented in earlier section, the discrete choice model assumes respondents' utility maximization behavior. The utility of choice alternative i for respondent n is a function of consumer-desired attributes **A** and respondent related attributes **S**. Consumer-desired attributes **A** of both the chosen and not chosen alternatives are necessary to estimate the coefficients in a choice model. Since, in this case, ratings are given only for the chosen vehicles, they cannot be *directly* used in discrete choice model.

3.4 INTEGRATED MIXED LOGIT MODELING PROCEDURE

To address the common issues of CSS data presented in Section 3.3, an integrated mixed logit modeling procedure is proposed. This integrated mixed logit modeling (IMLM) approach incorporates consumer satisfaction index to replace the subjective ratings from a CSS in choice modeling. It takes into account heterogeneity in consumer preferences in the form of the mixed logit model. The modeling procedure is shown in Figure 3.3.

As shown in Figure 3.3, the 4-step IMLM procedure starts with identifying key consumer-desired attributes **A**. In designing a large scale complex engineering system such as a vehicle, hundreds of engineering attributes may enter the model as design variables; however, including a large number of attributes in the model is unrealistic due to high computational cost and colinearity among multiple attributes. Therefore, a set of most important consumer-desired attributes **A** needs to be identified first. In the vehicle application, this is identified based on the answers to the survey question - “Which of the following were the most important factors in your choice of make and model?”. As shown in Figure 3.4, “Reliability / Durability” is identified to be the most important followed by the “Interior comfort”. “Exterior styling”, “Quality of workmanship”, “Performance”, “Like the image this vehicle portrays”, “Gas mileage”, and “Low Price”.

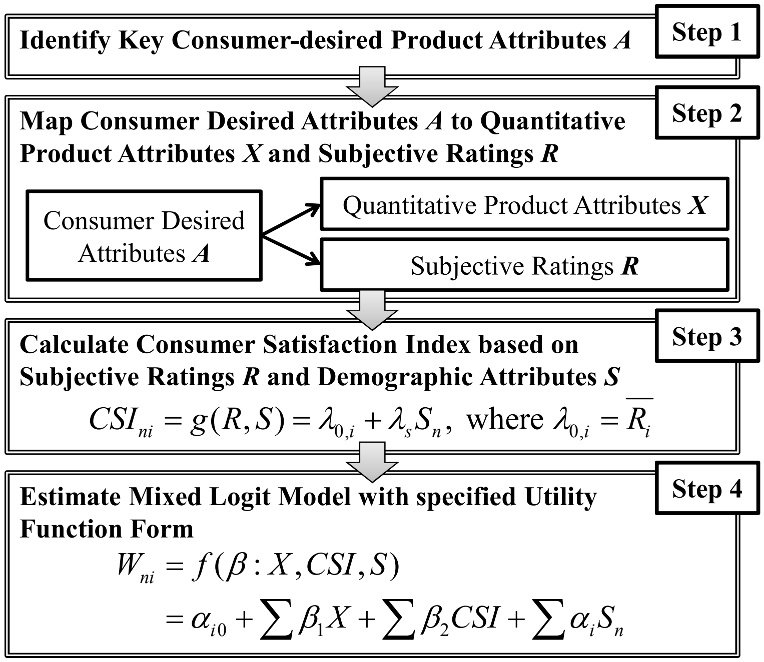


Figure 3.3: Integrated Mixed Logit Modeling Procedure

In Step 2, along with available rating data and product attributes, consumer-desired attributes **A** are mapped to either the subjective ratings **R**, representing the consumers' perception of qualitative attributes, or the quantitative product attributes **X**. During the mapping process, the principles of the Product Attribute Function Deployment (PAFD) method developed in our earlier work (Hoyle and Chen, 2009) are followed to link consumer-desired attributes to engineering attributes of interest. In the VQS example, subjective ratings **R** are presented in the form of rating score for styling, quality, etc., while quantitative product attributes **X** consist of engineering design attributes, for example front headroom dimensions, horsepower, etc., as illustrated in the house of quality matrix structure in Figure 3.4. Six quantitative product attributes include price, length/width, front headroom, rear legroom, torque, and Mileage Per Gallon in highway, and the remaining two are subjective ratings for styling and quality in the VQS data.

Consumer-desired Attributes A	Quantitative Product Attributes X						Subjective Ratings R	
	Front Headroom	Rear Legroom	Length / Width	Torque	Mileage per gallon	Price	Styling	Quality
Reliability / Durability								x
Quality of workmanship								x
Interior comfort	x	x						
Exterior styling			x					
Performance				x				
Like the image this vehicle portrays							x	
Gas mileage					x			
Low Price						x		

Figure 3.4: Step 2 - Mapping Consumer-desired Attributes A to Quantitative Product Attributes X and Subjective Ratings R

The consumer satisfaction indices **CSI** are introduced in Step 3 to alleviate the issues associated with rating data presented in Section 3.3. In the simplest case, the consumer satisfaction index is calculated using linear regression on ratings over all respondents with brand specific constants. Further, the consumer satisfaction index can be a more complex function of ratings, which represents individual likes and dislikes. In this work, we introduce **CSI** as a function of both brand specific constants, $\lambda_{0,i}$, to capture the average satisfaction for a given vehicle brand, and demographic coefficients, λ_s, to capture individual differences in satisfaction due to consumer demographic differences:

$$\mathbf{CSI}_{in} = \mathbf{C}(\mathbf{R}_i, \mathbf{S}_n) = \lambda_{0,i} + \lambda_s \mathbf{S}_n, \text{ where } \lambda_{0,i} = \overline{\mathbf{R}}_i \quad (3.1)$$

In Eqn. (3.1), the brand specific constants $\lambda_{0,i}$, one for each brand, are estimated as an average of subjective rating scores $\mathbf{R}_i$, representing consumers' aggregated preference to a specific brand. The demographic coefficients λ_s are obtained through regression analysis to evaluate individual differences as a function of **S** attributes. The **CSI** function in Eqn. (3.1) alleviates the ownership bias issue by including the average satisfaction; it also offers a way of estimating consumer ratings for all products in the choice set by introducing the brand specific constants, meanwhile the individual differences in satisfaction are captured by creating **CSI** as a function of consumer demographic attributes **S**.

In Step 4, the deterministic portion of the choice utility function form of W is specified as a function of both the consumer satisfaction index **CSI** and quantitative product attributes **X**. To explicitly model systematic heterogeneity in consumer preference, demographic attributes **S** are included in the form of alternative specific variables, i.e. the coefficients of demographic attributes are different for each alternative. The deterministic part W_{in} in the utility function of alternative i for respondent n is shown as follows:

$$W_{in} = W(\beta : \mathbf{X}, \mathbf{CSI}, \mathbf{S}) = \alpha_{i0} + \sum \beta_1 \mathbf{X} + \sum \beta_2 \mathbf{CSI} + \sum \alpha_i \mathbf{S}_n \,. \qquad (3.2)$$

In Eqn. (3.2), α_{i0} is brand-specific constant variable for alternative i and α_i is the coefficient of consumers' demographics $\mathbf{S}_n$, which together with β s, are coefficients to be determined based on the collected survey data. In the process of model estimation, the choice set J_n needs to be identified for each respondent. A choice set can be determined by related questions in CSS, for example a list of alternative vehicles considered by each respondent. Due to the lack of such information in the VQS data, seven vehicles representing the most popular sedans in the midsize segment, such as Ford Five Hundred, Toyota Camry, and Honda Accord, are considered as the common choice set in our case study. From the historical data collected by J.D. Power and Associates, it is found that compared to other car segments, the consumers of midsize cars often pay more attention to vehicle performance during their purchase, thus facilitating the test of our approach. With a specified form of utility function W and the choice set J_n, a mixed logit model is estimated to allow taste variation cross respondents.

3.5 Case Study of Vehicle Purchase

3.5.1 Model Estimation Results

In model estimation, 1773 respondents, with seven choice alternatives each, comprise 12411 observations in total. The mixed logit model with the utility function shown in Eqn. (3.2) is estimated in R-project (R-project) to demonstrate the proposed IMLM procedure. A hierarchical Bayesian model with a non-informative prior was implemented in the BUGS-project (Lunn et al., 2000) by utilizing Markov Chain Monte Carlo Methods with a Gibbs sampler to estimate the mixed logit model. In addition, a mixed logit model without consumer satisfaction indices is estimated to serve as a comparison basis. In the following sections, the former MXL with consumer satisfaction indices is denoted as MXLw/CSI, while the comparison basis is denoted as MXLw/oCSI. Table 3.4 lists 6 product attributes X, 2 consumer satisfaction indices CSI and 5 demographic attributes S included in the MXLw/CSI models. The model MXLw/oCSI includes all engineering attributes and demographic attributes as in MXLw/CSI, but it does not include CSI.

The consumer satisfaction index for *Styling* and *Quality* comes from estimating a linear regression model based on the corresponding rating data, as shown in Eqn. (3.1). The same five demographic attributes, gender, age, height, weight and income, are included in the model. All brand specific constants $\lambda_{0,i}$ are statistically significant.

Table 3.4: List of Attributes in the Mixed Logit Model

MXL w/ CSI
Engineering Attributes X
Manufacturer Suggested Retail Price (MSRP) ($)
Vehicle Length / Vehicle Width
Front Head Room (in)
Rear Leg Room (in)
Torque (ft-lbs)
Mileage Per Gallon in Highway (mile)
Consumer Satisfaction Index CSI
Styling Rating
Quality Rating
Demographic Attributes S
Gender (0=male, 1=female)
Age (years)
Height (in)
Weight (lb)
Household Income Last Year ($)

The choice set (7 mid-size vehicle alternatives with highest market share) is assumed to be identical for every respondent in this model. All vehicle attributes are normalized and all demographic attributes are mean-centered, while the consumer satisfaction indices keep the original units of rating scores. Table 3.5 shows a sample data sheet used to estimate the choice model. In the table, *caseid* in the first column represents each respondent, while *altnum* stands for alternative number which ranges from 1 to 7 corresponding to 7 vehicles in the choice set. *Chosen* is a dummy variable indicating whether this alternative is chosen by the respondent. *Price* is normalized price, and therefore lies between 0 and 1 (1 being the highest). *Styling* represents the CSI predicted using linear regression of the vehicle styling rating over the whole sample population. *Gender* is coded 1 for female and 0 for male.

Table 3.5: Sample Data Sheet for the Choice Model

caseid	altnum	chosen	Price	Styling	Gender	…
1	1	0	0.7554	3.49	1	…
1	2	0	0.2536	3.01	1	…
1	3	0	0.5576	3.33	1	…
1	4	1	0.4679	3.10	1	…
1	5	0	0.0000	3.21	1	…
1	6	0	1.0000	3.57	1	…
1	7	0	0.0737	3.00	1	…
2	1	0	0.7554	3.51	0	…
2	2	0	0.2536	3.09	0	…
…	…	…	…	…	…	…

Table 3.6: Model Statistics for MXLw/oCSI and MXLw/CSI Models

	MXL w/o CSI	**MXL w/ CSI**
No. of Parameters	48	52
Log Likelihood at Zero	-3450.10	-3450.10
Log Likelihood at Convergence	-1771.56	-1424.15
Model Fit ρ^2	48.65%	58.72%
Hit Rate	62.32%	77.89%

In both mixed logit model estimations, 4000 MCMC draws were taken, because it requires approximately 2000 draws for the posterior distributions to become stable. The statistics for the estimated models are summarized in Table 3.6. The model fit ρ^2 changed from 48.65% in the MXLw/oCSI model to 58.72% in the MXLw/CSI model. Meanwhile, the hit rate (percentage of correctly predicted choices) increased from 62.32% in the MXLw/oCSI model to only 77.89% in the MXLw/CSI model. This clearly shows that introducing consumer satisfaction index is beneficial.

Table 3.7: Posterior Mean of Random Coefficient Estimates from the Mixed Logit Model

	MXL w/o CSI		MXL w/ CSI	
Posterior Mean	Mean	StDev	Mean	StDev
Price	-0.953	0.286	-0.944	0.251
Length/Width	1.160	0.236	0.371	0.263
Front headroom	2.164	0.271	2.855	0.227
Rear legroom	0.453	0.213	0.590	0.170
Torque	1.063	0.257	0.751	0.187
MPG highway	1.771	0.209	1.776	0.154
Styling Index	n/a	n/a	3.136	0.485
Quality Index	n/a	n/a	0.271	0.126

The random coefficient estimates for the MXLw/oCSI and MXLw/CSI models are presented in Table 3.7. The coefficients for the 6 vehicle attributes and 2 consumer satisfaction indices are assumed to be normally distributed. A non-paired t-test is conducted (Tamhane and Dunlop, 2000) on the model coefficients. All six random coefficients in MXLw/oCSI and eight random coefficients in MXLw/CSI are significant at the 0.05 level, and all of them have reasonable signs: negative for Price; positive for Length/Width, Front Head Room, Rear Leg Room, Torque, MPG Highway, Styling Index, and Quality Index. The significantly negative coefficient of Price means that in general consumers prefer a lower price, as expected. Because the vehicle attributes are normalized, we can see the relative importance of attributes by comparing the magnitude of coefficients. Based on the comparison, Styling has the largest impact among 8 attributes included in MXLw/CSI, while Front Headroom and MPG Highway follows, and the Styling Index is about eleven times as important as the Quality Index. Comparing random coefficient estimates for six vehicle attributes in two models, we can see that their magnitudes are roughly on the same scale, while little discrepancy exists including different ranking order, etc.

By introducing a dummy variable for each of the 7 brands, brand specific constants (BSC) are estimated to understand consumers' inherent preference for different brands. As shown in bold font in Table 3.8, all the BSC are significant at the 0.10 level. Meanwhile, 19 out of 30 demographic attributes' coefficients are statistically significant (bold). Vehicle A is used as the base line in the model estimation (i.e. all BSCs and demographic attributes' coefficients are 0 for Veh. A). For instance, all six coefficients for age have negative signs. That means that consumers of higher age tend to choose vehicle A among all seven, *ceteris paribus*. Similarly, gender also has an impact on choice. Females (coded as 1 in gender) prefer Vehicle A to Vehicle B, compared with male consumers, *ceteris paribus*.

Table 3.8: Coefficient Estimates of Demographic Attributes from the Mixed Logit Models

	BSC	Gender	Age	Height	Weight	Income
Veh.A	***0***	***0***	***0***	***0***	***0***	***0***
Veh.B	**0.747**	**-0.944**	**-3.943**	**-0.792**	-0.035	**-1.471**
Veh.C	**0.850**	-0.427	**-5.235**	-0.563	-0.522	0.401
Veh.D	**2.310**	**-0.803**	**-5.196**	**-1.679**	-0.433	-0.487
Veh.E	**0.608**	-0.617	**-6.910**	**-1.022**	**-2.038**	-0.558
Veh.F	**0.622**	**-0.797**	**-5.996**	-0.560	**-1.799**	**1.073**
Veh.G	**1.222**	-0.532	**-6.345**	**-0.984**	**-1.894**	**-0.844**

In both mixed logit models, one normal prior distribution is assumed for all random coefficients, since we expect that taste variation cross respondents follows an approximately normal distribution. The mean and variance of the posterior distribution are calculated. Random coefficients of vehicle attributes for the MXLw/CSI model are plotted in Figure 3.5. Price, Front Headroom, Torque, MPG Highway, and Quality Rating are approximately normally distributed. The variances of Length/Width, Rear Legroom and Styling Rating are large. One explanation is that there exists large variation among respondents in their styling preferences, as well as length/width and rear legroom perception.

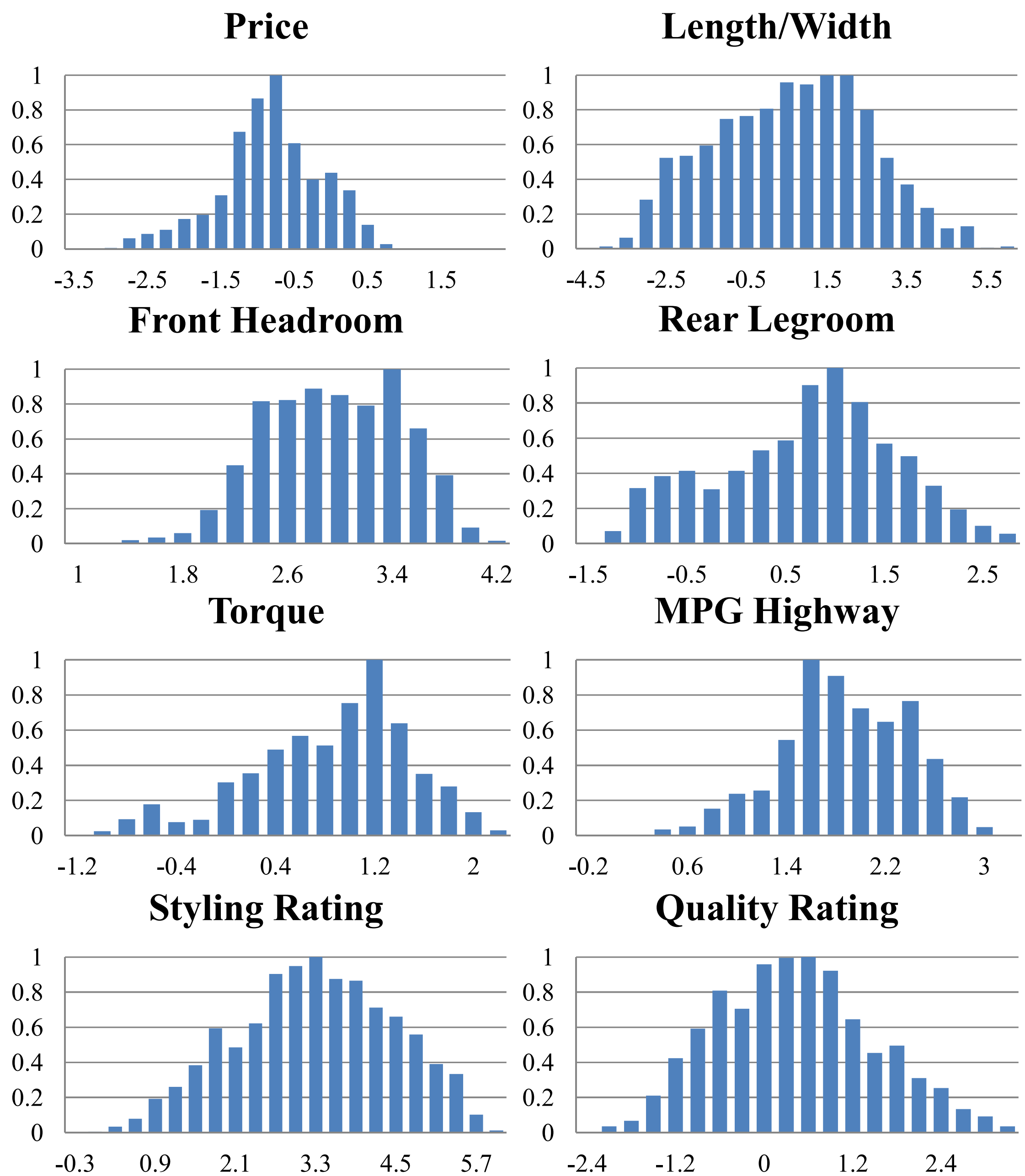


Figure 3.5: Posterior Distributions in MXLw/CSI Displaying Heterogeneity

3.5.2 Choice Share Prediction

One way to validate our mixed logit modeling procedure is to test its prediction accuracy in different market segments within the data set. Therefore, a multinomial logit (MNL) model with the same specification as MXLw/CSI is estimated in STATA (Stata Corporation, 1996-2009) and a choice share prediction test is performed for both the MNL and MXLw/CSI model results. The choice share prediction test was conducted for nine segments of combined Stature (small, medium, large) and Income (low, medium, high). Figure 3.6 plots the average prediction error in nine market segments using the MNL (dark blue columns) and MXLw/CSI (light red columns). For 8 out of 9 segments, average prediction error using MXLw/CSI is smaller, in some cases much smaller, than using MNL. In other words, MXLw/CSI exhibits better prediction ability in market segmentation tests. When the population is divided into smaller and smaller market segments, the superior prediction capability of mixed logit model will be clearer. When the population is broken down to individual levels, the choice share prediction result is identical to the hit rate.

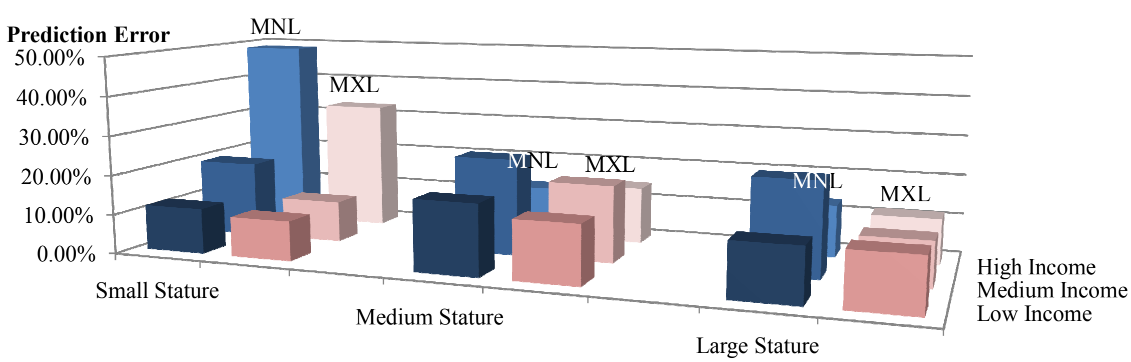


Figure 3.6: Stature / Income Market Segment Prediction Error

Using the MNL and MXLw/ CSI

3.5.3 What If Analysis In Support Of Engineering Design

To illustrate how discrete choice models can be used to study the impact of engineering design changes, "what if" analyses in two scenarios are demonstrated here.

In the first case, we consider the following "what if" design scenario: the Front Headroom of Vehicle B is increased by 1.0 inch, where the front headroom is the distance from the seating reference point to the ceiling, as marked with red oval in Figure 3.7. For a highly integrated complex system, multiple product attributes have strong dependencies which have to be accounted for in "what if" analysis. Take the vehicle design as an example, a change of front headroom may lead to a larger cross-section and therefore impact the fuel economy. Meanwhile, the consumers' perception of exterior styling and quality are also closely related the interior dimension. To take the dependency into account, regression models are built in our case study based on the historical vehicle data. Specifically, MPG is modeled as a function of price, front headroom, rear legroom, length/width, torque; *CSI* for styling and quality are modeled as functions of price, front headroom, rear legroom, length/width, torque, and MPG, respectively. Close to 100 vehicles data covering coupes, sedans, vans and SUVs are collected from Edmunds.com for regression analysis. The R square values of the three regression models are all above 0.89, indicating an acceptable goodness of fit. One thing to point out here is that the dependency relationships and the regression analysis described above are for demonstrating the what-if-analysis and therefore is not the primary focus of this work, as separate research effort is needed to accurately model the interdependencies and causal relationships of product attributes in a highly integrated complex system such as vehicles. When applied in real design problems, a separate engineering model linking vehicle attributes is needed to establish solid causal relationship between them, and at the same time, address possible issues such as multiple collinearity.

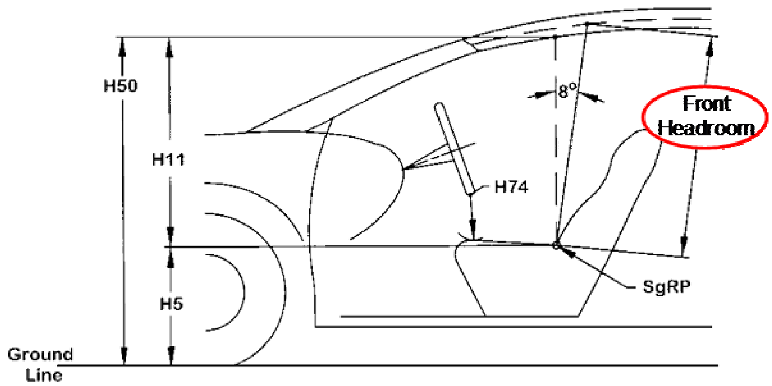


Figure 3.7: Side View of Vehicle Packaging Dimension

The new choice share is predicted using the MNL and MXLw/CSI models. As shown in Table 3.9, Vehicle B's predicted choice share change estimated with the mixed logit model is 47.20%, while the multinomial logit model predicts a smaller change of 42.41%. As discussed in Section 5.2, the MNL model is less accurate than MXLw/CSI model, because it does not take into account the taste variation across consumers. Therefore, predicted changes from mixed logit models are more trustworthy when used to predict choice share change given a change in vehicle attributes.

Table 3.9: "What If" Scenario 1 Results

	MNL			MXLw/R		
Choice share	Before	After	Change	Before	After	Change
Veh. A	0.1297	0.1304	0.54%	0.1351	0.1419	5.03%
Veh. B	**0.0790**	**0.1125**	**42.41%**	**0.0678**	**0.0998**	**47.20%**
Veh. C	0.1134	0.1100	-3.00%	0.0967	0.0952	-1.55%
Veh. D	0.1512	0.1452	-3.97%	0.1603	0.1534	-4.30%
Veh. E	0.1455	0.1370	-5.84%	0.1393	0.1258	-9.69%
Veh. F	0.1844	0.1762	-4.45%	0.1909	0.1833	-3.98%
Veh. G	0.1968	0.1887	-4.12%	0.2098	0.2007	-4.34%

Similarly, in the second case, we consider the following "what if" scenario: the Torque of Vehicle E is increased by 10 foot pounds (ft-lbs), which results in an increase in price and style rating and a decrease in quality rating, based on the regression analysis mentioned earlier. The new choice share is predicted using the MXLw/oCSI and MXLw/CSI models. As shown in Table 3.10, Vehicle B's predicted choice share change estimated with the MXLw/CSI model is 19.74%, while the MXLw/oCSI model predicts a negative change of 0.46%. The inconsistency in prediction comes from differences in the coefficients' magnitude estimated from both models, as shown in Table 3.7. As discussed in Section 3.5.1, the MXLw/oCSI model is less accurate than MXLw/CSI model, because it does not take into account the consumer satisfaction indices which measure consumers' perception of qualitative attributes. Therefore, predicted changes from MXLw/CSI models are more trustworthy when used to predict choice share change given a change in vehicle attributes.

Table 3.10: "What If" Scenario 2 Results

	MXL w/o CSI			MXL w/ CSI		
Choice share	Before	After	Change	Before	After	Change
Veh. A	0.1331	0.1329	-0.15%	0.1351	0.1323	-2.07%
Veh. B	0.0686	0.0693	1.02%	0.0678	0.0671	-1.03%
Veh. C	0.0981	0.1002	2.14%	0.0967	0.0832	-13.96%
Veh. D	0.1640	0.1641	0.06%	0.1603	0.1635	2.00%
Veh. E	**0.1517**	**0.1510**	**-0.46%**	**0.1393**	0.1668	**19.74%**
Veh. F	0.1852	0.1838	-0.76%	0.1909	0.1810	-5.19%
Veh. G	0.1993	0.1986	-0.35%	0.2098	0.2060	-1.81%

As a further refinement, a separate model which links the consumer satisfaction index *CSI* to demographic attributes **S** and product attributes **X** (rather than just brand specific constants) can be established so that we can predict change of consumer satisfaction indices with respect to engineering design changes using linear regression or the ordered logit model, as shown in (Hoyle et al., 2010). For example, an increase of the exterior dimension may result in an increase or a decrease in the styling index which also has an impact on consumers' choice. With this improvement, the advantage of the mixed logit model with consumer satisfaction indices will be further demonstrated.

3.6 Summary and Discussions

In this work, a close examination of CSS data with respect to its applicability to consumer choice modeling is first provided. Missing choice attributes in such data, the use of subjective measures such as ratings by consumers to describe product attributes only for the products they own, and multiple collinearity among many of the product attributes, are major challenges in building choice models based on CCS data. As shown in the J.D. Power and Associates' Vehicle Quality Survey data example, rating data are subjective measurements by nature and therefore face the issues of ownership bias as well as differences in rating style. High correlations of rating responses and lack of ratings for choice set alternatives prevent the direct inclusion of disaggregate (individual) rating data in a choice model.

To alleviate the limitations of CSS data, an IMLM procedure is developed; key consumer-desired attributes **A** are identified through importance ranking data and mapped to a set of corresponding quantitative product attributes **X** and subjective ratings **R**. Moreover, the consumer satisfaction indices *CSI* are introduced to incorporate subjective rating **R** in the choice model. This new approach alleviates the ownership bias issue by including the average satisfaction, allows estimations of consumer ratings for all products in the choice set, and captures the individual differences in satisfaction by creating *CSI* as a function of consumer demographic attributes **S**. Meanwhile, demographic attributes **S** are also incorporated into the utility function to capture systematic heterogeneity. Finally, a mixed logit model is implemented to allow taste variation across respondents.

The case study and subsequent discussion demonstrate many of the key advantages of the proposed approach. Most importantly, it is shown that the mixed logit model with consumer satisfaction indices successfully models the influence of rating data on consumer preference in comparison with a mixed logit model without a consumer satisfaction index. The model fit improved from 0.48 to 0.58, accompanied by a prediction error reduction from 38% to 22%. All random coefficients are statistically significant and have the correct signs as expected, which generally supports our selection of attributes in the model. In a market segment prediction test, the MXLw/CSI model has the better prediction accuracy compared to the MNLw/oCSI model. In both "what if" scenarios, the choice shares predicted by the MXLw/CSI model, MXLw/oCSI and MNL models differ significantly when used to predict the impact of product attribute changes on choice share. As the MXLw/CSI model better fits the data compared to the MXLw/oCSI and MNL models, predicted changes using the MXLw/CSI are more trustworthy.

Chapter 4

CHOICE MODELING FOR USAGE CONTEXT-BASED DESIGN

4.1 INTRODUCTION

Usage Context-based Design (UCBD) has become an area of growing interest in engineering design research. The usage context of a product refers to "all factors characterizing the application and environment in which a product is used that may significantly impact consumer preferences for product attributes" (Green et al., 2005). In other words, usage context is the set of scenarios in which a product (or service) is to be used, including the environments in which the product is used, the types of tasks the product performs, and the conditions under which the product will operate. It is proposed in this work that usage context should also be a part of the primary descriptors in the definition of a consumer profile, in addition to consumer's socio-economic status, anthropomorphic attributes, and previous product experience (Kumar et al., 2009b). Because a product will perform or be viewed differently for various usage contexts, their impacts on consumers' preferences and choice behaviors need to be studied. Even though previous works in marketing (Ratneshwar and Shocker, 1991) and engineering (Green et al., 2006) have illustrated the significance of usage context in a consumer's choice process, a general framework for quantitatively incorporating intended product usages to predict product performance and consumer's choice, does not exist, which is the focus of this research.

In this chapter, we propose the founding principles underlying a choice modeling approach to UCBD, where usage context is considered as a critical part of driving factors behind consumers' choice, in addition to consumer demographic attributes and product design attributes. In this chapter, we first provide a review of usage context influence, based on the literature from both market research and engineering design (Section 4.2). A taxonomy for UCBD is proposed in Section 4.3 by defining the important terms and their relations. Next, we discuss how the taxonomy is integrated into a step-by-step choice modeling procedure to support UCBD which captures the impact of usage context by explicitly modeling its influence on both product performances and consumer preferences (Section 4.4). Findings from both a jigsaw case study with stated preference data (Section 4.5) and a hybrid electric vehicle (HEV) case study with revealed preference data (Section 4.6) demonstrate the needs and benefits of incorporating usage context in choice modeling. Conclusions and future work are summarized at the end.

4.2 LITERATURE ON USAGE CONTEXT STUDY

Accurately capturing consumer choice is essential because it allows for the prediction of future product demand as a function of engineering design across a heterogeneous consumer population, characterized by multiple market segments. While previous works in DCA have laid the foundation for modeling the heterogeneity in consumers' choice behavior, the potential of disaggregate choice modeling in engineering design has not been fully realized due to an overreliance on marketing and demographic attributes (gender, age, income etc.) to approximate the complex drivers behind heterogeneous consumer choice. Existing choice modeling methods lose their effectiveness and fail to offer insights into *why* choices were made, because of the limited scope of consumer attributes included in the model. For this reason, it is necessary to investigate the reasons behind and the situations under which a product is being *used* to fully understand and model heterogeneous choice behavior. Hence, we must delve into a more proactive modeling approach to discover driving factors underlying consumers' choices by taking into account the *usage context* of a product. In this work, the usage context of a product is defined as "*all aspects describing the context of product use that vary under different use conditions and affect product performance and/or consumer preferences for the product attributes*". Considering the impact of usage context on product performance is a special feature of the proposed approach, where existing methods often treat product performance as "constant" across all consumers and usage contexts in choice modeling.

Marketing researchers were among the first to recognize the importance of product usage context. As Belk pointed out, *use situation* has "a demonstrable and systematic effect on current behavior" (Belk, 1974). Dickson (1982) proposed a person-situation (*usage context* in our terminology) framework in which the market is explicitly segmented by groups of consumers within usage situations. More recently, De la fuente and Guillen (De la Fuente and Guillén, 2005) studied the usage suitability of household cleaning products and their influences on purchase behavior. Although existing literature illustrated the significance of usage context in the consumers' choice process, the linkage between usage context and product performance as well as product design is missing, which calls for an innovative way to explicitly model usage context's impact with analytical methods.

4.2.1 Usage Context Literature In Market Research

In market research, use situation is defined by Belk (1974) as follows: "all those factors particular to a time and place of observation which do not follow from a knowledge of personal (intra-individual) and stimulus (choice alternative) attributes, and which have a demonstrable and systematic effect on current behavior." Belk later proposed a revised stimulus-organism-response

(S-O-R) paradigm (Belk, 1975) in which the stimulus is divided into an object and a situation, or usage context in our terminology. Relating to Belk's S-O-R paradigm, we propose here an UCBD influence diagram as illustrated in Figure 4.1.

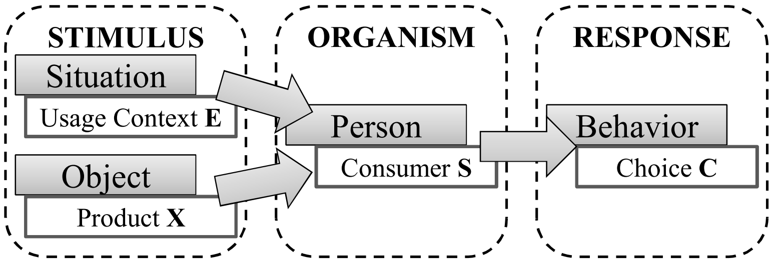


Figure 4.1: UCBD Influence Diagram based on Belk's S-O-R Paradigm

In the context of UCBD, *object* refers to product and *situation* refers to usage context. Both *usage context* and *product* act as stimulus to a consumer which leads to his/her *choice behavior*. Besides the conceptualization, Belk's categorization of five groups of situational characteristics (named *usage context attributes* in this work) (Belk, 1975) serves as the foundation for developing and classifying the usage context attributes for choice modeling (see Section 4.4).

The need for considering situational (usage contextual) variables in market segmentation was first recognized in the 1980s. Dickson (1982) pointed out that usage situation is overlooked in market segmentation and presented a person-situation segmentation framework in which the market is explicitly segmented by groups of consumers within usage situations. The work by Christensen et al. (2005) recommends stopping the common practice of segmenting consumers based on their demographics and replacing it with ways that reflect how consumers actually live their lives. The "substitution in use" approach by Stefflre (1971) was developed based on the premise that consumers think about product category instances within their functional roles in various possible usage contexts. As a further validation of this premise, in a more recent case study of snack foods, Ratneshwar and Shocker (1991) showed that products which do not belong to the same category could be considered as comparable in certain usage context, which brings up the need for constructing different choice set alternatives based on consumer profile and usage context. More recently, De la fuente and Guillen (2005) analyzed consumer perceptions with regard to the suitability of household cleaning products to anticipated usage contexts, as well as their influences on purchase behavior. In the case of multiple usage context scenarios, Berkowitz, et al. (1977) suggested aggregating an individual's given usage situation demand weighted by the situation's frequency of occurrence or importance. While their approach demonstrated the influence of usage suitability on consumer choice, the linkage between usage context and product performance, as well as product design is absent.

In the recent year, there is a growing interest on context of use in the Human-Computer Interface (HCI) field (Schmidt et al., 1999, Maguire, 2001). With prevalent use of mobile devices, context-aware applications (Green et al., 2005) as well as location-based services (Küpper, 2005) are developed to enhance the human-machine interface design of electronic products.

4.2.2 Usage Context Literature In Engineering Design

Even though the study of usage context in consumer behavior and HCI has been prevalent for years, it had not been applied to engineering design until 1990s. In Ulrich and Eppinger's product design and development book (Ulrich and Eppinger, 2003), the need for designers to envision a product's "use environment" in identifying consumer needs is emphasized. Methods have been suggested to observe a product in use as a way of gathering raw data from consumers. More recently, Green et al. published three successive papers (Green et al., 2004, Green et al., 2005, Green et al., 2006) on a frontier design method for product usage context, which is defined as a combination of application and environment in which a product will be used. A broader concept of *product design context* is constructed, consisting of three contexts that influence consumer preferences: *usage context*, *consumer context* and *market context*. Their work supports the idea that context can be differentiated based upon functional attributes, indicating a link between engineering parameters and perceived usefulness, which occurs under the influence of different usage contexts. Most recently, study on usage context attributes of HEVs (He et al., 2011a) suggested that usage context should be treated as an additional dimension of the consumer characterization process to reflect their preference heterogeneity.

Previous works on UCBD are mainly focused on qualitative analysis to support concept generation. However, the benefits of understanding usage context have great potential in an analytical design process as well. Through a choice model, we can understand the impact that usage context has on product performance and consumer preferences, and therefore optimize product design to maximize the consumer demand, or profit contributed by the product. In this work, we propose a comprehensive choice modeling framework for UCBD to quantitatively incorporate usage context into the product design process.

4.3 Taxonomy in Usage Context-based Design

As shown in the literature review section, previous works in marketing research and product design fields have employed different definitions and terminologies of usage context related variables. For instance, usage context is also called *use situation*; a usage context attribute is also referred to as a situational variable. To establish a common foundation for choice modeling in UCBD, this section is devoted to lay out our taxonomy in UCBD. The list follows the established classification in the market research domain and the specific needs associated with choice modeling. To illustrate the concepts, a jigsaw design problem is used as an example throughout this section.

4.3.1 Usage Context Attributes – E

Usage context attributes **E** refer to the characteristics or attributes used to describe the usage context. For defining "usage context", Belk (1974) stated that use situation includes all factors that influence the consumer behavior at a given time and place, except for the consumer profile and product attributes. Unlike Belk, Green et al. (2006) narrows down the scope of "usage context" to two major aspects, the application context and the environment context, and limits the influence of usage context to consumer preferences only. Usage context in real life varies significantly across product categories. In our view, its influence on consumer behavior includes the impact on product performance, choice set, and consumer preference. Hence, we define the usage context in our work as "*all aspects describing the context of product use that vary under different use conditions and affect product performance and/or consumer preferences for the product attributes*". This definition emphasizes two concepts key to our approach. First, usage context covers all aspects related to the use of a product, but excludes consumer profile (demographic attributes) and product attributes. Second, usage context influences consumer behavior through product performance, the choice set, as well as consumer preference.

One consideration to note is that, under many circumstances, it is difficult to draw a clear distinction between the consumer profile and usage context as separate sources of influence on consumers' choice. As a guideline, we refer to consumer profile attributes as those stable characteristics of a consumer that do not vary over a set of usage contexts, while those temporal, transitory characteristics of a consumer that do vary over usage contexts belong to the area of usage context. In other words, these usage context attributes change from application to application or from environment to environment over time. For example, the skill of the consumer to successfully accomplish a cutting task using a power tool, brand loyalty, and positive or negative experiences with a particular brand (Goldberg, 1995), are considered as *consumer profile attributes*, since they are more stable over time than the *usage context attributes*.

Specific to the choice modeling process, we can divide usage context attributes **E** into *performance*-related and *preference*-related, according to the way in which they impact consumer behavior. These usage context attributes **E** can be either continuous or discrete. While performance-related attributes $\mathbf{E_Y}$ influence product performance **Y**, preference-related attributes $\mathbf{E_W}$ have an impact on the choice set and consumer preference. In some cases, performance-related and preference-related usage context attributes are not mutually exclusive. For example in using a jigsaw, if the saw is to be used for cutting outdoors, the density of saw dust experienced by the user may be different than if the saw were used indoors (performance impact), whereas the user may prefer a bright saw color for outdoor use so that the saw will be easily identified if placed on the ground (preference impact). Prior knowledge of a usage context attribute's influence on preference can be used to reduce the complexity of estimating a choice model, and hypothesis testing of a usage context attribute in the choice model estimation process can be used to confirm this knowledge.

4.3.2 Usage Context Scenarios – U

Usage context scenarios **U** refer to the most common combinations of usage context attributes **E** describing common usage scenarios, which can be identified through survey and using data analysis techniques such as cluster analysis. Identifying common usages can significantly simplify the data collection process compared to surveying all factorial combinations of usage context attributes **E**. In addition to the situation that each consumer has one primary usage scenario, there are cases that multiple-usage scenarios need to be considered. The idea of usage importance indices, denoted as *F*, emerges from the need for considering multiple-usage scenarios, where a single product is used under a series of different usage scenarios. In this case, multiple usage scenarios are weighted by their usage importance indices in the range of [0, 1]. Eqn. (4.1) shows that the usage context attributes **E** and usage importance indices *F* together define the usage context scenario U.

$$\mathbf{U} = (\mathbf{E}, F) \tag{4.1}$$

The usage importance indices can be either specified by a user, or determined based on the observations of user choices under multiple-usage scenarios. In the former case, a user is asked to provide the best estimate of the importance of a particular usage in the survey. In the latter situation, the survey questionnaire is designed to identify the importance indices through choice model estimation.

4.3.3 Consumer Profile Attributes – S

The consumer profile **S** includes all stable or permanent aspects of consumer profile attributes impacting consumer choice behavior, for example, *gender*, *age*, *income* bracket, etc. In the choice modeling of UCBD, consumer profile attributes **S** may have a direct impact on consumers' preference and therefore may influence their choices. Similar to usage context attributes, consumer profile attributes **S** can be categorized into performance-related $\mathbf{S_Y}$ and preference-related $\mathbf{S_W}$ to differentiate their impact. For example, household income belongs to

$\mathbf{S_W}$, as it is expected to have a large impact on consumers' sensitivity on price: the more they earn, the less they care about the price. On the other hand, skill level of the consumer operating jigsaw is considered as a performance-related $\mathbf{S_Y}$, because jigsaw performances vary when operated by a beginner, intermediate, or experienced user.

4.3.4 Product Design Variables – X

Product design variables describe the engineering decisions involved in product design. In the jigsaw case, blade tooth height, stroke frequency, step distance between two teeth, etc., all belong to the product design variables **X**.

4.3.5 Consumer-desired Product Attributes – A

Consumer-desired product attributes **A** are defined as key product characteristics that influence consumers' choices in selecting a product. In a market survey, consumers are usually asked to rate these consumer-desired product attributes. They include not only engineering performances **Y**, but also non-engineering attributes **M**.

Engineering performance **Y** refers to all performance-related engineering attributes. Since **Y** plays a critical role in the engineering design process, engineering performance **Y** is our focus in this work. In the jigsaw example and other similar cases, engineering performance **Y** is further divided into *performance of service results* and *performance of service delivery* or transformations. The performance of service results $\mathbf{Y_r}$ represents the measures of the end performances of the resulting service, such as cutting precision, planarity, etc. On the other hand, the performance of service delivery or transformations $\mathbf{Y_t}$ represents the measures of the performances related to the delivery of the service, such as linear speed, noise, vibration, safety conditions, etc. The performances of the service delivery are no longer visible in the results once it has been delivered.

Non-engineering attributes **M** include all non-engineering aspects of consumer desires attributes, for example, price, brand, aesthetics and other common marketing measures. Price is one of the most influential non-engineering attributes **M** in consumers' choice. In practice, price can enter the utility function as a single term, or can be scaled by income or log income to reflect the connection between income and price sensitivity, as shown in the case study.

4.3.6 Consideration Set – J_n

The product choice set J_n is defined as a group of product alternatives consumers consider during their choice procedure. Simonson and Tversky (1992) showed that choices are made in the context of a consideration set, i.e. a choice set. Since only differences in utility matter due to the nature of choice models, the selection of a product choice set exhibits great impact on consumer choice. Methods for determining the appropriate choice set considering usage context are described in Section 4.4.2.

4.4 Usage Context-based Choice Modeling

In order to capture the impact of usage context attributes and utilize usage context information in a design process, a framework for choice modeling in UCBD is presented in this work. In this section, we focus on the procedure for implementing choice modeling in UCBD and discuss the potential issues involved in each phase. Our discussion follows the sequence of the four major phases for implementing choice modeling.

Phase I Collect usage context information and identify usage context attributes **E**. (**usage context identification**)

Phase II Collect consumer choice data together with choice set information J_n, consumer profile **S** and their usage context attributes **E**. In a stated preference survey, a choice experiment representing different combinations of consumer profile and usage context is designed where each respondent makes the selection among a choice set for given usage scenarios. For revealed preference data collection, all data from real consumer purchases are recorded. (**data collection**)

Phase III Create a physics-based model or a human-appraisal-based ordered logit model for predicting engineering performance **Y** as a function of usage context attributes **E**, consumer profile $\mathbf{S_Y}$, and design variables **X**. (**linking performance with usage context and consumer profile**)

Phase IV Create a choice model for market share and demand estimation (**choice model estimation**)

In the rest of this section, details for each phase are provided.

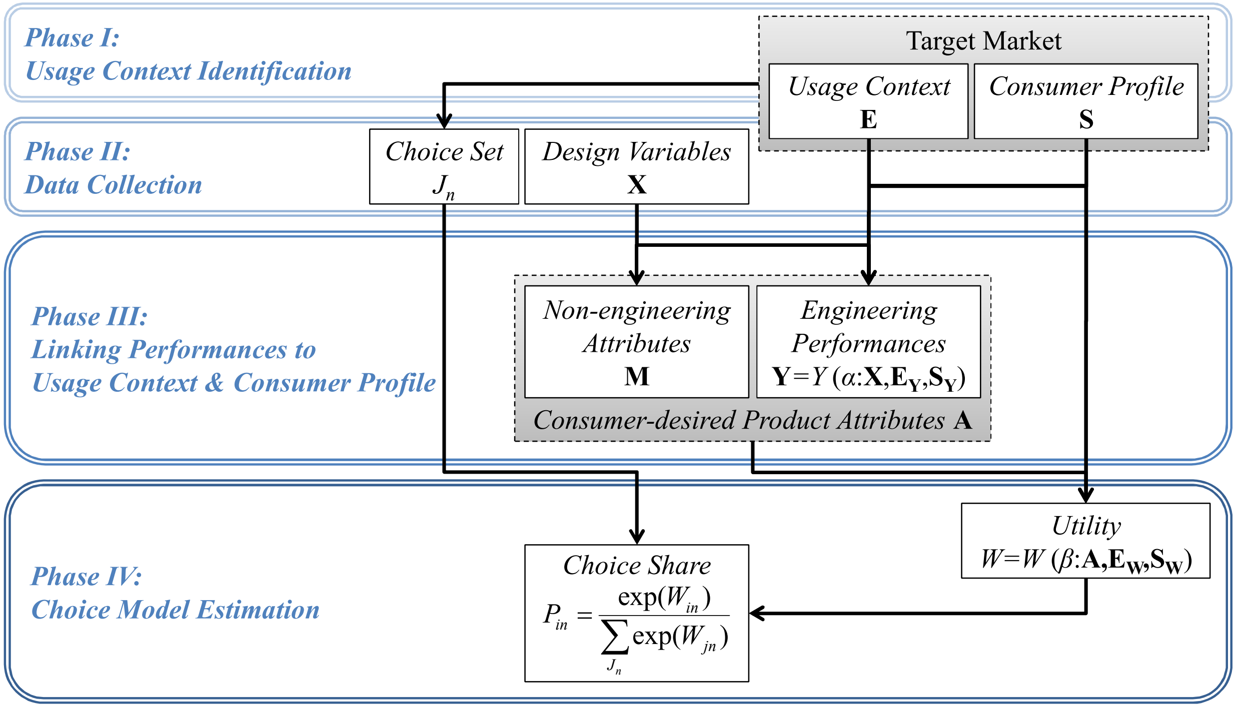


Figure 4.2: Choice Modeling for Usage Context-based Design

4.4.1 Phase I: Usage Context Identification

A successful product design requires an understanding of consumers' needs so that the products produced will match consumers' interest. Widely used survey methodologies such as focus groups, one-on-one interviews of experienced users, and observations of the product (Ulrich and Eppinger, 2003) can be used to identify important usage context attributes and common usage context scenarios among target consumers.

Following Belk's classification (Belk, 1975), usage context can be categorized into five types: *physical surroundings*, *social surroundings*, *temporal perspective*, *task definition*, and *antecedent states*. In Table 4.1, we use the jigsaw example to illustrate how the usage context attributes can be defined by following these five basic categories. It should be noted that based on Belk's classification, the scope of the usage context attributes is beyond the act of using the product, but also includes the context of purchase.

Table 4.1: Five Categories of Usage Context

Usage Context Type	Jigsaw Example
Physical surroundings	Location of cutting, Accessibility of an outlet, Availability of workbench.
Social surroundings	Presence of children, neighbors.
Temporal perspective	Expected process duration, Estimated time needed to purchase the tools in a nearby DIY store.
Task Definition	Material type, Board thickness, Minimal linear speed, Maximal vibration level, Noise and safety conditions.
Antecedent states	Set of saw tools already in possession, New life conditions or projects, Cash at disposal.

Physical surroundings are the most apparent characteristics of a usage, which include geographical location, weather condition, lighting, and other physical characteristics of a usage. In the case of using a jigsaw for cutting a board, the location where the operation must take place (indoor/outdoor), the accessibility of a power outlet, the availability of a workbench are typical examples of physical surroundings.

Social surroundings provide additional information about the social situation of a usage. Whether another person is present, his/her influence on the user, and other social characteristics belong to this category. For instance in cutting a board, one may prefer a jigsaw to a circular saw often used under these conditions, for reasons of safety and noise because of the presence of children nearby.

Temporal perspective refers to those aspects of the purchasing situation or to those of a given usage which are specific for a given range of time. For instance, the expected duration of the cutting process may be a reason for preferring a circular saw to a jigsaw or a powerful jigsaw to a more basic one (faster linear speed). In terms of purchase situation, the time and emergency aspect for buying a new tool in a nearby DIY store may also be a deciding factor under certain circumstances.

Task definition covers all features that explain the purpose of the purchase. For instance, one must consider the type of material to cut (wood, steel, etc.), the thickness of the board to cut (beyond a certain thickness, the cut is impossible), the minimal linear speed that is acceptable when the user delivers the maximal amount of pushing arm forces and wrist torques, the maximal vibration level that is tolerable, or the admissible noise and minimal safety conditions.

Antecedent states define a dimension of usage which is antecedent to the purchase. The factors for a new jigsaw acquisition may be the set of saw tools one already possesses (circular, chain, panel, bow, miter, etc.) and their respective age and expected remaining lifetimes, a new life condition or project (moving from an apartment to a house, or a house remodeling), and the cash at one's disposal.

The above-mentioned five categories of usage context attributes can be used as a checklist in the process of determining the potential usage context attributes. For stated preference surveys, as will be demonstrated in the jigsaw case study (Section 4.5), a user survey is used to collect the set of primary usage context attributes **E**. For problems with a large number of **E**, a cluster analysis becomes essential to reduce the possible combinations of **E** to a manageable size, and focus the study on a set of common usage context scenarios, i.e. the most common combinations of usage context attributes **E**.

4.4.2 **Phase II: Data Collection**

Due to the nature of data collection, surveys can be divided into Stated Preference (SP) (Kroes and Sheldon, 1988) and Revealed Preference (RP) (Samuelson, 1948). SP refers to choice experiments where respondents are presented with a set of simulated product options from which they make a choice. This approach is attractive for model building because a high level of control can be exercised over the collected data, resulting in a data set optimized for choice modeling. However, SP data does not represent real purchase behavior and such surveys require significant time and additional cost to administer, thus resulting in a limited size and quality (Louviere et al., 2000) dataset. For these reasons, it is sometimes desirable to utilize actual purchase data and consumer satisfaction surveys collected.

In a stated preference survey, a choice experiment is conducted in which consumers are asked to make a choice among several available alternatives under given usage context scenarios. Since the number of products available is usually much larger than the number of products a consumer can use and compare in a choice experiment within a reasonable amount of time, an optimal experimental design can be applied to reduce the number of products in the choice set to a feasible level. For example, a nested design of experiments (DOE) on $(J_n \mid \mathbf{S},\mathbf{E})$ can be applied to find the optimal set of choice alternatives for respondents based on their consumer profile **S** and usage context **E**. The D-optimal experiment design algorithm for human appraisal surveys (Hoyle et al., 2009) can be used to select the products to include in the choice set for best model estimation.

A try-it-out survey is highly recommended for collecting SP data, in which consumers are asked to use the products under given usage scenarios, rate the performances, and make a choice of one of the products. There are many advantages to conducting a try-it-out survey: first, handson experience is very important as it often simulates a real purchase process better for products that are typically tested prior to purchase; second, consumers experience the product under certain usage contexts, which ensures the relevance; and third, assessments of the product performance reflect consumers' perceived product performance. On the other hand, the try-it-out survey often requires more resources than a paper survey, where photos (or images) and data specification sheets are commonly used to present the products.

RP data has the advantage in that it reflects consumers' real choice behavior; however, RP data may present issues with collinearity, endogeneity, and lack of randomization. An examination of the information matrix from choice model estimation could identify the possible orthogonality issues in RP data. In some cases, consumers' ratings of product performances are also collected in RP surveys to capture consumers' perceived product performances. Such level of detail is required to model the impact of usage context on product performance and consumer preference as introduced next.

4.4.3 Phase III: Linking Performance with Usage Context and Consumer Profile

This is a unique phase for UCBD applications in which product performances $\mathbf{Y}$ are formulated as a function of performance-related usage context attributes $\mathbf{E_Y}$, performance-related consumer profile attributes $\mathbf{S_Y}$, and product design variables $\mathbf{X}$, as shown in Eqn. (4.2):

$$\mathbf{Y}_{in} = Y(\alpha : \mathbf{E}_{\mathbf{Y}n}, \mathbf{S}_{\mathbf{Y}n}, \mathbf{X}_i)\,. \tag{4.2}$$

where the coefficients α can either be established by a physics-based model or determined through a human-appraisal-survey-based regression model. The physics-based model is constructed based on the physical relations. Taking the jigsaw design as an example, a system of equations can be derived to calculate the translational force and the torque on the user's wrist to assess the user's comfort level during the cutting process as a function of wood type and thickness as well as of admissible force and torque depending on the user experience (Yannou et al., 2010). The second approach utilizes rating data given by consumers in a human appraisal survey and builds a regression model to predict the ratings of performances **Y**. While the physics-based model saves the time and cost of a survey, a human appraisal survey can be used to assess either quantitative or qualitative performance perceived by the consumers. Such surveys can be integrated into the try-it-out survey for choice modeling, as described in Phase II. The ordered logit model (McCullagh, 1980) is used for modeling the discrete rating data in the HEV case study in this work.

4.4.4 Phase IV: Choice Model Estimation

As shown in Figure 4.2, in Phase IV a predictive model of demand Q is established using Discrete Choice Analysis (DCA), a statistical technique of building probabilistic choice models (Ben-Akiva and Lerman, 1985, Koppelman and Bhat, 2006). DCA is based upon the assumption that individuals seek to maximize their personal *consumer choice utility*, u, when selecting a product from a choice set. With discrete choice analysis (Koppelman and Bhat, 2006), 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):

$$u_{in} = W_{in} + \varepsilon_{in} \,. \tag{4.3}$$

The observed part of utility for respondent n and for alternative i, W_{in}, is expressed as a function of consumer profile attributes **S**, consumer-desired vehicle attributes **A**, usage context attributes **E**, and the β coefficients, which are estimated by observing choices respondents make, As shown in Eqn. (4.4),

$$W_{in} = W(\beta : \mathbf{S}_{\mathbf{W}n}, \mathbf{A}_{in}, \mathbf{E}_{\mathbf{W}n}), \qquad (4.4)$$

where $\mathbf{A}_{in}$ denotes the consumer-desired product attributes of respondent n, alternative i, $\mathbf{E}_{\mathbf{W}n}$ and $\mathbf{S}_{\mathbf{W}n}$ denote the preference-related consumer profile attributes and usage context attributes of respondent n. The coefficients β are estimated based on the data collected in Phase III. From the observed utility, W_{in}, the probability P_{in} of an individual i choosing a given alternative n can be estimated. By following the information flow in the four-phase diagram (Figure 4.2), we can see clearly how product design variables **X,** together with the definitions of usage context **E** and consumer profile **S**, are first mapped to product performance **Y** (Eqn. (4.2)), then to deterministic utility W (Eqn. (4.3)), and finally to the probability of choice P_{in} , which can be aggregated to the total market share based on predictions for a population. This flow creates a mathematical link between product design decisions, represented by **X**, to consumer demand, represented by P_{in}.

As seen in Figure 4.3**,** the hierarchical choice modeling framework is illustrated using vehicle case study as an example to consider the impact of usage context attributes **E** by including them as the input of the choice model at the top level together with their bottom-up influence on the vehicle performance attributes which also serve as input of the choice model. Typical consumer profile attributes **S** include *gender*, *age*, *household income*, etc., while *local/highway driving condition* and *miles driven daily* are two of the commonly considered usage context attributes **E**. Consumer-desired vehicle attributes **A** refer to key vehicle features that influence consumers' choice in selecting a vehicle. The inclusion of consumer profile attributes **S** and usage context attributes **E,** in addition to consumer-desired vehicle attributes **A,** in the estimation of demand to capture the heterogeneity of consumer preference and their usage context, is the key component in the hierarchical choice modeling framework.

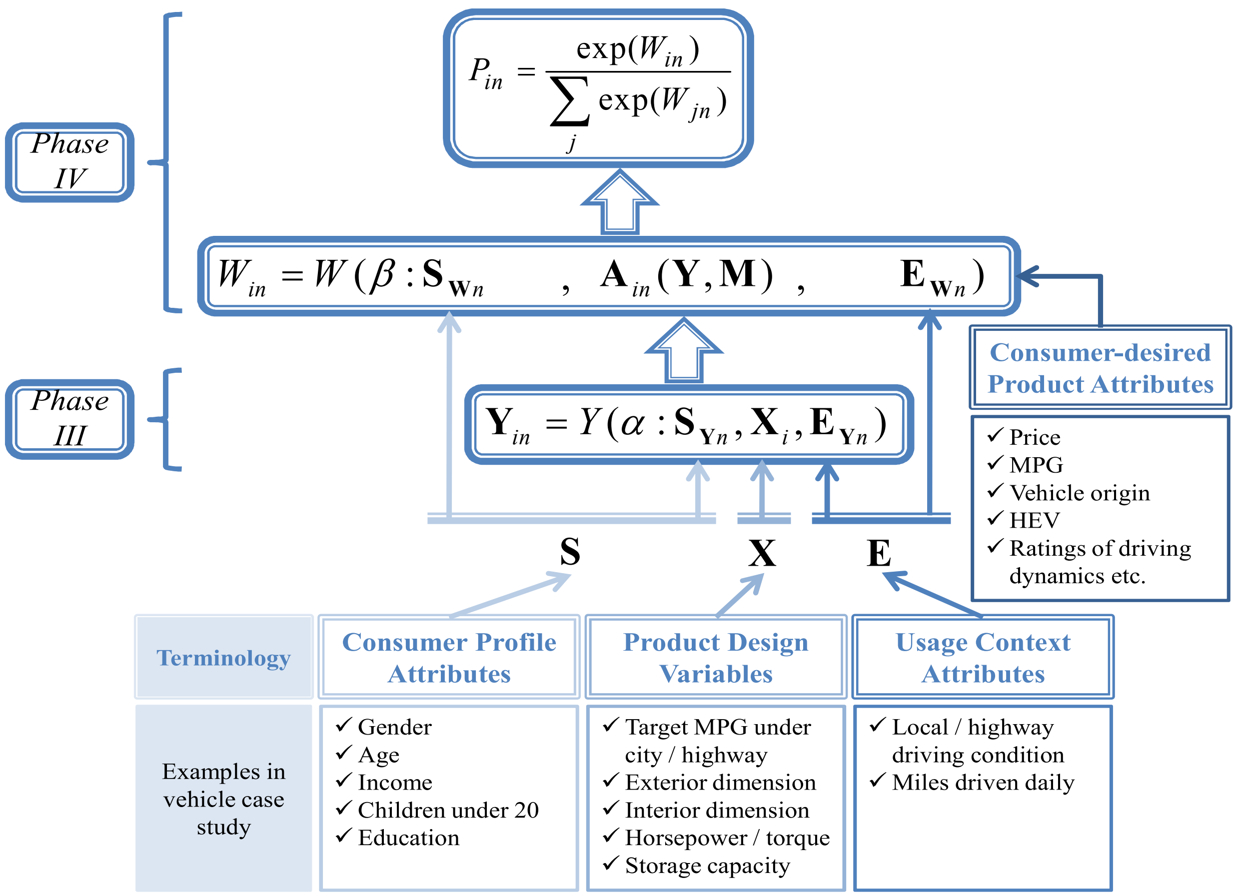


Figure 4.3: Hierarchical Choice Modeling for Usage Context-based Design

Phase IV in Figure 4.3 includes the choice probability prediction function in multinomial logit model with Type I extreme value error distribution (Train, 2003), and the deterministic portion of the choice utility as a function of **S**, **A**, and **E**. In Phase III, the bottom of hierarchy, a separate prediction model is established to link **A** with **X**, **S**, and **E**, as described in previous section. While product attributes are often fixed as constants for different consumers in conventional choice modeling, their dependence on consumers and usage context is considered here. For example, mileage per gallon (mpg) is one of the key vehicle attributes. Even though vehicle manufacturers provide target mpg measures under city and highway driving condition for each of their car models, the actual mpg value varies significantly from consumer to consumer because of the heterogeneous usage scenarios and consumer driving habits. Similarly, consumer ratings of vehicle performances are also influenced by individual profile attributes such as *gender* and *age*, as shown in Eqn. (4.2). For quantitative attributes, the above model can be expressed by physical equations. When ratings are used to measure qualitative attributes, an ordered logit model can be used due to its capability of handling discrete data (He et al., 2011b). By establishing a relation with vehicle design variables **X** through the hierarchical modeling framework**,** the obtained choice model can be used to support engineering design decisions (Wassenaar et al., 2005).

Furthermore, using the choice model created based on predominately single-usage surveys, the choice prediction can be expanded to multiple-usage cases using the following equations:

$$W_{in} = W(\beta : \mathbf{A}, \mathbf{U}, \mathbf{S}) = \sum_k W(\beta : \mathbf{M}, \mathbf{Y}^k, \mathbf{E_W}^m, \mathbf{S}) \cdot F^k , \tag{4.5}$$

$$\mathbf{Y}^k = Y(\mathbf{S_Y}, \mathbf{X}, \mathbf{E_Y}^k) , \tag{4.6}$$

where F^k is a importance measure indicating how important the usage scenario $\mathbf{U}^k$ is for the consumer, i.e., $\mathbf{U}^k = (\mathbf{E_W}^k, F^k)$; k is indicator of different usage scenarios. The above expansion is based on the assumption that the terms resulting from each usage scenario are independent from each other, and the utility function for the multiple-usage case can be treated as the weighted sum of individual usages, as Berkowitz (1977) suggested.

4.5 Jigsaw Example

In this subsection, the design of a jigsaw is used to demonstrate the implementation of the proposed usage context-based choice modeling approach with stated preference data. The jigsaw is a common power tool for cutting wood. Under different usage contexts, the jigsaw performances, as well as the consumers' preferences for the saw, change. The choice set considered in the user survey is formed by a few representative jigsaw products in the market. The four phases of choice modeling for UCBD are illustrated with the hypothetical saw design and a few representative attributes for demonstration. A choice model is built and estimated on synthetic survey data generated using a few key assumptions about consumer preferences. Results are discussed which demonstrate the proposed framework.

4.5.1 Phase I: Usage Context Identification

Phase I is completed with three tasks: *collect usage information*, *identify common usage contexts through cluster analysis*, and *identify usages context attributes*. We start with a user survey in which questions about primary usages are asked. It should be noted that the primary usage context is not limited to the most frequent usage context, and can be defined by the user. In some cases, for instance, a saw is expected to accommodate the most-demanding usage context. As described in Section 4.4.1, five categories of usage context can be used as a guideline for determining the usage context attributes. Figure 4.4 shows a small portion of the sample user survey questionnaire as an example. A few typical usage context questions for a jigsaw user would include wood type, working environment, etc.

Please answer the following question by choosing the best description of your primary saw usage context.

1. The woods you are cutting are:
   soft medium hard
2. The working environment of your cutting is:
   indoor outdoor

…

Figure 4.4: Sample User Survey Questionnaire for Phase I

In this case study, we select *wood type* and *working environment* as two usage context attributes **E** for demonstration purpose; wood type (it amounts to wood density in fact) is considered as a performance-related attribute $\mathbf{E_Y}$ that influences product performance **Y**, while both wood type and working environment are treated as preference-related attributes $\mathbf{E_W}$ with an impact on consumer preference. The *wood type* attribute is coded as 1 for soft, 2 for medium, and 3 for hard, while the *working environment* attribute is coded as 0 for indoor and 1 for outdoor. Based on the survey data of common usages, cluster analysis is performed. For our case study, indoor cutting for soft wood, outdoor cutting for medium wood, indoor cutting for medium wood, and outdoor cutting for hard wood are identified to be the most common usages (Table 4.2) based upon the results from a k-means clustering analysis (Hartigan and Wong, 1979) on the hypothetical survey data with two usage context attributes, *working environment* and *wood type*, and k (number of clusters) = 3.

Table 4.2: Common Usage Contexts Identified from Cluster Analysis

No.	Working environment E_1	Wood type E_2	Usage context description
1	0	1	indoor cutting for soft wood
2	1	2	outdoor cutting for medium wood
3	0	2	indoor cutting for medium wood
4	1	3	outdoor cutting for hard wood

4.5.2 **Phase II: Data Collection**

Various human-appraisal experiments can be utilized to collect consumers' preferences under different usage contexts as described in Section 4.4. The question lies in how to minimize the amount of surveys to cover the various attributes included in choice modeling. Here we assume that each respondent is surveyed for more than one usage context (but only one primary usage context at a time in the choice experiment). We also assume that all respondents have some level of experience with the product and are able to differentiate between the different usage contexts described in the survey questionnaire. Eight jigsaw products available in the market are considered, but only four products that are most relevant for a given usage context form the choice set J_n in each choice experiment. Three consumer profile attributes are included: *gender* (0 for male and 1 for female), *income* (annual income in \$1,000s), and *skill level* (1 for elementary user, 2 for experienced user, and 3 for professional user).

The synthetic data are simulated with 500 respondents, 4 choices alternatives, and 4 usage contexts (8,000 observations in total) based on a few key assumptions about consumer preferences. The suggested try-it-out survey questionnaire for user 1 under usage scenario 1 is shown in Appendix A: Sample Try-It-Out Survey Questionnaire for UCBCM (User 1, Usage Scenario 1). As each choice experiment has a different choice set, the products listed in the questionnaire might be different for each respondent. Table 4.3 presents three categories of attributes considered for choice modeling, including three consumer-desired product attributes $\boldsymbol{A}$ (*price*, *advance speed*, and *comfort*), two usage context attributes $\mathbf{E}$ (*working environment*, and cutting board *wood type*), three consumer profile attributes $\mathbf{S}$ (*income*, *gender*, and *skill level*), together with four design variables $\mathbf{X}$ (*blade tooth height*, *stroke frequency*, *blade translation*, and *step distance between teeth*).

Table 4.3: List of Attributes and Design Variables included in Jigsaw Case Study

Consumer Desired Product Attributes A		
M	Price	
Y_{t1}	Advance speed S_a	
Y_{t2}	Comfort level $P_{comfort}$	%
Usage Context Attributes E		
E_1	Working environment	indoor, outdoor
E_2	Cutting board wood type	soft, medium, hard
Consumer Profile S		
S_1	Income	uniform dist. , [50k, 150k]
S_2	Gender	male, female
S_3	Skill level	1, 2, 3
Product Design Variables X		
H_d	Blade tooth height	
F	Stroke frequency	
A	Blade translation	
s	Step distance between teeth	

It should be pointed out that the experimental design is not unique, and can be designed based on the number of respondents who are available (He and Chen, 2011, Hoyle et al., 2009). For example, when there are a large number of respondents, fewer choice experiments can be used for each respondent, than in experiments with fewer respondents. Pairing the usage contexts to consumers' primary usages is recommended, as it yields a better understanding of the influence of usage context attributes. If a two-stage (consumer) decision making process is considered (i.e., first the choice set is selected followed by the specific product), the survey will be designed for predicting the choice set for each consumer first.

For demonstration purpose, in this case study, each respondent is surveyed for more than one single-usage scenarios, but only single usage scenario at a time. This is based on the assumption that the respondents have some level of experience with the product under each usage scenario and are able to differentiate between the different usage scenarios described in the survey questionnaire. It should be pointed out that the DOE can be designed based on the number of respondents who are available. For example, when there are sufficient respondents, less choice experiments should be used for each respondent. Pairing the usage scenarios to consumers' primary usages is expected to yield the best understanding of the influence of usage context attributes.

If the two-stage (consumer) decision making described in Phase II of Section 4.4 is considered, the DOE should be redesigned together with the survey questionnaire. The respondent will be first presented with all products available in the market and asked to choose the ones that he/she will consider for purchase, given the primary usage scenario. A similar try-it-out survey follows to collect consumers' choice. In the choice modeling phase, a separate model will be built for predicting the choice set for each consumer.

4.5.3 Phase III: Linking Performances with Usage Context and Consumer Profile

In this study, the link between product variables **X**, performance-related usage context attributes $\mathbf{E_Y}$, performance-related consumer profile attributes $\mathbf{S_Y}$, and engineering performance **Y** (Eqn. (4.2), is established using a series of physics-based equations based on the functional principles of the jigsaw (Yannou et al., 2009, Yannou et al., 2010). Both engineering performances **Y** considered in this study, the advance speed S_a and comfort level $P_{comfort}$, belong to Y_t (performance of transformation). The advance speed S_a is calculated as follows:

$$S_a = \frac{2H_d f \cdot A}{s} \tag{4.7}$$

where H_d is the blade tooth height, f is the stroke frequency, A is the blade translation, and s is the step distance between two teeth. All variables in the equation (H_d, f, A and s) are product design variables **X**; usage context doesn't influence this particular performance. The comfort level $P_{comfort}$ is associated with the required wrist torque with respect to user' maximum wrist capability, as shown in the following equation:

$$P_{comfort} = 1 - \left| \frac{M_w}{M_{w-\max}} \right| \qquad (4.8)$$

where M_w is the wrist torque and $M_{w-\max}$ is the maximal wrist torque that can be delivered by the user. In (Yannou et al., 2009, Yannou et al., 2010), M_w is modeled as a function of both product design variables **X** and usage context attributes $\mathbf{E_Y}$ (i.e., wood type), while $M_{w-\max}$ is modeled as a function that depends on consumer profile attributes $\mathbf{S_Y}$. Details of the above physics-based equations can be found in references (Yannou et al., 2009, Yannou et al., 2010).

4.5.4 Phase IV: Choice Model Estimation

A multinomial logit model is estimated using STATA (Stata, 1996-2009). The goodness of fit, measured by the rho square is 0.82 with a log likelihood of -500.76. The coefficients the estimators, standard errors, and the significance of their *p* values are provided in Table 4.4. The price *M* is divided by the income S_1, as consumers with higher income are expected to be less sensitive to the price. The sign of the M/S_1 coefficient shows that price has a negative impact on the utility function. The coefficients of Y_{t1} and Y_{t2}, are both significant, showing that both performances are important in users' choice. Therefore, in-service and service performance results must be considered in the jigsaw design. The coefficient for S_2*Y_{t2} is significantly positive, which indicates that the female users tend to care more about the comfort than male users do. This is important to consider in the design process if the intended market for the saw has a sizable female population. Similarly, the coefficient for S_3*Y_{t1} is significantly positive, meaning that skilled users care more about the advance speed during cutting, compared with amateur users. As for the interactions between performance Y_{t1} and usage context variable E_1 (indoor / outdoor) and performance Y_{t2} and usage context variable E_2 (wood type), both coefficients are statistically significant, which indicates that both E_1 and E_2 belong to the category of preference-related usage context variable E_W. Moreover, the negative sign suggests that Y_{t2} (comfort) is less important when users are cutting outdoors (E_1=1), while the positive sign indicates that advance speed is more critical when users are cutting hard wood (E_2=3). This again provides direction in the design process: for example, if the intended market for the saw is users cutting soft to medium woods (e.g. framing materials), then advance speed is not as important in the design than if the intended market is for those cutting hard woods (e.g. hardwood flooring). The results from this case study are consistent with the general trend in consumer preferences assumed in data generation.

Table 4.4: Multinomial Logit Model Estimation Results in Jigsaw Case Study

Variables	Coef.	Std.Err.	P>|z|
Y_{t1}	5.39	1.42	0.00
Y_{t2}	27.30	1.98	0.00
M/S_1	-35.86	1.76	0.00
S_2*Y_{t2}	4.13	1.51	0.01
S_3*Y_{t1}	7.42	0.49	0.00
E_1*Y_{t2}	-4.94	1.62	0.00
E_2*Y_{t1}	4.06	0.49	0.00

With the estimated choice model, future demand of a target market (including target consumers and target usages) can be projected. Here we take the prediction of a single user' choice probability as an example to illustrate the difference between single usage and multiple usage scenarios. Considering the following two scenarios as shown in Table 4.5 for a female user with $70k annual income and skill level 3: 1) Single-usage: she uses the product solely under single usage 1, indoor cutting for soft wood; 2) Multiple-usage: she uses the product under usage 1, indoor cutting for soft wood, with 30% relative importance, and usage 4, outdoor cutting for hard wood, with 70% relative importance.

Table 4.5: Usage Importance Index F for Choice Prediction

	Usage 1	Usage 2	Usage 3	Usage 4
Single-usage	100%	0%	0%	0%
Multiple-usage	30%	0%	0%	70%

As shown in Table 4.6, in the single-usage case, the most preferred product 2 has a choice probability of 75.2%. However, in the multiple-usage case for the same user, product 3 has the highest choice probability of 55.4%, while the choice probability of product 2 (44.1%) is the second highest. It is interesting to note that the preference rank order of products may change when the usage scenario is different. Similar approach can be applied to forecast the choice probability of a group of target consumers each with different usage scenarios by aggregating individual's choice probability over a target population.

Table 4.6: Choice Share Prediction in Single-usage and Multiple-usage Scenarios

***Pr* (%)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Single-usage	0.4	75.2	24.2	0.0	0.0	0.2	0.0	0.0
Multiple-usage	0.4	44.1	55.4	0.0	0.0	0.1	0.0	0.0

4.6 HYBRID ELECTRIC VEHICLE EXAMPLE

Alternative fuel vehicles have drawn increasing attention in the past few years, because of their potential to reduce greenhouse-gas emissions and utilize renewable energy sources (Ehsani et al., 2009, Axsen et al., 2008, Shiau et al., 2009a). However, understanding consumer choices of alternative fuel vehicles is challenging because their preference construction process involves many aspects beyond traditional engineering considerations, which calls for a comprehensive modeling framework to incorporate usage context into engineering design. Taking HEV as an example, vehicle performances, such as mileage per gallon, often depend highly on their usage contexts, while consumers' attitudes towards new technology, especially "green" products, are strongly influenced by their intended usage. In this section, a HEV case study is used to illustrate the proposed usage context based choice modeling framework. Different from the jigsaw problem, the revealed preference data collected by JD Power and Associates for both HEVs and conventional vehicles (CVs) is utilized for model estimation. It should be noted that in our current study, the impact of HEV policies and other purchase incentives is not considered, because Diamond (2009) found that the impact of incentive policies on hybrid adoption is much weaker, compared to the strong relationship between gasoline prices and hybrid adoption.

4.6.1 Phase I: Usage Context Identification

Two usage context attributes are considered for HEV choice modeling: a *local/highway indicator* and *average miles driven daily*. While both attributes are treated as preference-related attributes $\mathbf{E}_W$, the local/highway indicator is also considered as a performance-related attribute $\mathbf{E}_Y$ in mileage per gallon (MPG) calculation, as detailed later. The local/highway indicator is assessed based on the combined MPG published by US Environmental Protection Agency (EPA, 2008) and the estimated MPG given by survey respondents. The indicator is a continuous parameter, ranging from 0 for pure local driving to 1 for pure highway driving. It is assumed that the local/highway indicator reflects the general driving condition a respondent faces, therefore the vehicle usage context. The local/highway driving condition not only greatly impacts vehicles' performances, e.g. MPG, but is also expected to influence consumers' choice preference for hybrid vehicles. The other usage context attribute considered is *average miles driven daily*, a commonly used descriptor of consumers' travel pattern which strongly influences the target performance of the batteries. The data is derived from the recorded miles driven in the first three months in the J.D. Power and Associates data.

4.6.2 Phase II: Data Collection

The Vehicle Quality Survey (VQS) conducted by J.D. Power and Associates belongs to the revealed preference data because the consumer satisfaction survey is strictly about the new vehicles respondents purchased instead of hypothetical design alternatives. In the 2007 VQS, vehicle purchase data from 90,000 nation-wide respondents on over 300 vehicles in the market are collected, including data for 11 HEV models. Further, respondents' demographic attributes and their usage patterns are recorded in the questionnaire. For model estimation, data collected from 8025 respondents, who reported their vehicle choice sets, are selected. The attributes and design variables included in the choice model are summarized in Table 4.7.

Table 4.7: List of Attributes and Design Variables included in HEV Case Study

***Consumer-desired product attributes* A**		
A_1	Price	Price paid, excluding tax, license, trade-in, etc.
A_2	MPG	Mileage Per Gallon under usage
A_3	Vehicle origin	Domestic / European / Japanese / Korean
A_4	Vehicle size	Compact / Midsize / Large / Premium
A_5	Vehicle type	Mini / Car / SUV / Minivan / VAN / MAV / Pickup
A_6	Hybrid electric vehicle	1 for hybrid, 0 for conventional
$A_{exterior}$	Exterior attractiveness rating	
$A_{interior}$	Interior attractiveness rating	
$A_{storage}$	Storage and space usage rating	Discrete rating on a scale from 1 to 10
A_{audio}	Audio rating	
A_{seats}	Seats rating	
A_{hvac}	HVAC rating	
$A_{dynamics}$	Driving dynamics rating	

A_{engine}	Engine and transmission rating	
A_{safety}	Visibility and safety rating	
***Usage context attributes* E**		
E_1	Local / highway indicator	0 – local, 1 – highway
E_2	Average miles driven daily	Unit: miles
***Consumer profile attributes* S**		
S_1	Gender	1 for male, 2 for female
S_2	Age	Age
S_3	Income	Household income last year
S_4	Children	Number of children under 20 living in the household
S_5	Education	Level of education completed
***Product design variables* X**		
X_1	Exterior dimensions	Vehicle length / width / height (unit: in.)
X_2	Vehicle weight	Unit: lbs
X_3	Interior dimensions	Front head / shoulder / hip / leg room (unit: in.) Rear head / shoulder / hip / leg room (unit: in.)
X_4	Storage capacity	Luggage capacity Cargo capacity
X_5	Engine specifications	Engine size Number of cylinders
X_6	Performance	Horsepower Torque
X_7	MPG targets	Target Mileage Per Gallon level under city / highway condition

There are 288 car models covered in the data set, each of them is chosen by at least one respondent. Fifteen consumer-desired product attributes **A** are selected including *price*, *vehicle origin*, *vehicle size*, *vehicle type*, *mileage per gallon* (MPG), *hybrid electric vehicle indicator*, and nine rating scores given by the respondents. The attribute "price" is the money respondents paid excluding tax, license, trade-in and etc. Since VQS only provides price for the purchased vehicles, the price data for other vehicles considered are estimated from a linear regression model based on vehicle make and model, and consumers' geographic locations. As shown in the third column in Table 4.7, vehicle origins are categorized as domestic, European, Japanese, and Korean; vehicle sizes are grouped into compact, midsize, large, and premium; vehicle type includes mini, car, sport utility vehicles (SUV), minivan, van, multi-activity vehicles (MAV), and pickup. The *hybrid electric vehicle indicator*, coded as 1 for hybrid vehicles, and 0 for conventional vehicles, reflects consumers' attitude toward new hybrid technology. Nine aspects of a vehicle, including exterior attractiveness, interior attractiveness (as stated by the average purchaser), storage and space usage, audio/entertainment/navigation system, seats, heating ventilation and air conditioning, driving dynamics, engine and transmission, and visibility and driving safety, are rated on a scale of 1 to 10, 10 being the most satisfactory. These discrete ratings are included in the choice modeling procedure, as they are considered to be a good measure of consumers' perceived vehicle performance (quality).

Meanwhile, *gender*, *age*, *household income*, *number of children under age 20 living together* and *education level*, are included as five consumer profile attributes **S**. Among the set of **S**, critical preference-related attributes $\mathbf{S_W}$ will be identified through choice modeling in Phase IV. All five **S** attributes are considered in the ordered logit regression for predicting the performance rating scores, as will be shown in Phase III.

4.6.3 Phase III: Linking Performances with Usage Contexts and Consumer Profile

Different from the jigsaw example in which physics-based modeling can be used to establish the relationship between performance and usage context attributes, in the HEV example, respondent survey data is used to create the relationship as shown in Eqn. (4.2) by using the ordered logit modeling method (Hoyle et al., 2011) for nine consumer desired product attributes (**A**) in the form of ratings. Here the ratings are used to represent product performances **Y**. Seven high level engineering design variables **X** are used in this case study, including *exterior dimension*, *interior dimension*, *performance*, *MPG targets*, etc. The obtained ordered logit models are also used to predict the ratings of other vehicle designs in the choice set as consumers only rate the vehicles they purchase. This limitation of the rating data in VQS may cause ownership bias in model estimation and potentially lead to inaccurate estimates of some coefficients due to the missing heterogeneity in owners' ratings. Further details for implementing the ordered logit model based on the VQS data by JD Power and Associates can be found in (He et al., 2011a). In addition to the design variables **X**, consumer profile $\mathbf{S_Y}$ such as *gender*, *age*, etc., are included to capture consumers' heterogeneity in rating. The coefficients estimators are later used for what-if-scenario analysis to forecast potential market share for targeting consumer and usage attributes.

Furthermore, the impact of usage context (local/highway indicator E_1) on the vehicle performance (A_2, *mileage per gallon*) is represented in the following equation:

$$A_2 = \frac{1}{\dfrac{1-E_1}{MPG_{city}} + \dfrac{E_1}{MPG_{highway}}}. \tag{4.9}$$

where MPG_{city} and $MPG_{highway}$ belong to the product design variables **X** listed in Table 4.7.

4.6.4 Phase IV: Choice Model Estimation

In Phase IV, as a result of choice modeling, interactions between consumer-desired product attributes **A**, usage context attributes **E**, and consumer demographics **S** are explicitly modeled in the utility function. The coefficients for all attributes and their interactions based on a multinomial logit model estimation (MNL with **E**) are listed in Table 4.8, together with the estimation results from a multinomial logit model without usage context attributes (MNL without **E**) as a comparison.

Table 4.8: Coefficients of MNL with E and MNL without E for HEV

	MNL without E		MNL with E	
Attributes	**Coeff.**	**Sig.**	**Coeff.**	**Sig.**
A_1/S_3	-0.0003	***	-0.0004	***
A_2	/		-3.1080	***
A_2 *city*	0.0456	***	/	
A_2 *highway*	-0.0791	***	/	
E_1*A_2	/		5.9454	***
E_2*A_2	/		0.0002	
A_3 *European*	1.9353	***	2.1886	***
A_3 *Japanese*	0.2314	***	0.5161	***
A_3 *Korean*	1.2617	***	1.3313	***
A_4 *Large*	-0.5636	***	-0.9111	***
A_4 *Medium*	0.0907	**	-0.1376	*
A_4 *Premium*	-0.2496	***	-0.3136	***
A_5 *MAV*	-0.9746	***	-1.3558	***
A_5 *Mini*	0.6717	***	1.3347	***
A_5 *Minivan*	-0.6733	***	-0.9826	***
A_5 *Pickup*	1.6354	***	1.4446	***
A_5 *SUV*	-0.0632		-0.2733	*
A_5 *Van*	1.2458		0.8806	
A_6 *hybrid*	2.8933	*	57.0667	***
E_1*A_6 *hybrid*	/		-105.84	***
S_5*A_6 *hybrid*	0.2875	***	0.1686	*
$A_{exterior}$	0.0593	***	0.0407	
$A_{interior}$	0.4835	***	0.4585	***
$A_{storage}$	0.5804	***	0.6253	***
A_{audio}	0.1741	***	0.1421	***
A_{seats}	0.1636	***	0.1046	*
A_{HVAC}	0.1242	***	0.1285	***
$A_{dynamics}$	0.2362	***	0.1640	***
A_{engine}	0.2565	***	0.3061	***
A_{safety}	0.0999	**	0.0455	

Significant with p value: * <=0.001; ** <=0.01; * <=0.05.**

From the results of the MNL including **E** attributes in modeling, we note that the coefficient for price/income is negative as expected. Only two consumer profile attributes, *household income* and *education level*, are statistically significant as preference-related attributes $\mathbf{S_W}$. A positive estimator for E_1*A_2 indicates that the usage context attribute E_1 (local/highway indicator) has a positive impact on consumers' preference on MPG measure. In other words, people primarily driving on highways tend to care more about the MPG value, in addition to the utility increase experienced by the change in MPG. Moreover, the attitude toward HEV itself has a fairly large coefficient estimator of 57.0667, which shows that people driving locally tend to favor HEV. Similarly as we expected, highway drivers do not prefer HEVs, as shown in the negative coefficient estimator of the E_1 and HEV indicator interaction (E_1*A_6). This finding presents an opportunity to design a HEV which performs well in highway driving to help overcome this issue. On the other hand, most coefficients from the MNL without modeling **E** have the same sign as the ones in MNL with **E**, but they are very different in magnitude, as the usage heterogeneity is not explicitly modeled. Inclusion of usage context will help designers more specifically target the vehicle design to the usage contexts of the intended market for the vehicle. It should be noted that the results shown in Table 4.8 are attained through sequential estimation of a hierarchical model and do not account for error propagation. Moreover, the negative coefficient of $A_2\_highway$ is due to high correlation between A_2_city and $A_2\_highway$. Multicolinearity between explanatory variables should be cautioned in model estimation.

4.6.5 Goodness-of-fit Measures

Goodness-of-fit measures based upon the log-likelihood of the converged model, such as the likelihood ratio index ρ^2 (also known as pseudo R-square), reflect how well the estimated model predicts actual individual choices in the data set. Higher values of ρ^2 indicate better predictions of the choices. As shown in Table 4.9, a significantly higher log-likelihood of -4825.26 and subsequently ρ^2 value of 0.5663 are achieved using the MNL model with usage context attributes **E** versus the MNL model without **E**. This implies that introducing the usage context attributes in choice modeling has captured the systematic taste heterogeneity of consumers under different usage contexts. This is important for designers so that they may best understand the preferences and usage contexts of the intended users.

Table 4.9: Model Statistics of MNL without E and with E

Multinomial Logit Model	without E	with E
Log likelihood at zero	-11125.01	-11125.01
Log likelihood at convergence	-6178.62	-4825.26
ρ^2	0.4203	0.5663

4.6.6 Cross-validation

For cross-validation of a choice model, the original data are divided into 5 subsets of samples. For each of the five cross-validation tests, a choice model is trained on 4 subset samples and later validated using the remaining hold-out sample. The likelihood ratio index ρ^2 values and hit rates (percentage of correctly predicted choices) are calculated and averaged out, as listed in Table 4.10.

Table 4.10: Cross-Validation Results for UCBCM of HEVs

Test	MNL without E		MNL with E	
Measure	ρ^2	hit rate	ρ^2	hit rate
1	58.72%	67.41%	68.10%	75.76%
2	57.10%	66.60%	66.87%	76.20%
3	56.10%	66.42%	65.61%	74.02%
4	56.35%	65.67%	65.84%	74.14%
5	55.77%	66.79%	65.98%	75.20%
Average	56.81%	66.58%	66.48%	75.07%

On average, the likelihood ratio index ρ^2 shows an over 17% improvement from 56.81% in the MNL without **E** to 66.48% in the MNL with **E**. The hit rate, though not theoretically consistent with random utility theory, is another commonly used measure of the prediction accuracy of an estimated model at the individual level. It is calculated by dividing the number of correctly predicted choices by the total number of respondents. Similar to ρ^2, the hit rate increases from 66.85% in the MNL without **E** to 75.07%, which shows that usage context greatly influences consumers' choice and should be modeled explicitly. As the choice model estimation and cross-validation are performed with choice set information from the VQS, future work is needed to address the challenges in predicting choice set construction for new market.

4.6.7 Market Segment Prediction Tests

The two models, MNL with E and MNL without E, are compared based upon the error in choice share prediction for conventional vehicles (CVs) and hybrid electric vehicles (HEVs). The market segment prediction test is conducted for three segments of Driving Conditions (local, combined, highway) and three segments of Age (low, medium, high). The results of Driving Condition market segment test and Age market segment test are shown in Table 4.11. 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):

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n_s}}, \qquad (4.\ 10)$$

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.

It is observed that the choice shares of HEVs vary more substantially by Driving Condition market segment than by Age segment. All of the choice share predictions are within the 95% confidence intervals. In the Driving Condition study, predictions by MNL with E are more accurate than those from MNL without E for local and highway driving condition segments, while both market segments have significantly different choice shares for HEVs, compared to the average in total. For two of the three Age market segments, the MNL model with E outperformed the MNL model without E.

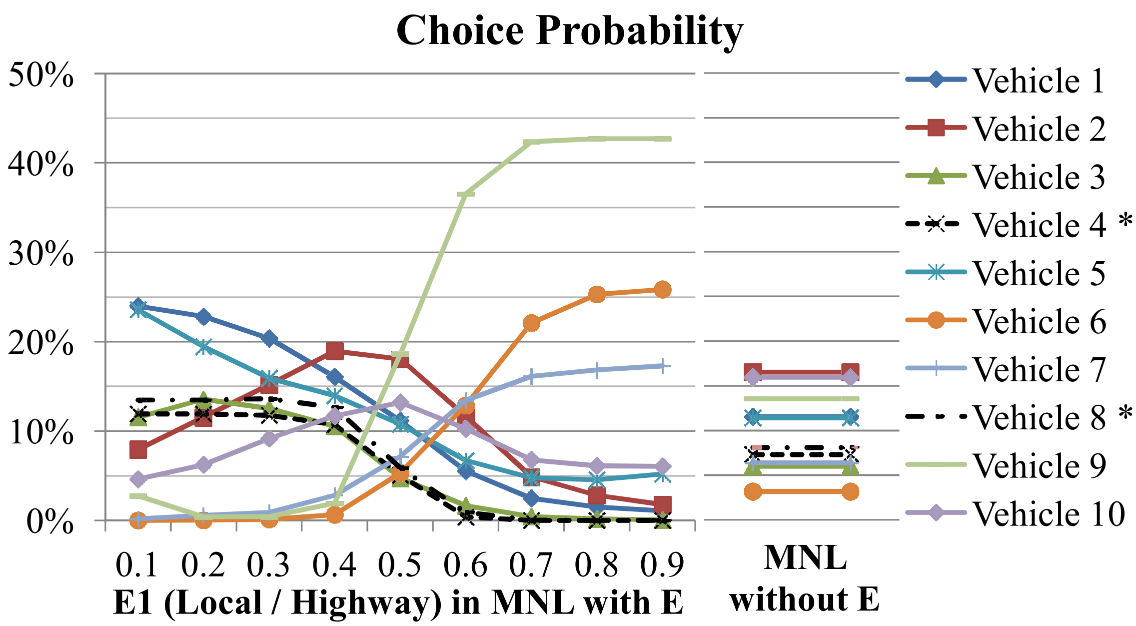
Table 4.11: Comparison of Choice Share Predictions for CVs and HEVs

		MNL without E				MNL with E			
Local Driving Condition Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	0.313	**0.315**	0.297	0.332	0.51%	**0.312**	0.295	0.330	0.31%
HEVs	0.021	**0.020**	0.015	0.025	7.47%	**0.022**	0.017	0.028	4.50%
Combined Driving Condition Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	0.332	**0.332**	0.314	0.350	0.06%	**0.333**	0.315	0.351	0.38%
HEVs	0.011	**0.011**	0.007	0.015	1.90%	**0.010**	0.006	0.014	11.23%
Highway Driving Condition Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	0.320	**0.318**	0.300	0.336	0.57%	**0.320**	0.302	0.338	0.09%
HEVs	0.002	**0.004**	0.002	0.007	80.90%	**0.003**	0.001	0.004	13.17%

		MNL without E				MNL with E			
Low Age Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	**0.317**	**0.316**	0.298	0.334	0.24%	**0.316**	0.298	0.334	0.18%
HEVs	**0.007**	**0.008**	0.004	0.011	11.01%	**0.008**	0.004	0.011	8.24%
Medium Age Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	**0.291**	**0.292**	0.274	0.310	0.18%	**0.292**	0.273	0.310	0.10%
HEVs	**0.013**	**0.013**	0.008	0.017	3.88%	**0.013**	0.008	0.017	2.22%
High Age Segment									
	Observed	**Predicted**	**95% C.I.**		**% Error**	**Predicted**	**95% C.I.**		**% Error**
CVs	**0.304**	**0.304**	0.286	0.321	0.01%	**0.304**	0.286	0.322	0.04%
HEVs	**0.012**	**0.012**	0.008	0.017	0.18%	**0.012**	0.008	0.016	1.07%

4.6.8 What-if-scenario Analysis

With the formulations described above and choice model results from MNL with **E**, a prediction model can be built to forecast the consumers' choice. For example, a target population of 260 consumers is simulated with consumer profile distribution drawn from the hybrid owner pool in VQS 2007 data set. Assuming that they are choosing a new vehicle to purchase from a choice set of 4 car models selected from 10 car models available in the market. The ten car models, among which two (vehicle 4 and vehicle 8) are HEVs, are selected based on their popularity in the choice set of consumers who considered at least one HEV. The choice set of each consumer can also be predicted using statistical learning or data mining methods with existing market data. Since not all consumers would consider a HEV when they shop for a new car, we assume that 40% of consumer would consider HEVs, while the rest of them don't. Additionally, we consider a series of nine different usage contexts: a uniformly distributed *local/highway indicator* with 0.2 range and mean value from 0.1 to 0.9 (with 0.1 interval), while average *miles driven daily* matches with the original dataset. Aggregated choice probability in target population calculated using our proposed framework is summarized in Figure 4.5.



*** Hybrid electric vehicle**

Figure 4.5: Choice Probability of Target Population under Different Usage Contexts

In Figure 4.5, the solid color lines (conventional vehicles) and dashed black lines (HEVs) on the left hand side show the predicted choice probability by MNL with **E**, while the lines on the right hand side represent the constant choice probability predicted by MNL without **E**. For instance, when the target population, on average, drives 40% under local conditions, the hybrid electric vehicle 4 and vehicle 8 have the predicted choice probabilities of 10.72% and 12.64%, respectively in MNL with **E**, as opposed to the constant 7.32% and 8.11% in MNL without **E**. According to the prediction from the MNL with **E**, their predicted market shares gradually decrease, as E_1 increases. When E_1 is less than or equal to 0.3, conventional vehicle 1 have the largest market share, closely followed by conventional vehicle 5. When E_1 increases to 0.5, each car model has its niche in the market. When E_1 is larger than or equal to 0.6, conventional vehicle 9 becomes the dominant car model, as it has the highest choice probability. In comparison, the predicted dominant vehicle choice by the MNL without **E** turns out to be conventional vehicle 2 with a choice probability of 16.51%, which is significantly different from the one predicted by the MNL with **E**. Since the missing usage information plays a key role in consumer choice and it is natural to expect that consumers make distinctive decisions when usage context changes, the MNL model with **E** is able to reveal relationships between usage context and consumer preference for product attributes that would not be revealed without E. Accurately predicting the choice probabilities (i.e. market share) for a given vehicle design, consumer population and set of usages are important to be considered by vehicle designers to tailor the vehicle design to the target market as closely as possible. Further research is needed to assess implications of model specification assumptions and to test external validity on predictions of choice shares for different usage contexts.

4.7 Conclusions and Discussions

In this chapter, a choice modeling framework for UCBD is proposed to quantify the impact of usage context on consumer choices. Previous works have illustrated the importance of considering usage context in design, but did not present a systematic and quantitative approach to choice modeling. The primary focus of this work is the development of a systematic UCBD taxonomy and a step-by-step procedure to quantitatively assess the impact of usage context on product performance and consumer preferences.

A taxonomy for UCBD is first defined by following the established classification in the market research domain and the needs associated with choice modeling. The step-by-step procedure for creating choice models in UCBD is then presented. To facilitate the identification of usage contexts in Phase I, it is recommended to elicit the usage context attributes from five categories of product usages including *physical surroundings*, *social surroundings*, *temporal perspective*, *task definition*, and *antecedent states*. In Phase II data collection, both the methods of Stated Preference and Revealed Preference surveys are presented to account for the choices respondents make conditional on the given usage context, which allows us to examine simultaneously the influence of product design, consumer profile, usage context, and their interactions, on consumer choices. Furthermore, Phase III is a unique step in a quantitative UCBD process in which the influence of usage context and consumer profile on product performance is analytically modeled. Additionally, in Phase IV, usage context enters into an individual's choice utility function directly to capture its influence on product preferences. In Phases III and IV of modeling, both consumer profile attributes $\mathbf{S}$ and usage context attributes $\mathbf{E}$ are further classified into performance-related $\mathbf{S_Y}$, $\mathbf{E_Y}$ and preference-related $\mathbf{S_W}$ and $\mathbf{E_Y}$ to differentiate their impact on product performance and consumer preferences, respectively. The usage context choice modeling approach in this work represents a significant expansion of traditional choice modeling approaches in the design literature.

Two case studies, a jigsaw design example with synthetic stated preference data and a HEV example with real revealed preference data, illustrate the proposed modeling framework. Both case studies follow the four-phase modeling procedure. The jigsaw case study emphasizes usage context identification, data collection with stated preference surveys, and the use of physics-based modeling to capture the impact of usage context on performance. On the other hand, more details of the modeling steps (Phases III and IV) are reported in the HEV case study based on its revealed preference survey data that reflects consumers' real choices and the use of ordered-logit modeling for predicting consumer ratings for system attributes. Results from both examples demonstrate the impact of usage context upon consumer preference as well as product performance. A set of validation tests are included for the HEV case study which demonstrate the necessity of expanding a traditional choice modeling framework to include usage context for improved model predictive capability. What-if-scenario analysis in the HEV example showed that predicted choice share in the target market changes in response to the change of performance ratings in distinctive usage contexts for given vehicle designs, which illustrates the potential of the proposed choice modeling framework in supporting engineering product design. An optimization problem can be formulated using the proposed framework to determine the optimal performance targets for engineering design. For example, in the case of HEV battery design, performance targets include both city and highway MPG as well as vehicle horsepower and torque.

Chapter 5

STATISTICAL ANALYSIS OF HYBRID ELECTRIC VEHICLE ADOPTION

5.1 INTRODUCTION

In this chapter, we analyze the vehicle usage and consumer profile attributes extracted from both National Household Travel Survey and Vehicle Quality Survey data to understand the impact of vehicle usage upon consumers' choices of hybrid electric vehicles. In addition, the key characteristics of hybrid vehicle drivers are identified to determine the market segmentations of hybrid electric vehicles and the critical attributes to include in the choice model. After a compatibility test of two datasets, a pooled choice model combining both data sources illustrates the significant influences of vehicle usage upon consumers' choices of hybrid electric vehicles. Even though the data-bases have in the past been used independently to study travel behavior and vehicle quality ratings, here we use them together. Based on the findings from data analysis and choice modeling, a study of vehicle design selection is carried out to evaluate target consumers' choices among conventional vehicle, hybrid electric vehicles, and plug-in hybrid electric vehicles.

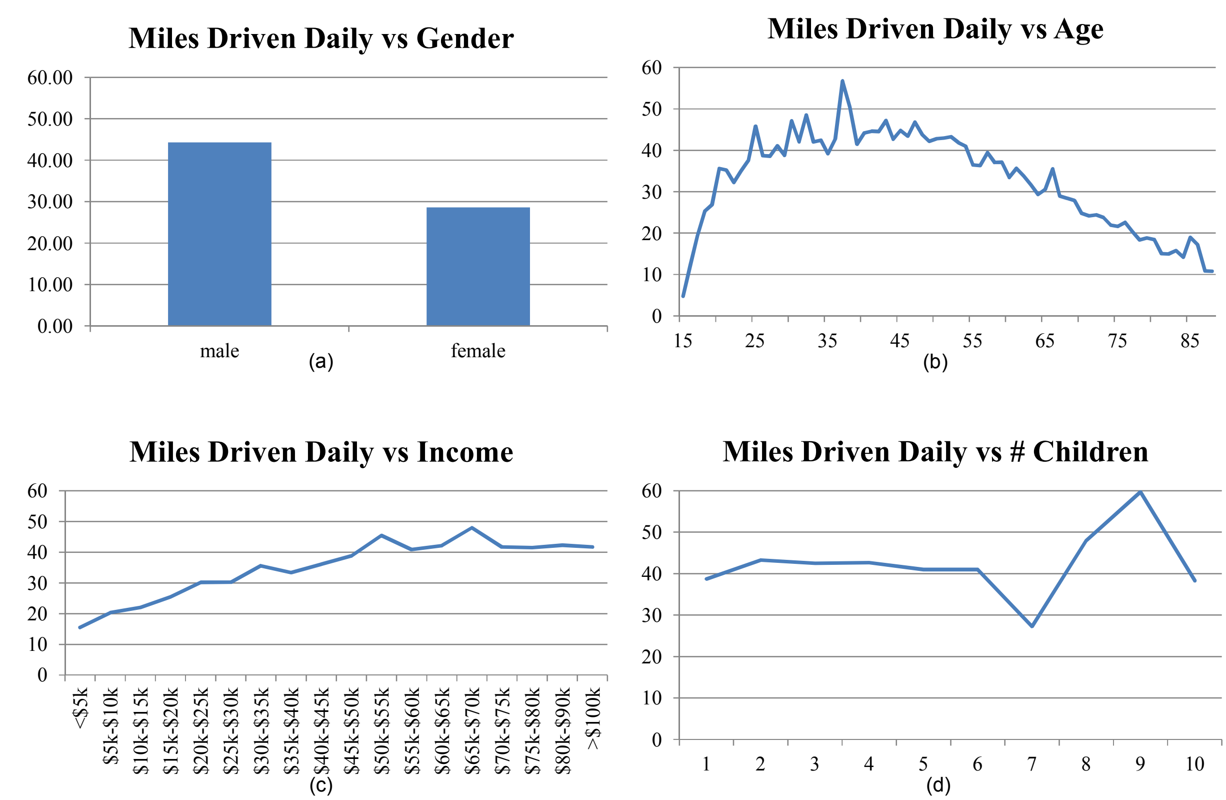
This chapter is organized as follows: In Section 5.2, we start with data analysis results for NHTS. A market segmentation study of HEV drivers is presented in Section 5.3, followed by a compatibility study of NHTS and VQS datasets in Section 5.4. A choice model is created in Section 5.5 to demonstrate the benefits of including usage context attributes and using combined NHTS and VQS data for studying consumer preferences of hybrid technology. The vehicle design selection study is presented in Section 5.6, followed by summary and conclusion in Section 5.7. It should be noted that the impact of HEV policies and other purchase incentives is not modeled in this work, as explained in Chapter 4.

5.2 Analysis Of Vehicle Usage Context And Consumer Profile

As discussed in previous chapter, usage context plays a critical role in consumers' choice because product performance and consumer preference change under different product usages. Nonetheless, the question about the relationship between usage context and consumer profile attributes remains: are they correlated with each other? If so, can we predict usage context based on consumer profile? Is it necessary to include both usage context and consumer profile attributes in choice modeling? To address these questions, we devote this section to investigate the relationship between usage context attributes and consumer profile attributes. The National Household Travel Survey (NHTS) data is used here to facilitate the data analysis and provide insights into vehicle usage context across the nation.

The NHTS is the nation's inventory of daily travel (FHWA, 2009). The survey includes demographic characteristics of households, people, vehicles, and detailed information on daily travel for all purposes by all modes (FHWA, 2009). NHTS data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and household attributes. The 2009 NHTS data set includes information on 150,147 households, 308,901 people, 309,163 vehicles, and 1,167,321 trips.

The usage context attribute *miles driven daily*, included in NHTS 2009 data, is of great interest in vehicle design, because it has a significant impact on choices of hybrid electric vehicles and plug-in hybrid electric vehicles (Shiau et al., 2009a, Shiau et al., 2010). Meanwhile, *gender*, *age*, household *income*, *number of children* living together, and *education level* are among the most commonly used consumer profile attributes (He et al., 2010). In Figure 5.1, the mean of usage context attribute *miles driven daily* is plotted against the five consumer profile attributes, respectively, to illustrate their relationships. It should be noted that the household weights and personal weights from NHTS data, calculated based on American Community Survey, are applied in data analyses throughout the dissertation to adjust the sampling and non-response bias.



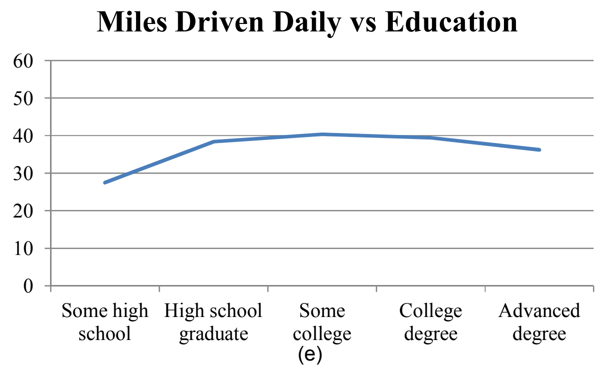


Figure 5.1: Miles Driven Daily versus

(a) Gender, (b) Age, (c) Household Income, (d) # Children, (e) Education Level

As shown in Figure 5.1(a), male respondents drive 44 miles on average, 54% more than their female counterparts. From (b), we see a bell curve in miles driven daily versus age. This indicates that the miles driven daily increases from 5 miles at age 16, peaks around 45 miles at age 35-45, and decreases slowly afterwards. Figure 5.1 (c) shows an increasing trend in miles driven daily with the increase of household income, while little conclusion can be drawn for relationship between miles driven daily and number of children, as seen in Figure 5.1 (d). Lastly, in Figure 5.1 (e), miles driven daily slowly increases with the increase of education level but drops at the end for advanced degree.

These observations of one-to-one relationship between usage context and consumer profile attributes can provide guidance for multivariate analysis such as Analysis of Variance (ANOVA) (Tamhane and Dunlop, 2000). In ANOVA, the usage context attribute *miles driven daily* is the dependent attribute, while the five consumer profile attributes discussed above are independent attributes. The ANOVA results are listed in Table 5.1. All five consumer profile attributes are statistically significant. However, they only contribute to about 10% of the total sum of squares, which suggests that the majority of the variance in usage context attribute *miles driven daily* cannot be explained by these five consumer profile attributes.

Table 5.1: Top Consumer Profile Influencing Usage Context

	Partial SS	F	
Gender	1.68E+12	8115.2	*
Age	1.35E+12	91.7	*
Income	9.54E+11	271.3	*
Children	3.07E+10	14.9	*
Education	2.62E+11	316.4	*
R^2	0.1009		

*** Significant with p value <=0.05.**

The ANOVA analysis indicates that the usage context attribute *miles driven daily* cannot be accurately predicted by a function of consumer profile attributes, as they do not fully explain the variances of the usage context attribute. Instead, it should be treated as an additional dimension of consumer classification measure, or consumer attributes. Similar to the consumer profile attributes, usage context attributes are traits of the consumers which can be used to categorize consumers into groups. For example, *miles driven daily* can be used to classify consumers into short distance, medium distance, and long distance drivers, just as consumers are usually divided into low, medium, high income segments based on their household *income*. Hence, both usage context and consumer profile attributes should be considered in market segmentation analysis, as will be illustrated in Section 5.3.

5.3 Identifying Characteristics Of Hybrid Electric Vehicle Drivers

As demonstrated in Section 5.2, in addition to consumer profile attributes, usage context should also be included in consumer preference studies, as it describes another dimension of market segmentation that cannot be explained by consumer profiles alone. But including all consumer profile and usage context attributes in the choice modeling process may cause problems in model estimation and lead to difficulties in model interpretation due to their correlations. In many cases, significant attributes identified directly by choice model estimation are unstable and sensitive to the list of attributes included in the estimation. To build a reliable choice model, a reduced set of key attributes needs to be identified to enter a choice modeling procedure. Moreover, identification of key characteristics of target consumers including both consumer profile and usage context attributes provides the foundation for a successful market segmentation analysis. Heterogeneous consumer preferences across market segments can be captured by creating separate choice models for each of the segments. Meanwhile, product design can be enhanced by adding features that meet the needs of a specific segmentation.

In our case study of HEVs, to identify characteristics of the hybrid electric vehicle drivers, we start with a complete list of consumer profile and usage context attributes extracted from NHTS. These attributes are at household and individual levels.

Individual-level attributes include, but are not limited to, demographic and socio-economic attributes of the survey respondents, such as *gender*, *age*, *race*, *education level*, *occupation category*, *working status*, *marital status*, *age of youngest child*, and *miles driven in the past 12 months*. *Miles driven in the past 12 months* belongs to usage context, while all others are consumer profile attributes. Dummy variables were generated for categorical attributes in the Person file of NHTS2009. These include *gender*, *age group* (0-15, 16-24, 25-34, 35-44, 45-54, 55-64, 65+ years), *race* (White, Black, Asian, Hispanic), *education level* (less than high school graduate, high school graduate, some college, college degree, graduate or professional degree), *occupation category* (service [sales or service]; clerical support [clerical or administrative support, blue collar [manufacturing, construction, maintenance, or farming]; professional [professional, managerial, or technical]). Three attributes, *working status* (working or retired), *marital status* (single or married), *age group of youngest child* (no child, age 0-5, age 6-15, age 16-21) are derived from *life cycle* classification in the Household file of NHTS2009.

Household-level attributes include basic information about the households, such as household income, household size, number of adults, number of workers, number of drivers, number of vehicles, home ownership, home type, size of metropolitan area, urban/rural status of the home address, and life cycle classification. All those attributes are extracted from the Household file of the NHTS2009, and are part of the consumer profile.

5.3.1 Market Segmentation Through Principal Component Analysis

Including all individual and household demographic and socio-economic related attributes discussed earlier in the choice model will be difficult since many of these attributes are highly correlated. For example, the number of vehicles and number of workers are highly correlated with the size of metropolitan area and the urban/rural status of a household. This high multi-collinearity between consumer attributes might create problems of identification of hybrid owners' characteristics and understanding the influence of vehicle usage on consumers' choice. Factor analysis, specifically, the principal component analysis provides a means to resolve this issue. In broad terms, factor analysis is a method for reformulating a set of natural or observed independent attributes into a new set (usually fewer in number) of independent attributes, such that the latter set has certain desired properties specified by the analyst. Principal component analysis searches through data to find the factors or components that may reduce the dimensions of variations and may be given a possible meaning (Stopher and Meyburg, 1979).

Forty explanatory attributes of consumer profile and vehicle usage context attributes extracted from NHTS data were chosen for principal component analysis. Multiple tests were done to extract the most reasonable factors from the data set. Models with eight to twenty factors were tested with different combinations of explanatory attributes. At the end, eighteen factors were extracted that explain 76.4% of variance in the data set. Each of the eigenvalues for the eighteen factors is greater than 1. The complete results of the principal component analysis using the Varimax rotation method are presented in Appendix B: Table B.1. The factor loadings of each of the explanatory attributes onto each of the factors provide a preliminary understanding of the interdependencies between each of the attributes.

Table 5.2 shows the definition of the identified factors (Column 3) and their correlations to attributes (Column 2). To better understand the factor analysis results, each factor is given a name based on the interpretation of its correlated attributes. Each factor is a linear combination of the attributes listed, and the sign of the coefficient for each attribute is shown in Table 5.2 just before each attribute. For instance, five dominant attributes forming factor 1 are household size, number of adults, drivers, vehicles, and non-single marital status. This factor is named as ***Drivers & Vehicles***, as it represents the driving dimension of the households. Dominant attributes in factor 2 include individuals between age 35 and 44, youngest children under 15, and household size. Thus, this factor indicates young to middle-aged households with children and is named as ***Mainstream Family***. Similarly, factor 4, named as ***High Income and Education***, is heavily influenced by professionals with advanced degrees and high income. Factor 13 represents ***Long Distance Worker*** who (mostly male) drive long distances on a daily basis. High correlation of 0.60 is seen between this factor and *blue collar occupation* (see Appendix B: Table B.1), which suggests that blue collar workers tend to travel more than their peers.

Table 5.2: Factors from Principal Component Analysis & T-test for HDs and CDs

Factor	Attributes	Definition	Mean Score

		#Obs	HDs 4912	CDs 89127	
1	+ household size, + # adults, + # drivers, + # vehicles, - single;	*Drivers & Vehicles*	-0.05	-0.02	*
2	+ age 35-44, - no child, + youngest children 6-15, + household size;	*Mainstream Family*	0.02	0.04	
3	+ # workers, + age 55 and 64, - age 65+, - retire;	*Working Class*	0.10	0.07	
4	+ household income, + graduate degree, + professional occupation;	*High Income & Education*	0.23	0.10	*
5	+ age 25-34, + youngest child under 5;	*Young with Child*	0.01	0.04	*
6	+ home ownership, + home type single family house, - home type townhouse;	*Home Owner*	0.10	0.06	*
7	+ home type duplex, + home type apartment/condo;	*Apartment Dweller*	-0.02	-0.01	
8	- white, + black;	*African American*	0.00	-0.05	*
9	+ Asian;	*Asian*	0.03	0.00	*
10	+ Hispanic;	*Hispanic*	0.00	-0.01	
11	+ size of metropolitan area, + rural area;	*Rural*	-0.05	-0.06	
12	+ age 45-54, - retire;	*Middle-Aged*	0.04	0.03	
13	+ miles past 12 months, - clerical occupation + blue collar occupation;	*Long Distance Worker*	-0.02	-0.12	*
14	+ age 16-21;	*Student*	-0.09	-0.04	*
15	+ college degree;	*College Graduate*	0.09	0.08	
16	+ some college experience;	*Some College*	0.01	0.04	*
17	- high school degree;	*Educated*	0.11	0.12	
18	+ sales & service occupation.	*Sales*	0.00	0.01	

5.3.2 **Characteristics of Hybrid Electric Vehicle Drivers**

Based on the principal component analysis results shown in Section 5.3.1, factor scores are calculated and used to identify the key characteristics of HEV drivers. The *hybrid* attribute (1 for HEVs, 0 for CVs) is used in categorizing and identifying potential HEV shoppers. Individual consumers and households are labeled based on the *hybrid* attribute of the vehicles they drive: if a consumer drives a HEV, he/she is labeled as *hybrid driver* (HD); if not, he/she is labeled as *conventional driver* (CD). Through a comparison of HDs and CDs, hybrid owners' characteristics can be identified. A series of t-tests is used to investigate if statistically a significant difference exists between HDs and CDs in the eighteen factors introduced above. The results are shown in Table 5.2. The asterisk (*) in the last column means that the null-hypothesis testing is significant with p value less than or equal to 0.10. In other words, a factor with asterisk is considered as influential on hybrid choice.

From Table 5.2, we can see that the factor *High Income & Education* (Factor 4) has a highly significant positive impact on hybrid drivers because their mean scores are higher for HDs than for CDs. This suggests that *household income* and *education level* may contribute to the choice behavior. The factor *Long Distance Worker* (Factor 13) also has a higher mean factor score for the hybrid drivers, which may result from its positive correlation with *gender* and *miles driven in the past 12 months*. This indicates that people's attitude toward hybrid electric vehicles may relate to their gender and their vehicle usage context. Moreover, minorities such as *African American* (Factor 8) and *Asian* (Factor 9) tend to have more hybrid drivers, which suggests that consumer profile attributes *race* could be a critical attribute in modeling consumer choice. On the other hand, the factor *Young with Child* (Factor 5) seems to have a lower mean factor score for hybrid drivers than for conventional drivers, which suggests that consumer profiles, such as *age*, *marital status*, and *number of children* may play a role in consumers' choice.

To sum up, based on the key characteristics of hybrid drivers, such as *High Income & Education*, identified through principal component analysis and t-test, a list of consumer profile and usage context attributes including *gender, age, household income, number of children, education level, marital status, race,* and *miles driven daily* are selected for further investigation in studying the compatibility of the two data sets and choice modeling, as detailed in the following section.

5.4 Compatibility Of NHTS With Vehicle Quality Survey Data

While NHTS data provides rich and detailed information about households, individuals, vehicles, and daily trips, information about the choice set each consumer considered during the vehicle purchasing process is not available. VQS data collected by JD Power and Associates, on the other hand, contains the details of the vehicles considered by each respondent, and is well-suited for building the usage context-based choice model. As the size of VQS data is much smaller than NHTS, a combination of both datasets would improve the predictive capabilities of choice modeling. To ensure that the choice model built upon NHTS and VQS data reflects and can predict the vehicle choice behavior of consumers throughout the nation, the compatibility of NHTS (2009) and VQS (2007) data sets is tested here.

The commonality and difference of the consumer attributes considered in the two data sets are first examined. As shown in the Venn diagram (Figure 5.2), NHTS and VQS share common usage context attribute E^{com} *miles driven daily*, while each of them has their own usage context attributes: *trip purpose* in NHTS and *local/highway indicator* in VQS. The common consumer profile attributes S^{com} shared by the two datasets include: *gender*, *age*, household *income*, *number of children*, *education* level, *race,* and *marital* status. Individually, NHTS provides information about *home ownership*, *home style*, *urban/rural* location, etc., while VQS collects consumers' *height* and *weight*.

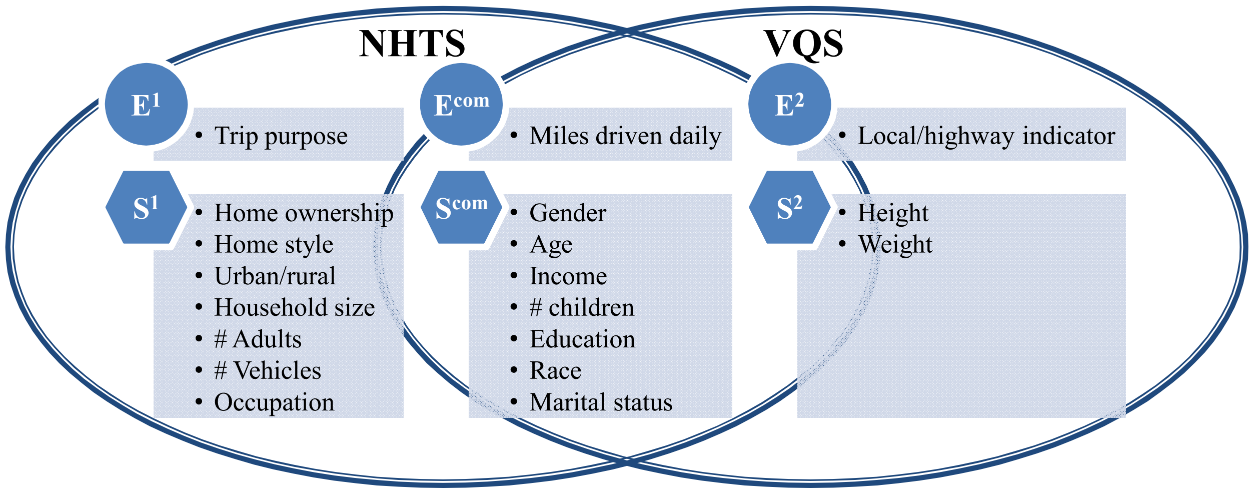


Figure 5.2: Vehicle Usage and Consumer Profile Attributes in NHTS and VQS

As determined by the factor analysis in Section 5.3, the key consumer profile attributes of interest to us include: *gender*, *age*, *household income*, *number of children*, *education level*, *marital status*, and *race*. The usage context attribute of interest is *miles driven daily* (*miles driven in the past 12 months* in NHTS). All attributes of interest are discrete data. The consumer profile attributes *household income* and *education level* are ordinal, while *gender* and *race* are categorical. The coding of ordinal and categorical data, for example *household income* brackets, in both datasets are adjusted to be consistent with each other. Two types of data analysis methods are used for numerical and categorical data, respectively, as detailed in the following discussion.

The numerical attributes are compared using Student t-test, as shown in Table 5.1. For each of the attributes, the first row in the table shows the mean attribute values for hybrid drivers and conventional drivers, respectively. The standard deviations are shown in the brackets of the second row. The asterisk mark indicates that the difference between HDs and CDs is significant with p value less than 0.05. Four attributes, *age*, *household income*, *education level*, *miles driven daily* (*miles driven in past 12 months* in NHTS divided by 365), are shown to have a positive impact on differentiating hybrid drivers from conventional drivers. The difference between HDs and CDs in *number of children* is not significant in the NHTS data. Comparing the mean values of each attributes, we see VQS data has much higher means for household income and education. This may result from the fact that the VQS samples mainly cover new vehicle owners in a specific calendar year, while the NHTS samples cover a broader range of vehicle owners in U.S.

Table 5.3: Comparison of Attributes in NHTS and VQS

	Attributes	**NHTS**			**VQS**		
	Mean	**HDs**	**CDs**		**HDs**	**CDs**	
S	*Age*	55.39 (14.3)	54.42 (15.4)	*	52.87 (13.2)	52.40 (15.1)	*
	Household income	5.94 (3.0)	5.42 (3.0)	*	7.54 (3.7)	7.21 (4.0)	*
	Number of children	0.52 (0.9)	0.50 (0.9)		0.62 (1.0)	0.58 (1.1)	*
	Education level	3.42 (1.2)	3.32 (1.1)	*	4.11 (1.7)	3.76 (1.7)	*
E	*Miles driven daily*	42.43 (38.1)	38.50 (33.8)	*	38.39 (23.6)	34.27 (23.0)	*

*** Significant with p value <=0.05.**

The 100% stacked columns in Figure 5.3 illustrate the composition of NHTS and VQS sample population. The NHTS sample is presented in darker shades, while the VQS sample is shown in lighter shades. Meanwhile, hybrid drivers are shown in red columns, and conventional drivers are shown in green columns. Along the horizontal axis, sample population is divided into groups based on their *gender*, *race* and *marital status*. The vertical axis shows the percentage of hybrid drivers (HDs) and conventional drivers (CDs) within the subgroup. For instance in Figure

5.3(a), gender distribution, about 94.5% of the male respondent in the NHTS are CDs, while the percentage of HDs in the VQS sample population is much smaller. This observation is confirmed by the other two plots in Figure 5.3. Further, as shown in Figure 5.3(a), no significant difference in percentage of hybrid drivers is noted between male and female population. Race composition is shown in Figure 5.3(b). In the NHTS sample, African American (Black) exhibits a higher percentage of hybrid drivers, compared with the other three race groups. However, the VQS sample disagrees: the black community has a much lower percentage of hybrid drivers in VQS. In Figure 5.3(c), sample populations are grouped into married or single individuals. No clear difference is seen between the two groups.

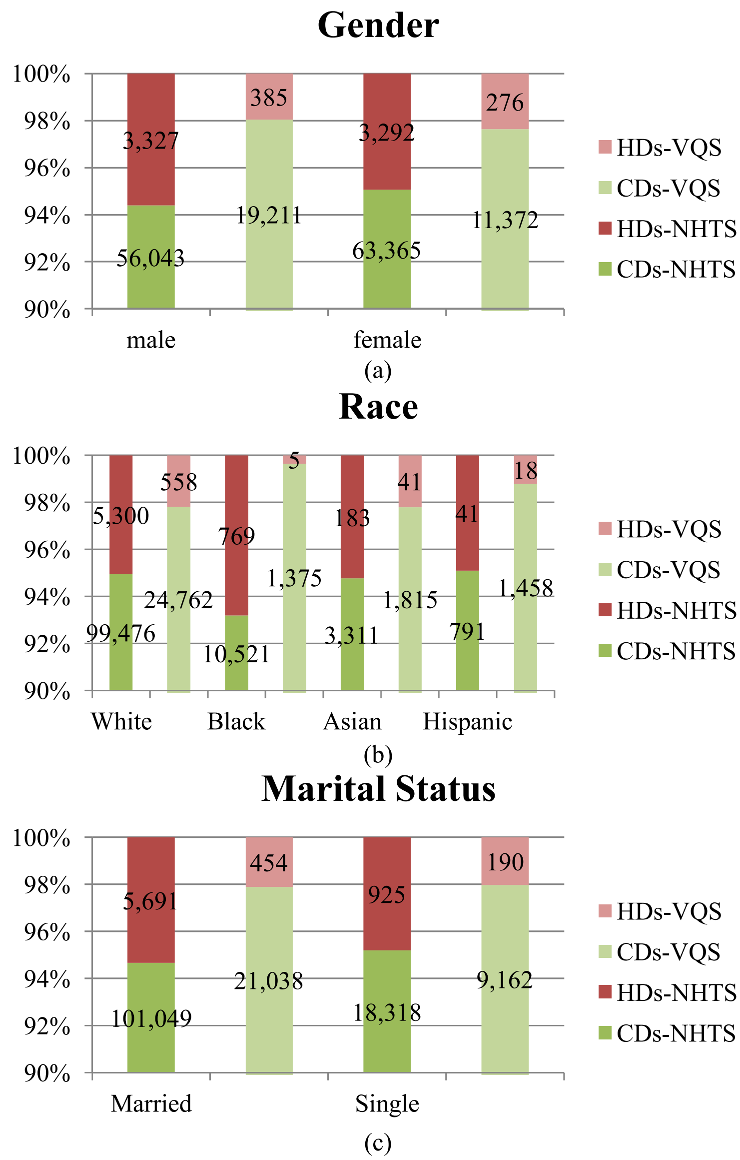


Figure 5.3: Comparison of Categorical Attributes in NHTS and VQS:

(a) Gender, (b) Race, (c) Marital Status

In summary, the NHTS data provides rich information about individuals, households, and vehicle usage, while VQS data collects details about consumers' choice and their satisfactions. Based on the compatibility study shown above, both data sets show consistent results, such as significantly higher *age*, *household income*, *education level*, and *miles driven daily* for hybrid drivers, in comparing hybrid electric vehicle drivers against conventional vehicle drivers, although discrepancies exist in the distributions of a few attributes among the sample population. Data analysis in this section yields fruitful results and brings valuable insights into the characteristics of HEV drivers. Such insights may provide guidance for future HEV designs to better serve the targeted market segments. In the following section, we will illustrate the process of using both the NHTS and VQS data for creating the usage context based choice model for HEV.

5.5 HEV Choice Modeling Based On Pooled Data

Based on the most important consumer attributes identified in Section 5.3, NHTS and VQS data are pooled for choice model estimation in this work to optimize data usage and improve the predictive capabilities of usage context-based choice modeling. While VQS is more suited for choice modeling purpose, NHTS is more representative of the real residents in the U.S. Once the model is built, the target population identified by the NHTS data in Section 5.3.2can be used for prediction. It should be noted that while a complete matching of attributes is desirable, there is no need to have the exact same set of attributes for model construction and prediction as the utility function underlying the choice model is expected to capture the impact of both usage context attributes **E** and consumer profile attributes **S** across multiple datasets.

In the 2007 Vehicle Quality Survey, vehicle purchase data from 90,000 nation-wide respondents of over 300 vehicles in the market are collected, including data for 11 HEV models. Further, respondents' demographic attributes and their vehicle usages are recorded. For model estimation, data collected from 8,025 respondents, who responded explicitly regarding three other vehicles considered in their choice set in addition to the vehicle they chose, are considered.

As for the NHTS 2009 data, 15,973 individuals with 2007 model-year vehicles are selected for pooled choice model estimation. Since this data provides no information about other vehicles considered by the respondents, three vehicles other than the one purchased are randomly selected from a set of 262 car models based on a uniform distribution to compose an individual choice set of four vehicles[1].

Table 5.4: List of Attributes in Usage Context-Based Choice Modeling for HEVs

Customer-desired product attributes A		**VQS**	**NHTS**
A_1	Price	*	*
A_2	MPG	*	*
A_3	Vehicle origin	*	*
A_4	Vehicle size	*	*
A_5	Vehicle type	*	*
A_6	Hybrid electric vehicle	*	*
$A_{exterior}$	Exterior attractiveness	*	
$A_{interior}$	Interior attractiveness	*	
$A_{storage}$	Storage and space usage	*	
A_{audio}	Audio	*	
A_{seats}	Seats	*	
A_{hvac}	HVAC	*	
$A_{dynamics}$	Driving dynamics**	*	
A_{engine}	Engine and transmission	*	
A_{safety}	Visibility and safety	*	
Usage context attributes E			
E_1	Local / highway indicator	*	
E_2	Miles driven daily	*	*
Customer profile attributes S			
S_1	Gender	*	*
S_2	Age	*	*
S_3	Income	*	*
S_4	Children	*	*
S_5	Education	*	*
S_6	Race	*	*
S_7	Marital status	*	*

[1] McFadden (1978) has shown that the uniform-conditioning property ensures that a multinomial logit model estimate using choice sets composed of randomly selected members drawn with a uniform distribution from the set of all choice alternatives will result in consistent estimates of the model parameters.

**** Driving dynamics surveys consumers' perception of ride smoothness in normal driving, quietness over harsh bumps, responsiveness / effort of steering system and braking, handling / stability on curves or winding roads and in adverse conditions.**

The attributes included in the choice model and their availability in VQS and NHTS are listed in Table 5.4. The asterisks in the last two columns indicate that whether the attributes are available in VQS and NHTS, respectively. Fifteen consumer-desired vehicle attributes **A** are selected including price, vehicle origin, vehicle size, vehicle type, mileage per gallon (mpg), hybrid electric vehicle indicator, and nine rating scores given by the respondents. The attribute "price" is the money respondents paid, excluding tax, license, trade-in value, etc. Vehicle origins are categorized as domestic, European, Japanese, and Korean; vehicle sizes are grouped into compact, midsize, large, and premium; vehicle type includes mini, car, sport utility vehicles (SUV), minivan, van, multi-activity vehicles (MAV), and pickup. The hybrid electric vehicle indicator, coded as one for hybrids, and zero for conventional vehicles, reflects consumers' attitude toward new hybrid technology. In VQS, nine aspects of the vehicle, including exterior attractiveness, interior attractiveness, storage and space usage, audio/entertainment/navigation system, seats, heating ventilation and air conditioning, driving dynamics, engine and transmission, and visibility and driving safety, are rated in an ascending scale from one to ten. These discrete ratings are included in the choice modeling procedure under **A**, because they are considered to be a good measure of consumers' perception of qualitative as well as quantitative vehicle attributes.

As for the vehicle usage attributes **E**, two most commonly considered vehicle usage attributes for HEV are included in the choice model: *local/highway indicator* and *miles driven daily*. It should be noted that the local/highway indicator is imputed by comparing the combined mpg published by the US Environmental Protection Agency (EPA, 2008) and the estimated mpg given by survey respondents in the VQS data. The indicator is a continuous parameter, ranging from zero (0) for local driving to one (1) for highway driving, and assumed to reflect the general driving conditions the respondents face, therefore the vehicle usage. Hence, as part of the level III in the hierarchical choice modeling framework, consumer desired attribute A_2, *mileage per gallon*, is shown in Eqn. (4.9). The other vehicle usage attribute considered is the *miles driven daily*, a popular descriptor of consumers' travel pattern. The data is derived from the recorded miles driven in the first three months from the market survey.

G*ender*, *age*, *household income*, *number of children under age 20 living together*, *education level, race, and marital status*, are included as consumer profile attributes **S**. Numerous combinations of consumer attributes are tested in the choice modeling process. From the final choice model estimation results, only two consumer profile attributes, *household income* and *education level*, are statistically significant. All consumer profile attributes are included in the ordered logit regression for predicting the performance rating scores. All correlations between consumer profile and vehicle usage attributes are between -0.34 and 0.32, justifying the inclusion of multiple attributes in choice model.

The utility function used is shown in Eqn. (5.11), where interactions between **A**, **E**, and **S** are explicitly modeled. W_1, W_2, and W_{pooled} stand for utility function in NHTS, VQS, and pooled data. As seen in (5.11), attributes in choice modeling include common consumer-desired attributes $\mathbf{A}^{com}$, VQS-specific $\mathbf{A}^2$, common vehicle usage attributes $\mathbf{E}^{com}$, VQS-specific $\mathbf{E}^2$, and common consumer profile attributes $\mathbf{S}^{com}$. A scale parameter μ is introduced in the pooled utility, W_{pooled}, to account for variation difference of error terms from two datasets. The structure of pooled utility function for choice modeling is similar to that used in nested logit model

(Koppelman and Bhat, 2006). Alternative specific constants (ASC) for each of the car models are not included in the utility function. While this may decrease the goodness-of-fit of the model, it allows choice prediction of newly introduced vehicle.

$$
\begin{aligned}
W_1 &= \beta_{1,A} \cdot \mathbf{A}^{com} + \beta_{1,A\cdot S}\mathbf{A}^{com} \cdot \mathbf{S}^{com} + \beta_{1,A\cdot E}\mathbf{A}^{com} \cdot \mathbf{E}^{com} \\
W_2 &= (\beta_{2,c}\mathbf{A}^{com} + \beta_2\mathbf{A}^2)(1 + \gamma_S\mathbf{S}^{com} + \gamma_E\mathbf{E}^{com} + \gamma_2\mathbf{E}^2) \\
W_{pooled} &= W_1 + \frac{1}{\mu}W_2
\end{aligned}
\tag{5.11}
$$

Table 5.5: Coefficients of pooled choice model estimation

Attributes	Coefficients
Price(k)/income	-0.1194*
100/miles per gallon	-0.5338*
100/miles per gallon * local/highway indicator (VQS only)	-0.5613*
100/miles per gallon * miles driven daily	-0.0012*
Vehicle origin (Domestic as base)	
European	0.4729*
Japanese	0.1544*
Korean	0.2816*
Vehicle size (Compact as base)	
Large	-0.1855*
Medium	-0.0167*
Premium	0.1240*
Vehicle type (Car as base)	
MAV	-0.3463*
Mini	0.1222*
Minivan	-0.2939*
Pickup	-0.0110
SUV	-0.1680*
VAN	-1.8446*
Minivan	0.0713*
Hybrid electric vehicle	1.6985*
Hybrid electric vehicle * local/highway indicator (VQS only)	-1.9586*
Hybrid electric vehicle * miles driven daily	-0.0010
Hybrid electric vehicle * education level	0.0518
Rating (VQS only)	
Exterior attractiveness	0.0131*
Interior attractiveness	0.1069*

Storage and space usage	0.1297*
Audio	0.0363*
Seats	0.0333*
HVAC	0.0273*
Driving dynamics	0.0495*
Engine and transmission	0.0737*
Visibility and safety	0.0154*

*** Significant at 0.01;**

From the results of MNL including E attributes in modeling shown in Table 5.5, we note that all coefficients are significant with p value less than 0.05. The coefficient for price/income is negative as expected. The second attributes $100/A_2$ reflects the amount of gasoline needed to drive 100 miles. A negative estimator for E_1*100/A_2 indicates that the usage context attribute E_1 (local/highway indicator) has a negative impact on consumers' preference on inverse of MPG measure. In other words, people primarily driving on highways tend to care more about the MPG value. Moreover, the attitude toward HEV itself has a positive coefficient estimator of 3.1476, which shows that people driving locally tend to favor HEV. Similarly as we expected, highway drivers' preference towards HEVs are weaker, as shown in the negative coefficient estimator of the E_1 and HEV indicator interaction (E_1*A_6). The positive estimator of education HEV indicator interaction (S_5*A_6) suggests that people with higher education are more likely to prefer HEVs, which is consistent with the finding in Section 5.3. All correlations between HEV indicator and usage context attributes are small, within range of [-0.02, 0.05].

Goodness-of-fit measures based upon the log-likelihood of the converged model, such as the likelihood ratio index ρ^2 (also known as pseudo R-square), reflect how well the estimated model predicts actual individual choices in the data set. Higher values of ρ^2 indicate better predictions of the choices. A multinomial logit model without usage context attributes is used here as a comparison. A slightly higher log-likelihood of -38957.776 and subsequently ρ^2 value of 0.2983 are achieved using the MNL model with usage context attributes **E** versus -39420.684 and 0.2900 in the MNL model without **E**. This implies that introducing the usage context attributes in choice modeling has captured the systematic taste heterogeneity of consumers under different usage context.

5.6 HEV/PHEV Design Selection

Plug-in hybrid electric vehicle technology is considered a potential near-term approach to address global warming and U.S. dependency on foreign oil in the transportation sector as the cost, size, and weight of batteries are reduced. PHEVs use the large battery packs to store energy from the electricity grid and propel the vehicle partly on electricity instead of gasoline. Under the average mix of electricity sources in the U.S., vehicles can be driven with lower operation cost and fewer greenhouse gas (GHG) emissions per mile when powered by electricity rather than by gasoline. PHEVs have the potential to displace a large portion of the gasoline consumed by the transportation sector with electricity since approximately 60% of U.S. passenger vehicles travel less than 30 miles each day. Several automobile manufacturers have announced plans to produce PHEVs commercially in the future, including General Motors' Chevrolet Volt, which will carry enough battery modules to store 40 miles worth of electricity and Toyota's plug-in version of the Prius, which will carry enough batteries for approximately 13 miles of electric travel. Shiau (2009a) built a vehicle engineering model based on Powertrain System Analysis Toolkit (PSAT) simulation data, in which the total lifetime cost and fuel economy of the vehicle design are modeled as functions of battery capacity.

In this design example, the vehicle performance measures including lifetime cost and fuel economy are integrated into consumers' choice through the proposed Choice Modeling for Usage

Context-based Design. Due to the focus on vehicle fuel economy performance and lifetime cost measure, we consider the following selected attributes from Table 5.5: price(k)/income (P/INC), 100/miles per gallon (MPG), 100/miles per gallon * local/highway indicator (100/MPG*HWY), 100/miles per gallon * miles driven daily (100/MPG*MDD), hybrid electric vehicle (HEV), hybrid electric vehicle * local/highway indicator (HEV*HWY), hybrid electric vehicle * miles driven daily (HEV*MDD), hybrid electric vehicle * education level (HEV*EDU). Because only difference matters in the choice utility function, we assume that other attributes are equal. The choice utility function listed in Eqn. (5.11) can be simplified and substituted with coefficients from Table 5.5, as shown in the following Eqn. (5.12).

$$\begin{aligned} W = & -0.1194 \times P / INC - 0.5338 \times 100 / MPG \\ & -0.5613 \times 100 / MPG * HWY - 0.0012 \times 100 / MPG * MDD \\ & +1.6985 \times HEV - 1.9586 \times HEV * HWY \\ & -0.0010 \times HEV * MDD + 0.0518 \times HEV * EDU \end{aligned} \tag{5.12}$$

Now let us consider an average consumer with household income level 5, without higher education (college, or graduate school), and mixed driving condition (half local, half highway). Meanwhile, a range of miles driven daily from 0 to 100 miles is considered here to study consumers' choice among CV, HEV, and PHEV. Based on the pricing of current market offerings, vehicle base costs of \$17,600, \$25,600, and \$35,600 are used for CV, HEV, and PHEV, respectively. Figure 5.4 shows the trend of *exp(W)* with respect to miles driven daily for PHEV10 (PHEV design with AER of 10 miles), PHEV20, PHEV30, PHEV40, PHEV50, PHEV60, HEV, and CV. According to utility maximization theory, the vehicle design with highest *exp(W)* has the highest choice probability. From the figure, we can see that PHEV10 dominates the market from 10 miles driven daily to somewhere less than 20 miles driven daily, as PHEV20 takes over. As electricity is more cost effective than gasoline, the ideal vehicle choice would have an AER close to the miles driven daily to maximize the advantage of electric vehicle. However, as the battery capacity increases, the cost of battery comes in and drags down the benefits of PHEV. At about 21 miles driven daily, the utility of HEV surpasses the one of PHEV20 and continues to be the most likely choice for increasing miles driven daily.

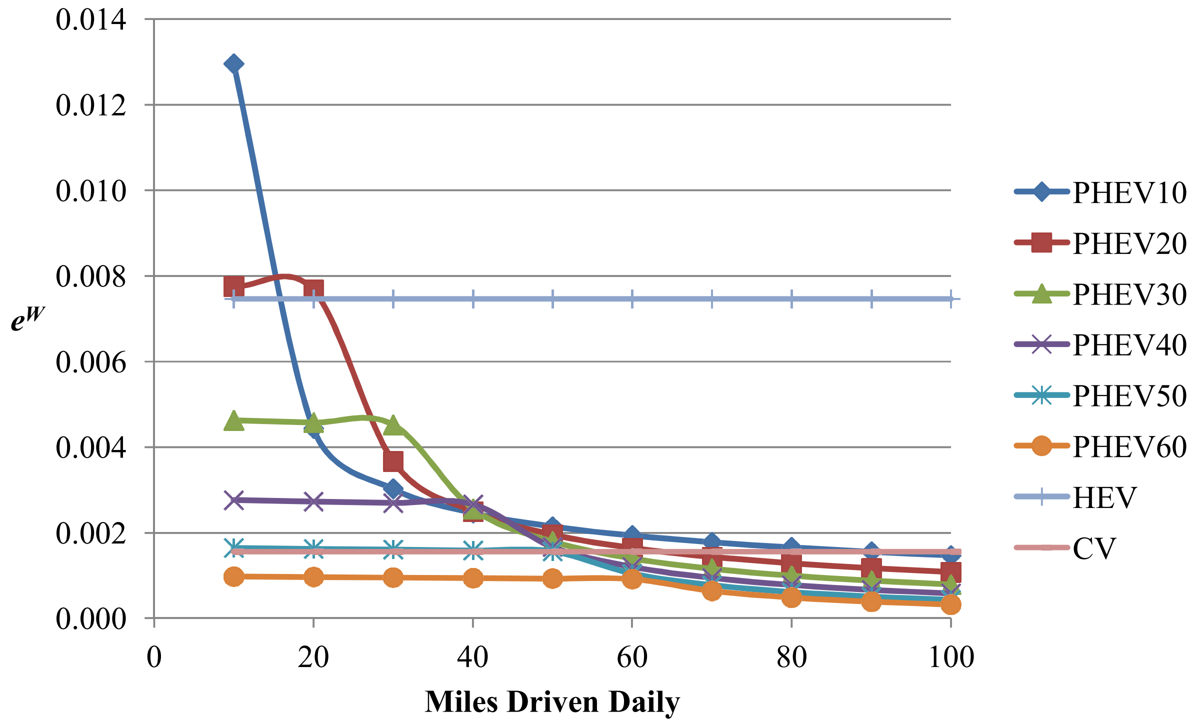


Figure 5.4: Predicted Utility for PHEV, HEV, and CV

Moreover, if consumer were to choose from a CV, a HEV, and a PHEV, the predicted choice probability are plotted in Figure 5.5. For example, when the consumer drives 20 miles on a daily basis, the predicted choice probability of PHEV20 would be 45.95%, compared with 44.71% for HEV, and 9.34% for CV. Interestingly, when the miles driven daily increases to over 50 miles, PHEV with large battery capacity is no longer desirable probably due to their higher cost.

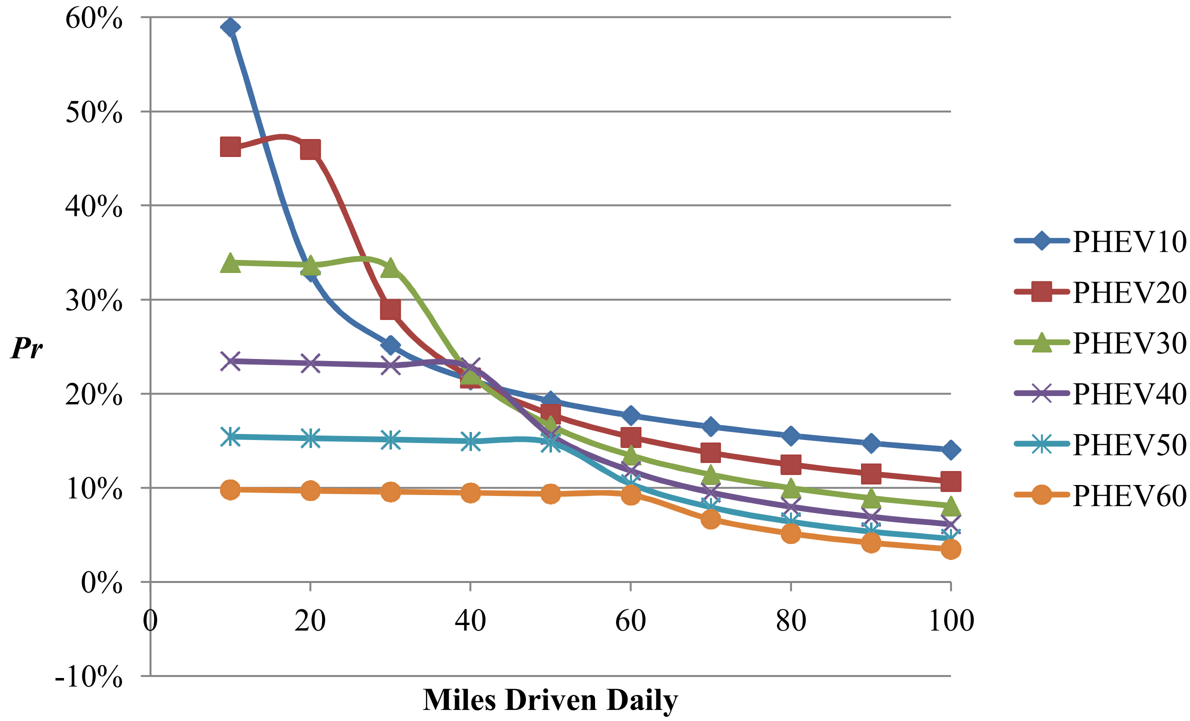


Figure 5.5: Predicted Choice Probability for PHEVs

5.7 Conclusions and Discussions

In this chapter, a systematic procedure and the associated data analysis techniques are presented for choice modeling of complex engineering systems that involve a large number of consumer attributes including both the usage context attributes and consumer profile attributes. For choice modeling of HEV, both the NHTS and VQS data are studied. To understand the relationship between usage context and consumer profile, an analysis of variance is performed on the usage context attribute *miles driven daily*. Interesting trends are seen in the one-to-one plots of the usage context attribute versus consumer profile attributes. The ANOVA result suggests that usage context is an additional dimension of overall consumer classification since the consumer profile attributes cannot fully explain the difference in usage context. Through the Principal Component

Analysis, 40 consumer profile and usage context attributes are grouped into 18 distinctive factors such as drivers and vehicles, mainstream family, working class, high income and education, etc. Key characteristics of hybrid electric vehicle drivers are identified by comparing the mean factor scores within the hybrid driver and conventional driver groups, which laid the foundation for segmenting the market based on both consumer profile and usage context.

While NHTS collects rich and detailed data on consumer profile and vehicle usage context, VQS data provides information of individual's choice set and their ratings towards multiple vehicle attributes. The compatibility of the two datasets was tested by comparing common attributes shared by both. The results confirm that the differences in consumer profile and usage context attributes between hybrid and convention driver groups are consistent in NHTS and VQS, even though some discrepancies exist due to the difference in the sample population. By including the list of key attributes identified through the principal component analysis, a usage context based choice model is created using pooled NHTS and VQS data. All coefficients from the choice model estimation are statistically significant with expected signs, which provide valuable insights into consumer preference for HEVs.

The HEV study in this work illustrates how NHTS and VQS can be "pooled" together to maximize the predictive capabilities and benefit the study of consumer preferences for hybrid technology and therefore to support the design of HEVs. Key HEV driver characteristics identified in this study provide guidance to market segmentation studies. Moreover, choice model results based on pooled NHTS and VQS data demonstrate the potential of modeling consumer preferences by linking their choices to both usage context and consumer profiles. As demonstrated in the vehicle design selection study, the consumers' preference toward HEVs modeled in this case study is used as a basis for the choice modeling of Plug-in Hybrid Electric

Vehicles (PHEVs) as HEVs and PHEVs share many common consumer desired product attributes associated with the new vehicle technology. The findings presented in this chapter illustrate how the choice modeling framework proposed in Chapter 4 can be applied to the study the early adoption of the new products such as alternative fuel vehicles by evaluating consumers' choices among conventional vehicles, hybrid electric vehicles and plug-in hybrid electric vehicles.

Chapter 6

AGENT-BASED CHOICE MODELING CONSIDERING SOCIAL CONTEXT

6.1 INTRODUCTION

While the use of Discrete Choice Analysis (DCA) is prevalent in capturing consumers' preferences and describing their choice behavior in product design (Li and Azarm, 2000, Wassenaar and Chen, 2003, Frischknecht et al., 2010, Michalek et al., 2006a, Williams et al., 2008), individuals' choices are studied without their social contexts in most cases. Empirical studies show that social context, such as "neighbor" effects may impact consumer's choice behavior (Case, 1992). Often times, social context influences consumers' attitudes towards new products, such as those involving green technology. As an example, a consumer's decision in choosing an eco-friendly alternative fuel vehicle like hybrid electric vehicle (HEV) or plug-in hybrid electric vehicle (PHEV) may be largely influenced by neighbors and friends or others who share similar social status or profile. In the broad market of consumer products, a large amount of product reviews and recommendations are now made available through the rapid growing online shopping websites and social networking sites, which further accelerates the social impact on product adoption. Integrating social network information into consumer choice modeling and developing methods for predicting the social influence on consumer choices and their attitudes towards adopting new green products is the focus of this research.

Consumer choice modeling is essential in engineering design because it allows for the prediction of future product demand as a function of engineering design and the target market across a heterogeneous consumer population. Capturing heterogeneous choice behavior can be achieved using disaggregate demand modeling methods, with the probabilistic DCA (Li and Azarm, 2000) being the most widely used approach. Depending upon the degree of heterogeneity and the specific design problem, different types of DCA models, such as multinomial logit models (Hausman and McFadden, 1984), nested logit models (Koppelman and Sethi, 2000), and mixed logit models (Train, 2003), have been utilized in engineering design to capture the heterogeneity in consumer preferences resulting from many aspects beyond traditional engineering considerations. Hoyle et al. (Hoyle et al., 2011) modeled both systematic and random consumer heterogeneity using a hierarchical Bayes mixed logit model. This modeling framework is further expanded by He et al. (He et al., 2010) to consider usage context attributes as a part of consumer attributes in usage context-based design. Recent research in demand modeling for engineering design has extended general demand modeling methodologies to understand preference inconsistencies in consumer's choice of "green" products (MacDonald et al., 2009), optimal design under price competition (Shiau and Michalek, 2009), preferences for aesthetic forms (Orsborn et al., 2009), and the use of latent class analysis (Sullivan et al., 2011). While the existing work demonstrated the benefits of using DCA in modeling consumer choice behavior, the merits of DCA are limited due to its assumption of consumers making individual decisions isolated from each other, which is contradictory to the real world situation.

As many behavioral economists and psychologists have noted, *choice is social*. In other words, an individual's decisions are not immune from the influence of others. This is especially the case in forecasting the adoptions (first-time purchases) of new green products, which is a critical but challenging task. Because of its potential impact on energy and environment, there are a handful of research works on forecasting the HEV/PHEVs' market potential as the vehicle design evolves and the green technology matures. Rousseau et al (Rousseau et al., 2007) researched the impact of all electric range, drive cycle, and control strategy on battery requirements. Based on the study of PHEV batteries' effect on vehicle performance and cost, Shiau et al. (2009a) suggested to target the adoption of small-capacity PHEVs by urban drivers who can charge frequently. While these works focused on the cost-performance models, a few pilot projects have been conducted to better understand consumers' knowledge and awareness of PHEV (Axsen and Kurani, 2008). For choice modeling of alternative fuel vehicles, He et al. (He et al., 2010) quantitatively assessed the impact of vehicle usage on HEV choice and demonstrated that local driving consumers tend to prefer HEV more, compared with their counterparts. Sullivan et al. (2005) suggested that consumers make purchasing decisions based on their own personal attributes as well as vehicle attributes, they later developed an agent-based simulation for modeling market penetration of PHEVs under a variety of consumer, economic, and policy conditions (2009). Existing studies nevertheless mostly focus on understanding the impact of marketing attributes, such as price, but the linkage between product design and consumers' behavior under social impact is missing. An enhanced choice modeling framework is therefore needed to bridge the gap among engineering, marketing, and social science domains.

The research objective of this work is to develop an agent-based choice modeling framework considering the social impact on new product adoption by integrating methods rooted in social network theories, agent-based modeling, and discrete choice analysis. A new agent-based choice modeling framework to capture the dynamic influence from social network on consumer adoption of new products is presented. By introducing the social influence attributes into the choice utility function, the social network simulation is integrated with the traditional discrete choice analysis by following the procedure of social network construction, social influence evaluation, and choice model estimation.

This chapter is organized as follows: In Section 6.2, we provide a review of the literature on social network theories and existing work on integrating social interaction in choice models. The proposed agent-based choice modeling framework considering social impact is presented in Section 6.3, followed by a case study of modeling hybrid electric vehicle ownership in California in Section 6.4. Discussions and conclusions are included in Section 6.5.

6.2 Social Network Theories And Integration With Choice Modeling

With the growing public awareness of the complex "connectedness" of modern society, the idea of *social network*, in which a group of people are connected to some or all of the others following a random or particular pattern in graph, has been gaining more attention (Easley and Kleinberg, 2010, Faust and Wasserman, 1994). For example, the leading online social networking site Facebook has so far attracted more than 800 million active users (Facebook), demonstrating the power of interpersonal connections in our daily lives. There are two key elements of a social network which includes *nodes*, representing members of the network, i.e. consumers in the context of product design, and *links*, illustrating the connections between members, i.e. linked consumers. Depending on the specific network structure, distinctive influences through social network can be observed, modeled, and researched in numerous domains including social science, humanities, etc. The meaning of "connectedness" encompasses two related issues in social network modeling and simulation: one is the network structure – the *media* of social impact; the other is behavioral interactions– the *mechanism* of social impact.

How to integrate these two key elements of social network into consumer behavior simulation and choice modeling are further discussed in Section 6.3.

Among the various social behavior theories proposed in literature, the *contagions theory* is the most relevant to our interest in product design. Contagion theory is based on the assumption that the opportunities for contact provided by social networks serve as a mechanism that exposes individuals, groups, and organizations to information, attitudinal messages, and the behavior of others (Burt, 1987, Contractor and Eisenberg, 1990). This exposure increases the likelihood that social network members will develop beliefs, assumptions, and attitudes that are similar to others in their network (Carley, 1991). In (Manski, 1995), Manski defined the endogenous effects, contextual effects, and correlated effects in explaining the observation that individuals belonging to the same group tend to behave similarly. McFadden (McFadden, 2010) decomposed the causes of this sociality of choice by stating that choice is influenced by information from a peer group, heuristics rooted in the behavior of others, analogies or anecdotal information garnered from associates, and constraints imposed by others. Both scholars, along with many others, emphasized the importance of incorporating social influence in choice models.

The economists are among the pioneers in quantitatively considering interdependence of various decision-makers' choices (Aoki, 1995, Brock and Durlauf, 2001, Blume et al., 2003). They introduced social interactions in binary discrete choice models by allowing a given consumer's choice for a particular alternative to be dependent on the alternative's overall market share, i.e. the global social network effects. Later, the results on the behavior of binary logit models are extended to multinomial logit models (Brock and Durlauf, 2002) and nested logit with global effects (Dugundji and Gulyas, 2003a). Since the global social network effects would be perfectly correlated with a set of alternative-specific constants in a discrete choice model, Dugundji and Gulyas (Dugundji and Gulyas, 2003b) presented a more general framework for studying local social network effects in discrete choice models, where social network effects are calculated within each market segment.

In transportation field, Dugundji and Gulyas (Dugundji and Walker, 2005) utilized simulated data for modeling intercity travel behavior by considering local social network effects into the choice utility with varying network density. Páez et al. (Páez et al., 2008) presented a multinomial logit model of residential location choice using simulated network data with varying degrees of distributions and clustering parameters. To capture social influences without explicit knowledge of the individual networks, Walker et al. (Walker et al., 2011) introduced a local social network effect in choice model, i.e. the percent of population choose the specific alternative within the peer group defined based on socio-economic status (income, education, age) and spatial proximity of residential location.

While the flourishing publications in economics and transportation field provided the theoretical foundation to incorporate social interactions into the choice modeling, a major limitation of the above-mentioned methods is the assumption that the social impact is the only critical factor in the choice model. In many fields including product design, other factors are shown to be critical in consumers' choice including product performance, characteristics of consumers, as well as the product usage contexts. Hence, research is needed to address the unique challenges in incorporating social influences in choice modeling for product design applications.

6.3 AGENT-BASED CHOICE MODELING FRAMEWORK CONSIDERING SOCIAL IMPACT

To address the aforementioned limitations in existing work, an agent-based choice modeling framework is developed in this research for product design to quantitatively capture the impact of social network on consumer choice behavior. In this section, a three-stage agent-based choice modeling framework considering the social network impact is presented. The adoption of alternative fuel vehicles is used as an example to explain the proposed agent-based choice modeling framework.

The proposed choice modeling framework consists of the following three stages: I) social network construction, in which a virtual environment where consumers interact with their linked friends is created; II) social influence evaluation, where the network influence is evaluated in the form of social influence attributes by simulating how consumers communicate with their linked friends and the accumulated influence each receives; III) choice model estimation, in which consumers' rational decisions based on the utility maximization theory are studied to quantify the impact from social influence modeled in stage II together with other product and consumer attributes. The details of each stage are discussed as follows.

6.3.1 Stage I: Social Network Construction

Social network construction is the process of creating network links based on existing data - the predefined network structure, and the descriptive information associated with each node (consumer). Due to the complexity in social network data collection, it is often the case that there is limited data, if any, on the real social network structure among the sample population used in data collection for choice modeling. An alternative is to construct the network through simulations based on certain hypotheses of a network structure using collected consumer attributes such as the socio-demographic and usage context attributes. In our case study of HEV, the National Household Travel Survey (NHTS) is the data source from which the demographic and usage context attributes of the population are obtained. Examples of commonly used predefined social network structure include *small-world network* with short average path length (Watts and Strogatz, 1998), and *scale-free network* with power-law degree distribution (Barabási and Albert, 1999).

Within the network construct, links representing connections in a social network are generated to simulate interpersonal interactions. Rogers (1995) suggested that interpersonal influence can occur among individuals who are homophilous (i.e., similar to each other) or who are heterophilous (i.e., dissimilar to each other). Homophilous connections, or close links, represent neighbors, friends, and other regular contacts (e.g. coworkers) who are connected, whereas heterophilous connections, or distant links, represent acquaintances and other information sources (e.g. online reviewers). Foundational to simulating the social influence is the concept of social distance, which is defined as the distance between locations of two nodes (consumers) in a social geography, i.e. social space, the concept of which is inspired by Krugman's work on economic geography (Krugman, 1990). A social geography can be constructed based on attributes used to describe consumers' social dimensions. As shown in Eqn. (6.1), d_{ij} is defined as the p-norm distance in social space between consumer i and j, while x_i^m represents the attributes in the m-dimensional social space. The attributes of consumer social dimension include, but are not limited to, consumer profile attributes **S**, as well as usage context attributes **E**.

$$d_{ij} = \left(\sum_m \left| x_i^m - x_j^m \right|^p \right)^{1/p} \tag{6.1}$$

As suggested by Akerlof (Akerlof, 1997), two nodes with shorter social distance are more likely to be connected. Their method implements a distance-decay function in social space to reflect the hypothesis that the degree of influence between nodes should decrease as their opinions or behavior become more dissimilar (Festinger, 1954). The strength of a connection in social space is a function of the interacting distance between consumers, shown as follows (Páez et al., 2008):

$$l_{ij} = \begin{cases} \gamma_1 \exp(-\gamma_2 d_{ij}^2), & \text{for } i = j \\ 0, & \text{for } i \neq j \end{cases} \tag{6.2}$$

As shown in Eqn. (6.2), γ_1 is the parameter controlling magnitude of the effect and γ_2 is the one controlling the rate of decay. In many cases, the social links are binary, 1 for "with link", and 0 for "no link". In practice, the distance decay function in Eqn. (6.2) is replaced with a significance criterion of a given threshold, as shown in Eqn.(6.3).

$$L_{ij} = \begin{cases} 1, \text{if } l_{ij} \geq a_i \\ 0, \text{otherwise} \end{cases} \tag{6.3}$$

where a_i is the threshold value, which can be determined by the model analyst. More detailed discussion of the significance criterion function can be found in (Leenders, 2002). Other treatments such as relative connection strength can be used to model more complex network structures.

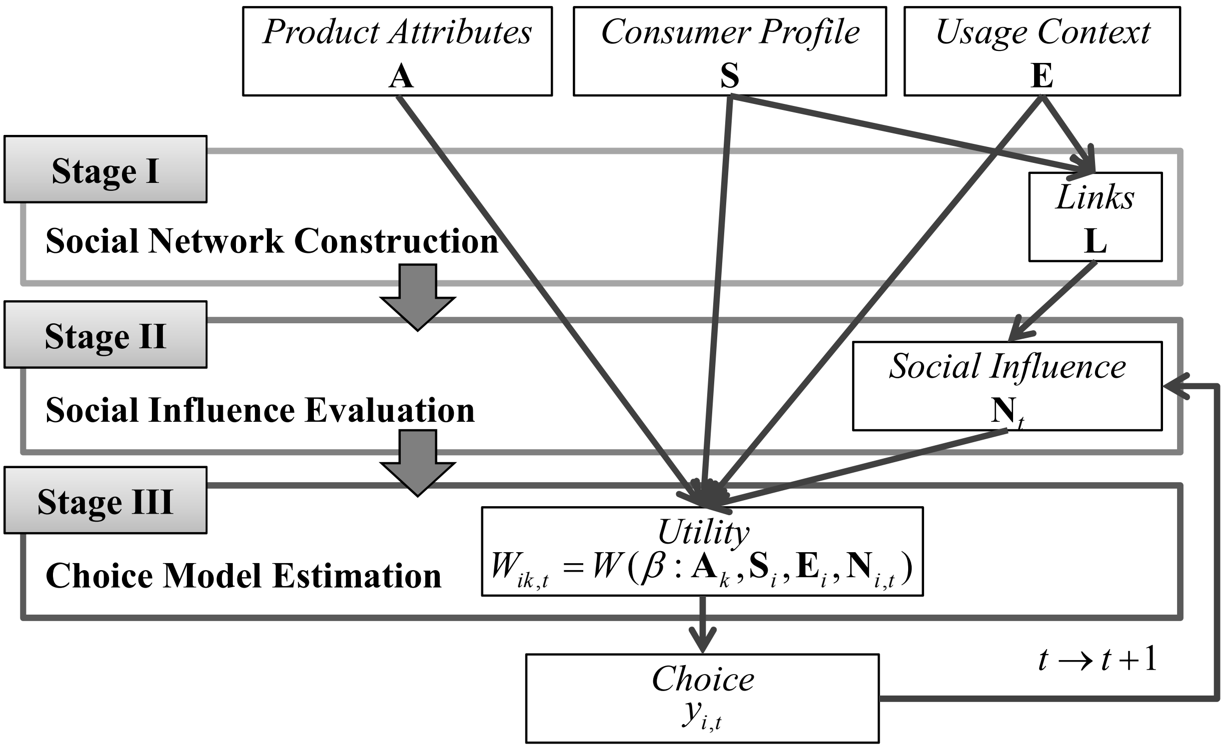


Figure 6.1: Agent-based Choice Modeling Framework Considering Social Impact

6.3.2 Stage II: Social Influence Evaluation

Once a network is constructed, social influences on consumer preferences are evaluated by simulating the interactions between consumers in the network. As a result, social influence attribute $\mathbf{N}_{i,t}$ is defined as the collection of influences from all other consumers linked to a focal consumer *i* at time *t*. Because the social influence is a function of time *t*, **N** is evaluated for each of the time periods. This attribute is introduced in the consumer choice utility function to capture the social influence on product adoption. Three types of social network influence are studied in literature (Snijders, 2001): structural effects on network dynamics, effects on network dynamics associated with covariates, and effects on behavior evolution. In this study we assume the structure of the social network constructed in Stage I is stable over time, i.e., links among consumers do not change over time. Hence, only the third type, effect on behavioral evolution, is relevant in this work. Formulations for evaluating the social influence attribute associated with several popular effects on behavior evolution are presented in Table 6.. For example, the average friend effect is defined as the average degree of impact from linked contacts with similar behavior. Because social behavior y (in this case choice behavior) of other consumers in the network changes over time, the average friend effect is updated at each time iteration t, as detailed in Section 6.4. Considering multiple popular effects in network modeling is cautioned due to the possible high correlations among these effects.

Market surveys or interviews are often needed to better understand the importance of social influences and which effect in Table 1 is the most relevant to the problem of interest. For instance, some individuals are highly influenced by their neighbors, coworkers, or other close contacts with whom they communicate daily, while other people are likely to trust suggestions and advices from their remote contact, such as online product reviews, or blog posts from people with the same lifestyle as his/hers. These effects are modeled as close links and distant links respectively in the small word network used in our case study, as detailed in Section 6.4.1.

Table 6.1: Commonly Considered Effects On Behavioral Evolution

Behavioral effects	Definition	Mathematical formulation
Tendency effect	Individual constants representing basic tendency.	$N_i = c_i$
Average similarity effect	Average degree of consumer being similar to their linked neighbors.	$N_i = \sum_j l_{ij} / \sum_j L_{ij}$
Total similarity effect	Total degree of consumer being similar to their linked neighbors.	$N_i = \sum_j l_{ij}$
Average friend effect*	Average degree of impact from linked contacts with similar behavior y.	$N_{i,t} = y_{i,t-1} \sum_j L_{ij} y_{j,t-1} / \sum_j L_{ij}$
In-degree effect	Number of neighbors linked to a consumer.	$N_i = \sum_j L_{ji}$
Out-degree effect	Number of neighbors linked from a consumer.	$N_i = \sum_j L_{ij}$

In the social network literature, contagion mechanisms have been used to explain the attitudes as well as the behavior of network members. Erickson (1988) offers a comprehensive overview of the various theories that address the "relational basis of attitudes" (p.99). She describes how various network metrics such as frequency, multiplicity, strength, asymmetry can shape the extent to which others influence individuals in their networks. She also describes cohesion and structural equivalences models that offer alternative, and in some cases complementary, explanations of the contagion process. *Contagion by cohesion* implies that the attitudes and behaviors of the other to whom they are directly connected influence network member. *Contagion by structural equivalence* implies that others who have similar structural patterns of relationships within the network influences consumers. Both mechanisms provide some insights to interpersonal influence on attitudes towards new technology. The contagion by cohesion is modeled by formulating social influence attributes **N** as a function of the binary link variable L_{ij} (Eqn. (6.3)), as will be demonstrated in the case study (Section 6.4).

6.3.3 Stage III: Choice Model Estimation

In the traditional choice modeling framework, a predictive model of demand Q is established using DCA, which is based upon the assumption that individuals seek to maximize their personal *consumer choice utility*, u, when selecting a product from a choice set. The choice utility is derived by assuming that the individual's (i) true choice utility, u, for a design alternative, k, consists of an observed part W, and an unobserved random disturbance ε (unobserved utility):

$$u_{ik} = W_{ik} + \varepsilon_{ik} \text{.} \tag{6.4}$$

As shown in Eqn. (6.5), the observed or deterministic part of utility W_{ik} is expressed as a function of consumer desired product attributes $\mathbf{A}_{ik}$, of respondent i, alternative k, usage context attributes $\mathbf{E}_i$, and consumer profile attributes $\mathbf{S}_i$ of respondent i,

$$W_{ik} = W(\beta : \mathbf{A}_{ik}, \mathbf{E}_i, \mathbf{S}_i), \qquad (6.5)$$

In this work, we introduce a new element, the social influence attributes $\mathbf{N}_{i,t}$, into the utility function as shown in Eqn. (6.6).

$$W_{ik,t} = W(\beta : \mathbf{A}_{ik}, \mathbf{S}_i, \mathbf{E}_i, \mathbf{N}_{i,t}) \qquad (6.6)$$

Also new to the above utility function is the time variable t, which reflects the dynamic nature of social network influence: in each time period, consumers make decisions, which may change the social influence attributes of their linked contacts. Hence, the coefficients β are estimated based on the collected market data over multiple time periods. From the observed utility, $W_{ik,t}$, the probability P_{ik} of an individual i choosing a given alternative k, and the resulting choice behavior $y_{i,t}$ can be estimated.

The information flow in the three-phase diagram (Figure 6.1) shows how the usage context **E** and consumer profile **S** are first mapped to the interpersonal links **L** (Eqn. (6.3)), then to social influences **N**. In the last stage of choice model estimation, the product attributes **A**, together with consumer profile **S** and usage context **E**, as well as newly introduced social influence attributes **N**, comprise the explanatory variables of a choice utility function. For each time period, stage II and stage III are repeated to capture the changing dynamics of social influence.

The parameters for discrete choice analysis are estimated with market data to capture consumer behavior and model consumers' choices for new products. The procedure is similar to conventional choice model estimation except that the coefficients of social influence attribute

$\mathbf{N}_{i,t}$, need to be calibrated to match with the sales data from multiple time periods. Determining the time dependent social influence attribute $\mathbf{N}_{i,t}$ for each individual is challenging as it is often the case that no empirical data of network links L_{ij} exists at the individual, i.e. disaggregate, level to support the model estimation process. Instead, aggregate sales data $M_{r,t}$ throughout multiple time periods is used to identify aggregated social influence attribute $\mathbf{N}_t$ to ensure that the integrated choice modeling reflects the real market at the aggregate level. The same approach can also be applied to sensitivity analysis with different parameter settings of the social network structure, such as the clustering coefficients and shortest path length in the small-world network simulation of our case study in Section 6.4.1.

To validate the created choice model, the one-step-ahead approach can be applied, where the sales data at the end of time period $M_{r,t}$ is excluded from the estimation process, but later is used to validate the accuracy of the model prediction. A comparison with the prediction from the existing widely accepted benchmark product diffusion models, such as Norton-Bass model (1987) can be a possible way to validate the model. After estimation and validation, the agent-based choice modeling framework can be used for product design optimization to explore potential outcomes of different new product designs in response to distinctive social impacts.

6.4 Case Study Of Green Product Adoption: HEV Ownership In California

National Household Travel Survey (NHTS) data is used as the market data in this study, including demographic characteristics of households, people, vehicles, and detailed information on daily travel in the U.S. for all purposes by all modes (FHWA, 2009). The data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and household attributes. Due to the popularity of HEV in

California state, California has the largest population of HEV owners, which is evident in the NHTS 2009 data. The sample population of 41,330 respondents from California is used as data source for the case study presented in this section. It should be noted that the impact of federal and state level HEV policies and other purchase incentives is not considered, because only California respondents are included in this study.

6.4.1 **Social Network Simulation**

In the first two stages of the agent-based choice modeling framework, a small-world network is simulated based on the California sample population from the NHTS data. The *Small-World Network* (Watts and Strogatz, 1998) is a type of mathematical graph in which most nodes are not neighbors of one another, but can be reached from every other by a small number of steps, as shown in Figure 6.2(a). Two other types of networks: regular nearest neighbor network in (b) and random network in (c), are provided in comparison. Small-Word Network offers a mechanism to represent interpersonal influences of both close and distant links within a social network – consumers are connected to their near "neighbors" (in their social space), as well as a small number of consumers far away from them. Such networks have two important characteristics: (1) they have a high clustering coefficient, i.e., two randomly chosen consumers in the network who happen to be linked to another consumer have a high probability of also being linked to each other; and, (2) they have a small path length, i.e., the average distance between any two consumers in the network, measured as the number of links of the shortest path (with the fewest links) connecting them, is small. The popular notion that any two people in the world are connected by short chains of connections (i.e. an average of six degrees of separation (Watts, 2004)) is a reflection of the short path lengths in the small-world networks. The interpersonal links L_{ij} are generated using the small-world network, where each individual or household is connected to many other individuals or households nearby with close links, and, at the same time, has distant links to other individuals or households who are different in terms of the demographics and usage patterns.

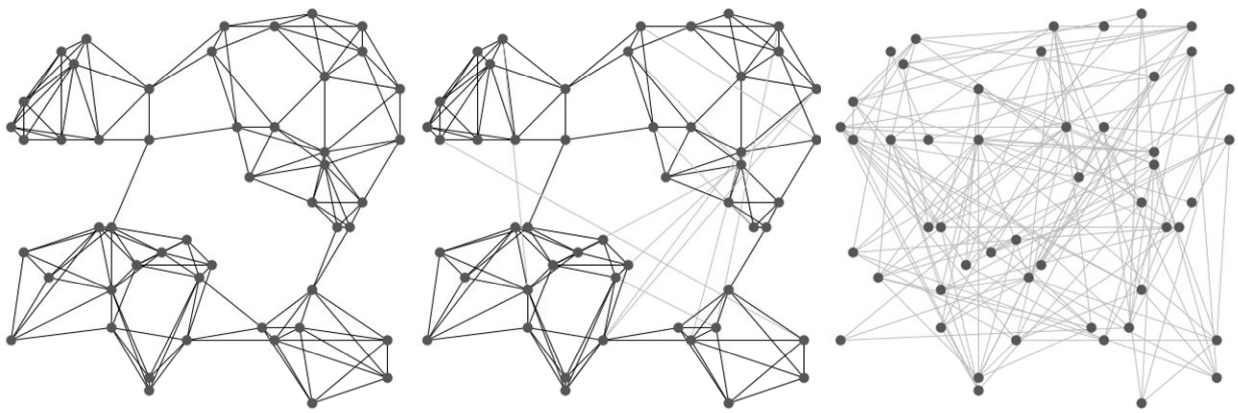
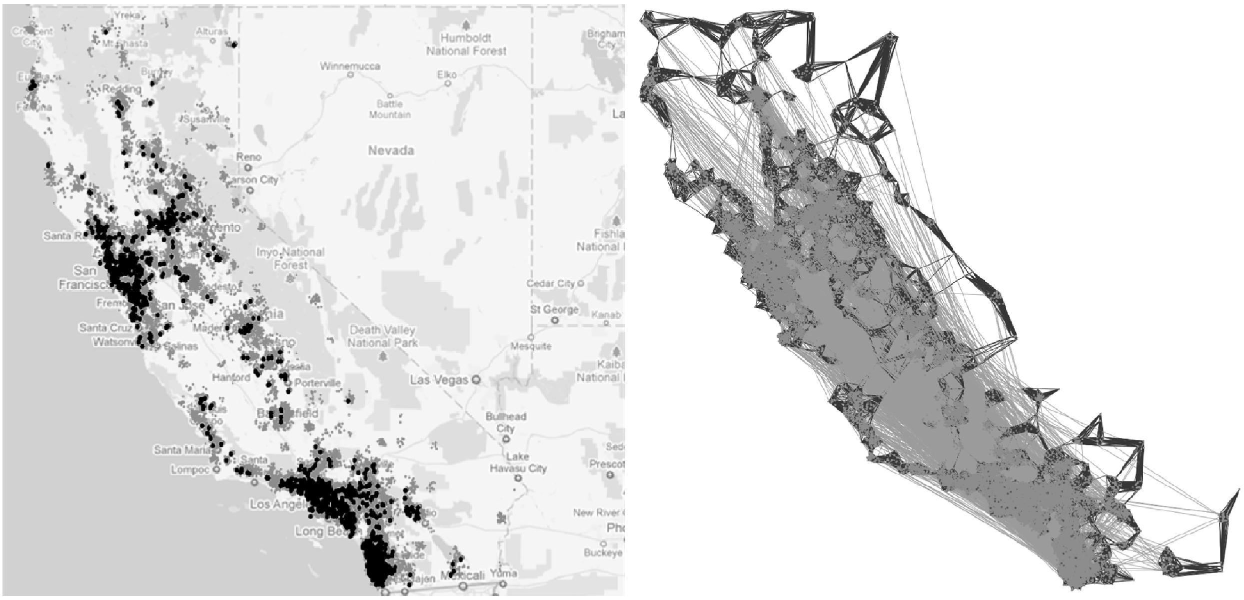


Figure 6.2: Social Network With Different Rewiring Probabilities: (a) Regular Nearest Neighbor (α=0.0), (b) Small World (α=0.1), (c) Random Network (α=1.0)

Following the Watts-Strogatz mechanism (Watts and Strogatz, 1998), a random graph generation model that produces graphs with small-world properties, including the short average path length and high clustering, a small world network[2] of California respondents from NHTS is adapted from the NetLogo Small-World Network model (Wilensky, 2005) using NetLogo (Wilensky, 1999). More details on this NetLogo model are included in Appendix D: Small-World Network Simulation Model in NetLogo. The geographic location, i.e. latitude and longitude data mapped from the zip code information, is used as the attributes of social dimension for consumer *i*. In Figure 6.3, hybrid electric vehicle owners are shown in black, while the conventional

[2] The original NetLogo model file can be found at:
https://docs.google.com/open?id=0Bz-d9_dIxQRAcWhpS2ZWcFVKQzQ

vehicle owners are shown in gray. In the right figure, the dark gray lines represent the original links to nearest neighbors, while the light gray lines represent rewired links to random consumers in network.



(a) (b)

Figure 6.3: California Sample Population from NHTS:

(a): Geographical Map; (b): Small-World Network With n = 10, α = 0.01

Depending on the predefined average number of friends n, each consumer is connected to its n nearest neighbors. All links are undirected, meaning all connections are mutual, $L_{ij} = L_{ji}$. Every existing link is then rewired to a random consumer with the rewiring probability α. Figure 6.4 shows the information flow in the social network simulation. The network parameters such as the number of friends and the rewiring probability belong to the inputs to the social network simulation, while social influence attributes and the network properties such as clustering coefficients and average path length are the outputs. The ***clustering coefficient*** is defined as the probability that two randomly selected friends of focal consumer are friends with each other. In other words, it is the fraction of pairs of friends that are connected to each other by links. In general, the clustering coefficient of a node ranges from 0 (when none of the consumer's friends are friends with each other) to 1 (when all of the consumer's friends are friends with each other). ***Path*** in a social network is simply defined as a sequence of nodes with the property that each consecutive pair in the sequence is connected by a link. Things often travel along a path – this could be a passenger taking a sequence of airline flights, or a trend of adopting a new technology being passed from person to person in a social network. The calculation of the average path length reflects the average distance between two nodes randomly selected in the network. In a social network with short average path length, product information is easily passed from consumer to consumer.

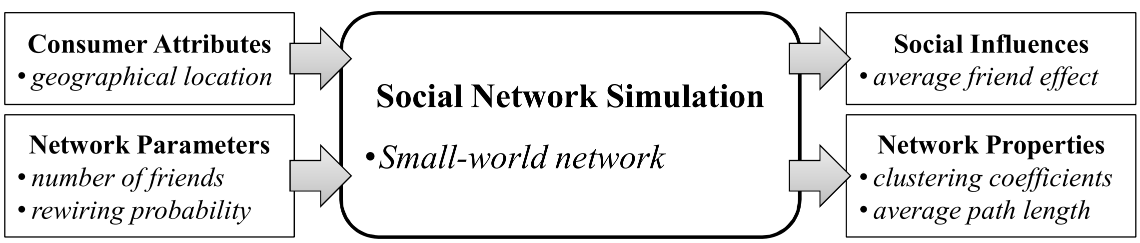


Figure 6.4: Influence Diagram in Social Network Simulation

Figure 6.5 shows the relation between network properties and the input parameter settings. Three n values, 5, 10, and 15 are plotted as curves with different marks and shades. As shown in Figure 6.5(a), the clustering coefficient increases with the decrease of rewiring probability and the increase of number of linked neighbors. On the other hand, the average path length decreases with the increase of rewiring probability and the number of linked neighbors, as seen in (b). In the following section, choice modeling results based on n = 10 and α=0.01 are presented, followed by the sensitivity analysis on the impact of varying number of linked neighbors and rewiring probabilities. The number of friends and rewiring probability are chosen based on the literature in (Kossinets and Watts, 2006) and (Watts and Strogatz, 1998), respectively.

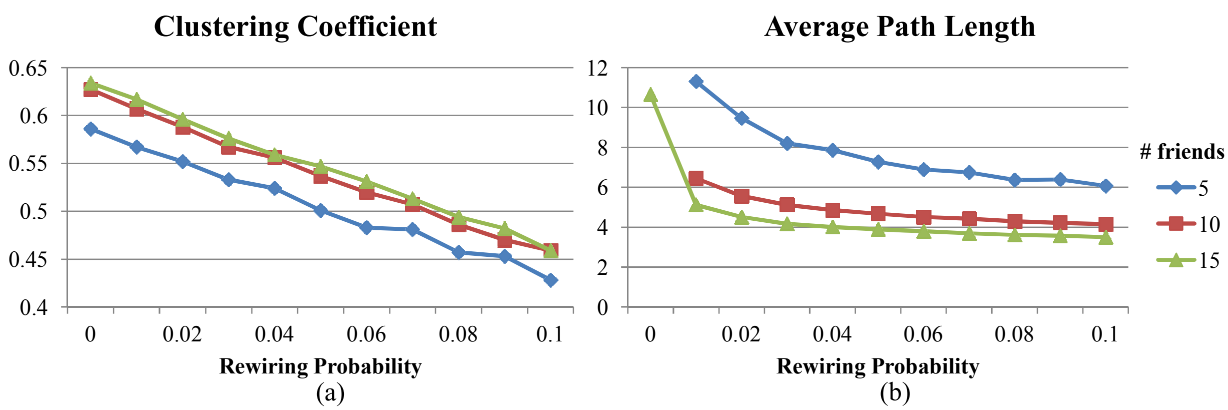


Figure 6.5: Social Network Properties In Response To Network Parameters: (a) Clustering Coefficient, (b) Average Path Length

In product design, this small-world phenomenon implies that consumers not only consider the choices of close friends, but are also influenced by remote contacts such as online reviews from people outside the regular social proximity. Literature has shown that many empirical networks exhibit the small-world phenomenon, for this reason, the Watts-Strogatz method for small-world network simulation is used in our case study to simulate the network structure, in lieu of the social network data. As for measuring the influence through social network, a variation of average friend effect described in Table 6. is used, as shown in Eqn. (6.7).

$$N_{i,t} = \sum_j L_{ij} y_{j,t\text{-}1} \Big/ \sum_j L_{ij} \tag{6.7}$$

where $y_{j,t-1}$ is a binary variable that represents the choice behavior at previous time period t-1: 1 for hybrid electric vehicle owner, and 0 for conventional vehicle owner. This social influence attribute can be regarded as the percentage of hybrid electric vehicle owners in the focal consumer's friends circle. The simulated value of social influence attribute in Eqn. (6.7) is later passed as an input to the choice model. The evaluation process iterates for each of the time periods $t=1...T$, to capture the dynamic nature of the social influence attribute.

6.4.2 Estimation of Discrete Choice Models

Data from 13,802 respondents in California who owned vehicles with model year from 2002 to 2009 are selected for building the discrete choice model. Note that 2002 is the first year the HEVs appeared in the NHTS data. While vehicles from multiple years are considered in choice modeling, the consumer profile and usage context attributes are based on the collected data in 2009. Therefore, the choice model estimation results shown below are valid under the assumption that there is no significant change in consumer profile and usage context from 2002 to 2009. Because the collected data provides no information about other vehicles considered by the respondents, three vehicles other than the one purchased are randomly selected from a set of 262 car models based on a uniform distribution to compose an individual choice set of four vehicles. McFadden (McFadden, 1978) has shown that a multinomial logit model estimate using choice sets composed of randomly selected members drawn with a uniform distribution from the set of all choice alternatives will result in consistent estimates of the model parameters. Later in the sensitivity analysis, the impact of the number of vehicles in a choice set is studied.

As shown in Figure 6.6, seven consumer-desired vehicle attributes **A** are selected including price, mileage per gallon (mpg), vehicle origin, vehicle size, vehicle type, footprint, acceleration (torque/vehicle weight), and the HEV indicator. The attribute "price" is the money respondents paid, excluding tax, license, trade-in value, etc. The mileage per gallon comes from the combined mpg published by the US Environmental Protection Agency (EPA, 2008). Vehicle origins are categorized as domestic, European, Japanese, and Korean; vehicle sizes are grouped into compact, midsize, large, and premium; vehicle type includes mini (such as compact vehicles), car (such as sedans), sport utility vehicles (SUV), minivan, van, multi-activity vehicles (MAV), and pickup. Vehicle footprint is defined as the product of vehicle length and vehicle width, reflecting the general size of the car, while the power, i.e. torque, divided by vehicle weight, is used as an approximate measure of the acceleration feature. The HEV indicator, is coded as 1 for hybrids, and 0 for conventional vehicles.

Gender, *age*, *household income*, *number of children under age 18*, *education level*, are included as consumer profile attributes **S**. Numerous combinations of consumer attributes are tested in the choice modeling process. In the final choice model estimation results, only three consumer profile attributes, *household income, number of children*, and *education level*, are statistically significant. As for the vehicle usage attributes **E**, the most commonly considered vehicle usage attributes *miles driven daily* is included in the choice model.

$$W_{ik,t} = W(\beta : \mathbf{A}_{ik}, \mathbf{S}_i, \mathbf{E}_i, \mathbf{N}_{i,t})$$

Terms	Consumer Desired Product Attributes	Consumer Profile	Usage Context	Social Influence
Examples	✓ Price ✓ Mileage per gallon ✓ Vehicle origin ✓ Vehicle size ✓ Vehicle type ✓ Footprint ✓ Acceleration ✓ HEV indicator	✓ Gender ✓ Age ✓ Household income ✓ Children under 18 ✓ Education level	✓ Miles driven daily	✓ Average friend effect on green attitude

Figure 6.6: Selected Attributes Included In Choice Utility Function

Table 6.2: Model Statistics and Coefficients of MNL with N and MNL without N (N=Social Influence)

	MNL with N		MNL without N	
Log likelihood at zero		-1.00E+07		-1.00E+07
Log likelihood at convergence		-9.97E+06		-9.97E+06
ρ^2		0.1459		0.1457
Attributes	Coeff.	Std.Err.	Coeff.	Std.Err.
Price(k)/income	-0.1598	0.0006	-0.1598	0.0006
100/miles per gallon	-0.8970	0.0013	-0.8960	0.0013
100/miles per gallon * miles driven daily	-0.0007	0.0000	-0.0007	0.0000
Vehicle origin(domestic as base)				
European	0.0740	0.0020	0.0743	0.0020
Japanese	0.8144	0.0011	0.8141	0.0011
Korean	-0.5864	0.0022	-0.5862	0.0022
Vehicle size (compact as base)				
Medium	-0.4954	0.0013	-0.4957	0.0013
Large	-0.5031	0.0025	-0.5026	0.0025
Premium	-0.6153	0.0017	-0.6156	0.0017
Vehicle type (car as base)				
MAV	-0.7805	0.0013	-0.7797	0.0013
Mini car	-0.4588	0.0031	-0.4600	0.0031
Minivan	-1.3492	0.0032	-1.3497	0.0032
Minivan * number of children	0.6068	0.0016	0.6073	0.0016
Pickup	-0.4422	0.0020	-0.4432	0.0020
SUV	-0.8448	0.0024	-0.8447	0.0024
SUV * number of children	-0.2108	0.0021	-0.2101	0.0021
Van	-1.7665	0.0065	-1.7677	0.0065
Footprint	0.0832	0.0006	0.0833	0.0006
Acceleration	18.7690	0.0600	18.7885	0.0600
Green attitude				
Hybrid electric vehicle (HEV)	-2.0804	0.0080	-2.1592	0.0079
Hybrid electric vehicle * high education level	0.6388	0.0034	0.6380	0.0034
Hybrid electric vehicle * fuel price	0.7560	0.0030	0.8022	0.0029
Hybrid electric vehicle * social impact	3.2379	0.0508	/	/

Structure of the utility function is shown in Eqn. (6.8), where interactions between **A**, **E**, **S**, and **N** are explicitly modeled. Alternative specific constants (ASC) for each of the car models are not included in the utility function. While this may decrease the goodness-of-fit of the model, it allows choice prediction of new vehicle in product design.

$$W = \beta_A \cdot \mathbf{A} + \beta_{A \cdot S} \mathbf{A} \cdot \mathbf{S} + \beta_{A \cdot E} \mathbf{A} \cdot \mathbf{E} + \beta_{A \cdot N} \mathbf{A} \cdot \mathbf{N} \qquad (6.8)$$

Following the choice modeling procedure in Stage III described in Section 6.3, interactions between consumer-desired product attributes **A**, consumer profile **S,** usage context attributes **E**, and the social influence **N** are explicitly modeled in the utility function. The coefficients for all attributes and their interactions based on a multinomial logit model estimation (MNL with **N**) are compared to the estimation results from a multinomial logit model without "social network influence" (MNL without **N**). "Green attitude", which reflects consumers' attitude toward new hybrid technology, is a collective effect of all utility terms involving the hybrid electric vehicle indicator, i.e. hybrid electric vehicle, interaction between hybrid electric vehicle and high education level, interaction between hybrid electric vehicle and fuel price, and interaction between hybrid electric vehicle and social impact, as shown in Table 6.2.

Goodness-of-fit measures based upon the log-likelihood of the converged model, such as the likelihood ratio index ρ^2 (also known as pseudo R-square), reflect how well the estimated model predicts actual individual choices in the data set. Higher values of ρ^2 indicate better predictions of the choices. As shown at the top of Table 6.2, a slightly higher log-likelihood of -9.97e+6 and subsequently ρ^2 value of 0.1459 are achieved using the MNL model with social influence attribute **N** versus the MNL model without **N**. The ρ^2 value of 0.1459 means that the MNL model with N have a 14.59% improvement in prediction, compared to the initial model with zero information. Even though the proposed choice modeling framework captures the systematic heterogeneity of consumers under social impact, the differences between two models shown are relatively small. A closer look into the data set revealed the underlying reason: the hybrid electric vehicle owners comprise a small percentage (7.59%) of the whole sample population, resulting into a limited number of observations with non-zero social influence attribute values (0.67%).

From the results of the MNL including **N** attributes in modeling, we note that all coefficients are statistically significant at the 0.01 level. The coefficients for *price/income* and *100/miles per gallon* are negative as expected. Only three consumer profile attributes, *household income*, *number of children*, and *education level*, are statistically significant. A negative estimator for *100/miles driven daily * miles driven daily* indicates that the usage context attribute has a negative impact on consumers' preference on the inverse of MPG measure, gasoline needed to travel 100 miles in this case. Similarly, the positive sign of *minivan * number of children* suggests that household with children prefer Minivan, which is consistent with common sense. As for the *green attitude*, it is shown that consumers have negative attitudes toward HEV indicator, excluding all interactions. This may be caused by uncertainty associated with the new technology and limited consumer knowledge. *Higher education level* positively impacts people's attitude toward HEV. Similar effect is seen with the average *fuel price* of a certain time period. As expected, the social impact has a large positive impact on the hybrid electric vehicle attitude.

6.4.3 Sensitivity Analysis

In the previous section, the social network is constructed based on the assumption that consumers have 10 linked friends on average, and the rewiring probability is equal to 0.01%. To better understand the model dependence upon these assumptions, the sensitivity of the agent-based choice modeling in response to the changes in average number of friends n and rewiring probability α is tested under three choice set scenarios: 1) choice set size = 4, 2) choice set size = 20, and 3) choice set size = 100. For each of the choice set scenarios, three discrete values, 5, 10, and 15, are selected for number of friends, while four values of the rewiring probability of 0%, 1%, 2%, and 10%, are tested. The results are summarized in Figure 6.7. The three figures (a), (b), and (c) represent results under different scenarios of choice set size. The vertical axis shows the value of social impact term N_i from the discrete choice model estimation, while the horizontal axis shows the value of rewiring probability. As shown in the legend, the light gray, medium gray, and dark gray lines correspond to average number of linked neighbors of 5, 10, and 15, respectively. For each of the parameter combinations, three random samples are drawn to minimize the individual bias from random sampling, as shown with different markers.

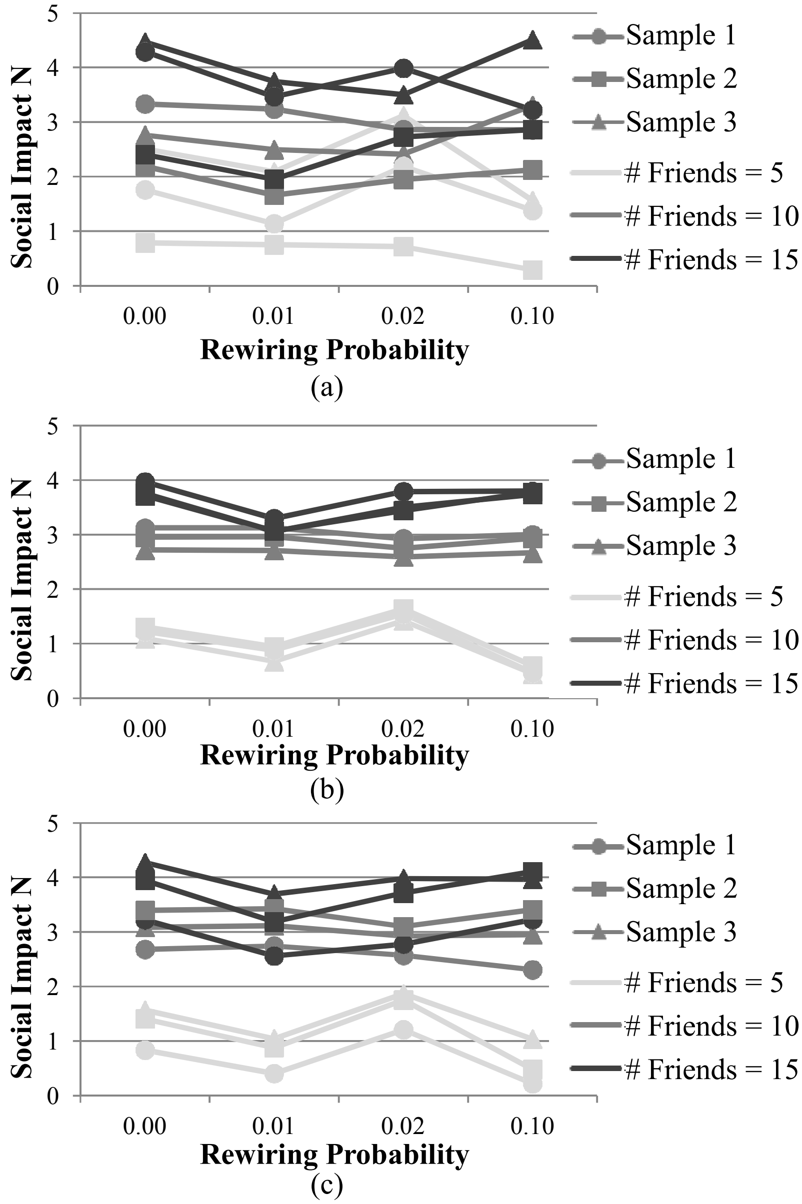
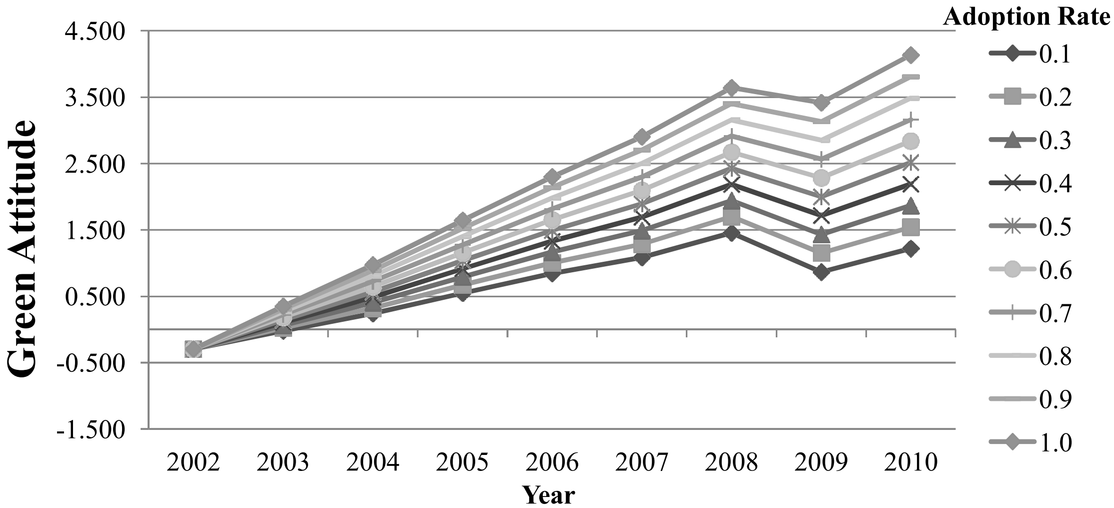


Figure 6.7: Social Impact In Response To # Friends & Rewiring Probability Under Different Scenarios of Choice Set Size: (a) = 4, (b) = 20, (c) = 100.

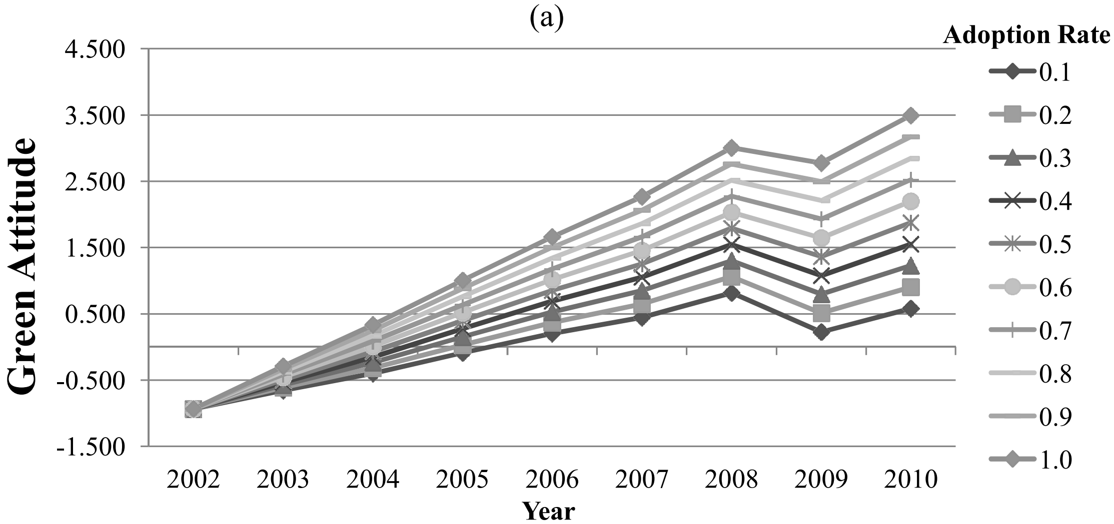
From Figure 6.7, we note that the increasing number of friends resulted into stronger social impact terms. This is because the social impact term in Eqn. (6.7) follows a percentage representation: the more contacts a consumer is linked to, the smaller the social impact term, with all other things being equal. This difference becomes clearer with the increasing choice set size, because the randomness decreases with the increase of number of alternatives in individual choice sets. No clear distinction in social impact value is seen between (a), (b) and (c), suggesting that the choice set size may not have a significant impact upon parameter estimator of social impact. However, a closer look into the log likelihood at convergence and under scenarios with different choice set size shows that it matters in terms of the model fit, with ρ^2 averages to 0.1463, 0.0929, and 0.0687 in (a), (b), and (c), respectively.

6.4.4 Green Attitude Forecasting

To illustrate the potential in forecasting new product adoption using the proposed approach, the green attitude from 2002 to 2010 under different scenarios are plotted in Figure 6.8. As mentioned earlier, green attitude measures a collective effect of all utility terms involving the hybrid electric vehicle indicator in choice model. The vertical axis in both plots represents the green attitude, which is the sum of the HEV related coefficients in the utility terms. The plot (a) represents consumers with higher education level (college graduates or higher), while the plot (b) illustrates the remaining population. A comparison of the two plots shows that consumer with higher education level tend to have more positive preference for HEVs. In both plots, ten scenarios with different adoption rate at year 2010 are tested, ranging from 0.1 (10%) to 1.0 (100%). As seen in the figure, the green attitude increases faster under higher adoption rate. In general, it follows an increasing trend except for the drop between 2008 and 2009 due to the decrease in gasoline price.



(a)



(b)

Figure 6.8: Attitude Towards Green HEV Technology Over Time: (a) Higher Education; (b) Non-higher Education

To illustrate the potential in forecasting new product adoption using the proposed choice model, the green attitude from 2002 to 2020 under different scenarios are plotted in Figure 6.9. As mentioned earlier, green attitude measures a collective effect of all utility terms involving the hybrid electric vehicle (HEV) indicator in the choice model. The vertical axis in the plot, the ***green attitude*** is the sum of the HEV related coefficients in the utility terms (Table 6.2). Take the

MNL model without social influence attributes **N** as an example, it is the summation of -2.1592 × *HEV* + 0.6380 × *HEV* × *high education level* + 0.8022 × *HEV* × *fuel price.* And in the case of MNL model with social influence attributes **N**, it equals to -2.0804 × *HEV* + 0.6388 × *HEV* × *high education level* + 0.7560 × *HEV* × *fuel price* + 3.2379 × *HEV* × *social influence.*

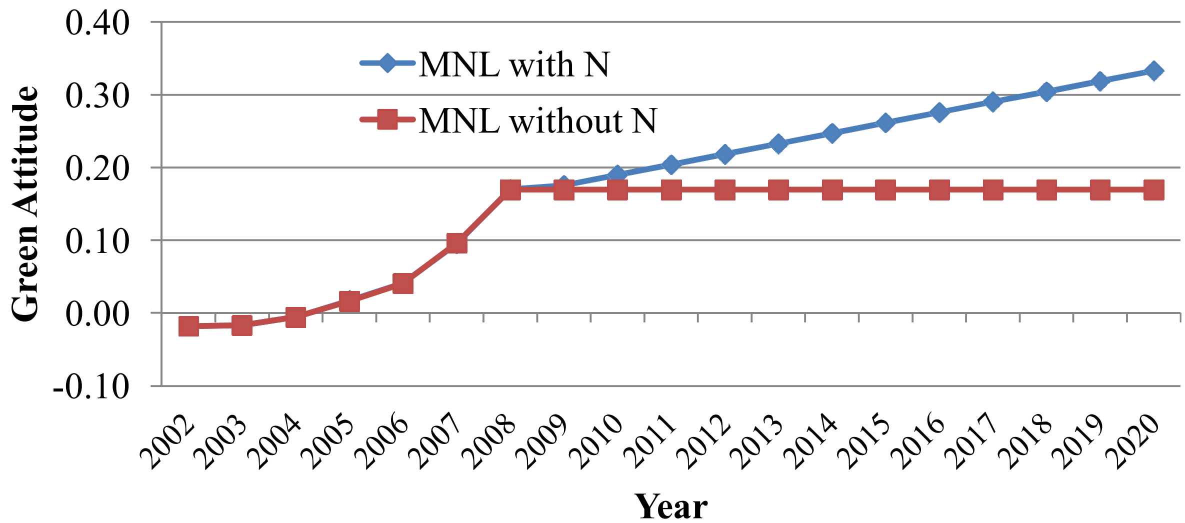


Figure 6.9: Forecast of Attitude Towards Green HEV Technology Over Time

The value of green attitude from 2002 to 2008 comes from the NHTS data and the remaining values since 2009 are forecasted and compared between using MNL with N and MNL without N. As shown in Figure 6.9, green attitudes from MNL with N and MNL without N are consistent from year 2002 to 2008, while the green attitude forecasting diverges significantly starting from 2010 because the MNL model without N doesn't take the social influence into account. In forecasting the green attitude after 2009, we assume that the value of social influence attribute increases linearly with respect to time – the assumption is supported by a linear regression model built using data from 2002 to 2008. With the increasing value of social influence attribute, the green attitude is forecasted to grow significantly in the next decade, while the forecast from

MNL without N stays constant due to the exclusion of dynamic social influence attributes in the utility function. Similarly, the predicted choice share of HEV using MNL with N and MNL without N are shown in Figure 6.10.

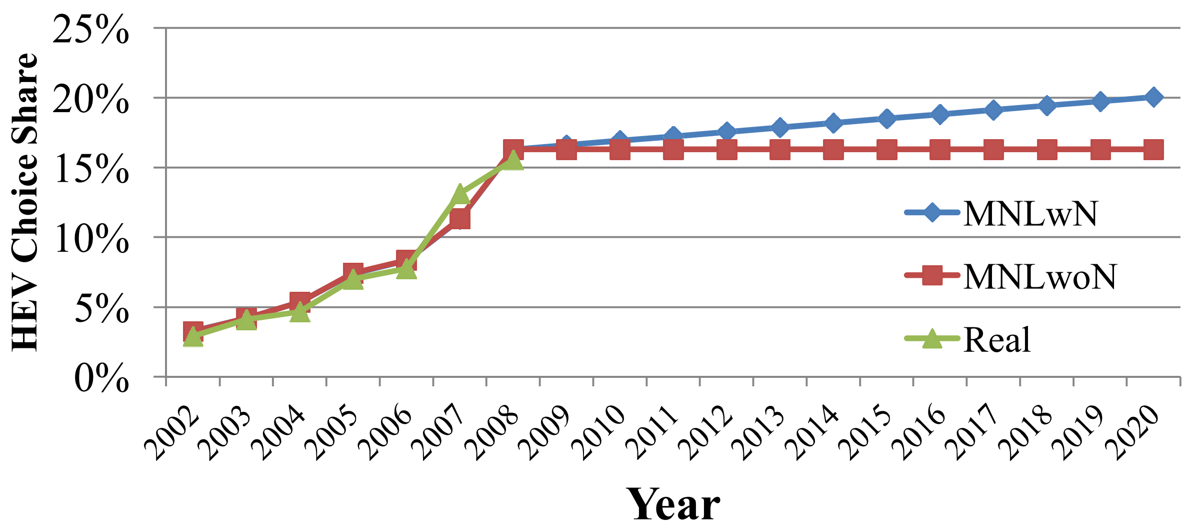


Figure 6.10: Predicted HEV Choice Share Over Time

6.5 Conclusions and Discussions

In this work, an integrated agent-based choice modeling framework considering social impact is proposed for forecasting new product adoption. To our best knowledge, this is the first attempt in engineering design research to develop analytical techniques that integrate discrete choice model with social network simulation at individual level to address simultaneously the interactions between product, consumer, and social network in product design. The primary research contribution of this work is the development of an integrated choice modeling approach that combines discrete choice models with social network simulations to support new product design, while considering the social impact upon consumers' choices at the individual level. Modeling social context within choice modeling framework introduces a new dimension for understanding consumers' attitudes toward new products, and at the same time, improves prediction accuracy by reducing market risks and uncertainties associated with consumer behavior studies. This integration offers a dynamic view of consumers' choice of new products, in which consumers' preferences may change over time due to social influence. To ensure that our proposed methodology reflects real market behavior, external data source, i.e. the National Household Travel Survey data, is used for choice model estimation to provide insights into how consumers make tradeoffs among different attributes.

The case study of hybrid electric vehicle owners in California illustrates the potential benefits of the proposed methods in supporting the design of green products. Alternative fuel vehicles have received wide attention lately due to the increasing awareness of environmental impacts among consumers and the incentives from government. The dramatic changes in automotive industry call for an innovation in technology and a switch of focus to alternative-fuel vehicles, among which HEV/PHEV are the most expected due to the social impact upon consumer choice of new product. As the first generation PHEV models launches into U.S. market, forecasting its market potential would be of great interests for the consumers, manufacturers, generation companies, and government agents. From a broader system point of view, such choice models can be further integrated into a multi-agent energy market simulation framework to study the impact of consumer vehicle choices on future electric generation needs. The agent-based choice modeling considering social impact could be utilized to address the above needs, providing an estimate of choice share with respect to consumers' dynamic preferences from the past to the future.

The case study of hybrid electric vehicle owners in California illustrates the potential benefits of the proposed methods in supporting the design of green products. Alternative fuel vehicles

have received wide attention lately due to the increasing awareness of environmental impacts among consumers and the incentives from government. The dramatic changes in automotive industry call for an innovation in technology and a switch of focus to alternative-fuel vehicles, among which HEV/PHEV are the most expected due to the social impact upon consumer choice of new product. As the first generation PHEV models launches into U.S. market, forecasting its market potential would be of great interests for the consumers, manufacturers, generation companies, and government agents. From a broader system point of view, such choice models can be further integrated into a multi-agent energy market simulation framework to study the impact of consumer vehicle choices on future electric generation needs. The agent-based choice modeling considering social impact could be utilized to address the above needs, providing an estimate of choice share with respect to consumers' dynamic preferences from the past to the future.

Chapter 7

CONCLUSIONS AND INTELLECTUAL MERITS

7.1 CONTRIBUTION OF THE DISSERTATION

The primary research contribution of this work is the development of an integrated Usage and Social Context-based Choice Modeling approach to support the engineering design for new product design, considering the heterogeneity of the consumers in their product usage, social network, and preferences. This dissertation represents the first attempt in engineering design research to develop analytical techniques that integrate engineering, marketing, and social science domains to address simultaneously the interactions between product, consumer, and usage/social context in product design. The proposed approach transforms the conventional product-centric paradigm to human-centered design by considering factors beyond the traditional engineering domain including consumers' perception, usage context, and social influence. Incorporating usage context into choice modeling greatly benefits studies of product family design and market segmentation, as market segments include not only heterogeneous consumers preferences, but also distinctive usage patterns as well. Moreover, modeling social context within an agent-based simulation framework introduces a new dimension for understanding consumers' attitudes toward new products, and at the same time, improves prediction accuracy by reducing market risks and uncertainties associated with consumer behavior studies.

The specific tools and methodologies are built upon the principles of the Decision-Based Design (DBD) paradigm, providing a tool to incorporate rating data into the DBD methods, a hierarchical framework to capture the impact of usage context on product performance and consumer preferences, and an agent-based choice modeling approach considering social impact to support engineering design for new product adoption. The proposed approach provides a rigorous choice modeling approach, which is suitable for use on various engineering products in a wide variety of market, and an estimate of choice share with respect to consumers' dynamic preferences from the past to the future

The specific research contribution of each of the new tools and methodologies comprising the usage and social context-based choice modeling approach is detailed as follows.

The **Integrated Mixed Logit Modeling** (IMLM) procedure is developed to offer a mathematically rigorous, decision-theoretic modeling tool for incorporating consumers' perception such as rating data into the choice modeling process. Such a method is needed based on an examination of key characteristics of rating data and the resulting challenges in modeling, which could lead to invalid estimation results. The IMLM procedure extends the DCA choice utility function to quantitatively establish the relationship between subjective measures of consumer perception such as rating data and consumer profile and product attributes while employing the DBD principles to provide vigorous quantitative assessments for design decisions. The IMLM procedure can be implemented for a real design problem, with an interdisciplinary team composed of marketing researchers and engineering experts.

The **Usage Context-based Choice Modeling** framework can be widely applied to assess the impact of usage context for any engineering product in which an interaction between a user (or a group of users) and a product (or product family) exists. In this work, the attributes of the usage context, such as driving condition in the case of vehicle, were believed to influence both the product performance and the consumer preferences. The proposed usage context-based choice modeling framework includes four phases: usage context identification, data collection, linking performance with usage context and customer profile, and choice model estimation. The hierarchical choice modeling method is necessary to complement the development of mass customization in product and product family design, in which the product design must meet the needs of a diverse consumer population with varying usage patterns.

The methods for **Statistical Analysis of Hybrid Electric Vehicle Adoption** can be applied to a wide variety of market data and the respective predictive models estimated using market data in the context of new product adoption. While the results presented focus on the hybrid electric vehicles, these methods can be applied to a wide range of products. These methods are developed specifically for identifying the key characteristics of early adopters of new products or new technologies. Multiple data sets are combined for modeling purposes to address the issue of limited data availability. Application of the methods will result in better predictive models for forecasting the impact of new designs or design improvements on consumer choices, as illustrated in the vehicles design selection study, and ultimately enterprise profitability.

The proposed **Agent-based Choice Modeling** approach provides the necessary comprehensive choice modeling methodology to guide the design of new products with considerations of interpersonal interactions among consumers. This approach explicitly captures the impact of social network upon consumers' choice behavior, and is formulated to address the challenges in modeling consumer heterogeneity in their decision-making process, such as limited data availability, and the need to quantify impact of social network. This methodology could find wide spread use in forecasting market penetration trend for new products or technologies, such as alternative fuel vehicles, and other eco-friendly products and devices, in which social interactions play a critical role in consumers' adoption of the new product and hence the successful market penetration. The approach is innovative in the sense that it captures the dynamic nature of the new product adoption, which is often neglected in the traditional choice modeling literature with the "rationale" assumption.

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 potential high profits. These types of breakthroughs are rare, however (Pine, 1993, Otto and Wood, 2001); it has been noted a few years ago 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 design to ensure consumer acceptance and profitability. Without solid methods for decision-making to manage innovation, enterprises must overly reply upon benchmarking successful competitors, creating superficial cosmetic changes to existing products to generate interest, or introducing a wide variety of disparate product 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 product 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 Usage and Social Context-based Choice Model provides a method to build consumer preference models under specific usage context and social impact at any time throughout the product design life cycle, and provide rigorous evaluation of design improvements.

7.2 Recommendations for Future Work

IMLM Method: The recommended future work for the IMLM method is to improve the currently sequential estimation of the rating model and the choice model into a single stage all-in-one process. The integrated all-in-one model estimation is needed to alleviate the issue of error propagation from the rating model to the choice model. Moreover, different formulations of rating model can be explored and compared to identify the suitable problem formulation in the context of choice modeling. Meanwhile, the integration of proposed IMLM method into the design optimization problem warrants further research efforts to demonstrate its capability in supporting engineering design.

Usage Context-based Choice Modeling: The primary research need for the usage context-based choice modeling is an extension to choice consideration set selection with respect to usage context. The potential influence of usage context upon the choice consideration set construction in a two-stage decision making process (Hauser et al., 2010) would be an interesting research topic to explore. Meanwhile, this work is limited to single primary usage of a product, which may not be true for many market offerings. Therefore, the expansion of the current framework to model multiple-usage contexts for product family design is another interesting directions for future work. Furthermore, the proposed framework creates performance models and preference models in two separate steps (Phase III and Phase IV, respectively), in which error from lower level model may cause issues in upper level model estimation. Hence, the all-in-one Hierarchical Bayesian choice modelling approach (Hoyle et al., 2011) can be used to improve the stability of model estimation in the proposed framework.

Statistical Analysis of Hybrid Electric Vehicle Adopters: Methods have been developed for preprocessing the market data for understanding the relationship between consumer profile and usage context attributes, identifying the key characteristics of hybrid electric vehicle adopters, and modeling consumers' choices by combining multiple datasets. However, the findings presented in this dissertation are limited to hybrid electric vehicles adoption and the vehicle design selection study is built upon the assumption that consumers' attitudes towards hybrid electric vehicle technology and plug-in hybrid electric vehicle technology is comparable. As the first generation PHEV models launches into U.S. market, we expect to see more PHEV market data. With these new data, either stated preference data or revealed preference data, research work is needed to understand consumers' attitude of PHEV and how the adoption of PHEV differs from the current HEV trend observed and discussed in this dissertation. Further, forecasting its market potential would be of great interests for the consumers, manufacturers, generation companies, and government agents, because of its promising potential in realizing the zero emission target.

Agent-based Choice Model considering Social Impact: Several areas for future work remain for the agent-based choice modeling approach. Mostly importantly, empirical study needs to be conducted to confirm the hypotheses used in social network simulation. The proposed agent-based choice modeling considering social impact could be utilized to provide an estimate of choice share with respect to consumers' dynamic preferences from the past to the future. However, due to its independence of irrelevant alternatives assumption, the multinomial logit model presented in this work has limitations, which warrant future research investigation by implementing other choice modeling techniques such as Hierarchical Bayes mixed logit model. Moreover, there are potential confounding issues because both social impact term and choice utility is a function of consumer profile attributes. To extend and refine the current work, the two-stage Berry, Levinsoh, and Pakes method proposed by (Walker et al., 2011) may be used to correct the endogeneity in a choice model with confounding attributes. The integration of choice modeling and agent-based simulation offers a dynamic view of consumers' choice of engineering products, in which consumers' preferences may change over time. To ensure that our proposed methodology reflects real market behavior, external data sources including aggregate and disaggregate market data, as well as existing benchmarking product diffusion models such as Bass model (Bass, 1969) from the literature can used for model calibration and validation to reduce the modeling uncertainties.

References

AJZEN, I. 2005. *Attitudes, personality and behaviour*, Open Univ Pr.

AKERLOF, G. A. 1997. Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, 1005-1027.

ALLENBY, G. M., ROSSI, P. E. & MCCULLOCH, R. E. 2005. Hierarchical Bayes Models: A Practitioners Guide. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=655541.

AOKI, M. 1995. Economic Fluctuations With Interactive Agents: Dynamic And Stochastic Externalities*. *Japanese Economic Review,* 46**,** 148-165.

AXELROD, R. 1997. *The complexity of cooperation: Agent-based models of competition and collaboration*, Princeton Univ Pr.

AXSEN, J., BURKE, A. & KURANI, K. 2008. Batteries for Plug-in hybrid electric Vehicles (PHEVs): Goals and the state of Technology circa 2008. Davis, CA: Institute of Transportation Studies, University of California Davis.

AXSEN, J. & KURANI, K. 2008. The early US market for PHEVs: Anticipating consumer awareness, recharge potential, design priorities and energy impacts. Davis, CA: Institute of Transportation Studies, University of California Davis.

AXSEN, J. & KURANI, K. 2009. Interpersonal Influence within Car Buyers' Social Networks: Five Perspectives on Plug-in Hybrid Electric Vehicle Demonstration Participants. Davis, CA: Institute of Transportation Studies, University of California Davis.

BARABÁSI, A. L. & ALBERT, R. 1999. Emergence of scaling in random networks. *science,* 286**,** 509.

BASS, F. M. 1969. A new product growth for model consumer durables. *Management Science,* 15**,** 215-227.

BELK, R. 1974. An exploratory assessment of situational effects in buyer behavior. *Journal of Marketing Research***,** 156-163.

BELK, R. 1975. Situational variables and consumer behavior. *Journal of Consumer Research***,** 157-164.

BEN-AKIVA, M. & BOCCARA, B. 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing,* 12**,** 9-24.

BEN-AKIVA, M. & LERMAN, S. R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand,* Cambridge, MA, MIT Press.

BERKOWITZ, E., GINTER, J. & TALARZYK, W. Year. An investigation of the Effects of Specific Usage Situations on the Prediction of Consumer Choice Behavior. *In:* GREENBERG, B. & BELLENGER, D., eds. Educators' Proceedings, 1977 Chicago. American Marketing Association.

BLUME, L., DURLAUF, S. & INSTITUTE, U. O. W.-.-M. S. S. R. 2003. Equilibrium concepts for social interaction models. *International Game Theory Review,* 5**,** 193-210.

BOLLOBÁS, B. & CHUNG, F. R. K. 1988. The diameter of a cycle plus a random matching. *SIAM Journal on discrete mathematics,* 1**,** 328.

BOUND, J., BROWN, C. & MATHIOWETZ, N. 2001. Measurement error in survey data. *Handbook of econometrics***,** 3705-3843.

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

BROCK, W. A. & DURLAUF, S. N. 2001. Discrete choice with social interactions. *Review of Economic studies,* 68**,** 235-260.

BROCK, W. A. & DURLAUF, S. N. 2002. A multinomial-choice model of neighborhood effects. *The American Economic Review,* 92**,** 298-303.

BURKHARDT, M. 1994. Social interaction effects following a technological change: A longitudinal investigation. *Academy of Management Journal,* 37**,** 869-898.

BURT, R. 1987. Social contagion and innovation: Cohesion versus structural equivalence. *ajs,* 92**,** 1287.

CARLEY, K. 1991. A Theory of Group Stability. *American Sociological Review,* 56**,** 331-354.

CASE, A. 1992. Neighborhood influence and technological change. *Regional Science and Urban Economics,* 22**,** 491-508.

CHEUNG, K., KWOK, J., LAW, M. & TSUI, K. 2003. Mining customer product ratings for personalized marketing. *Decision Support Systems,* 35**,** 231-243.

CHRISTENSEN, C., COOK, S. & HALL, T. 2005. Marketing malpractice: the cause and the cure. *Harvard business review,* 83**,** 74.

CLAUSING, D. & HAUSER, J. 1988. The House of Quality. *Harvard Business Review,* 66**,** 63-73.

CONTRACTOR, N. & EISENBERG, E. 1990. Communication Networks and New Media in Organizations. *Organizations and communication technology***,** 143-172.

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

CORPORATION, S. 1996-2009. StataSE 9.2. College Station, Texas.

CROCKETT, R. O. 2007. Honing the Razr Edge. *Business Week.*

DE LA FUENTE, J. & GUILLÉN, M. 2005. Identifying the influence of product design and usage situation on consumer choice. *International Journal of Market Research,* 47**,** 667.

DEFFUANT, G., HUET, S. & AMBLARD, F. 2005. An Individual Based Model of Innovation Diffusion Mixing Social Value and Individual Benefit1. *ajs,* 110**,** 1041-69.

DELRE, S., JAGER, W., BIJMOLT, T. & JANSSEN, M. 2007. Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of Business Research,* 60**,** 826-835.

DIAMOND, D. 2009. The impact of government incentives for hybrid-electric vehicles: Evidence from US states. *Energy Policy,* 37**,** 972-983.

DICKSON, P. 1982. Person-situation: Segmentation's missing link. *The Journal of Marketing,* 46**,** 56-64.

DONG, H. & WOOD, W. Year. Integrating Computational Synthesis and Decision-Based Conceptual Design. *In:* Proceedings of the 2004 ASME Design Engineering Technical Conference, September 2004 Salt Lake City, Utah. ASME, New York.

DUGUNDJI, E. & GULYAS, L. Year. Empirical estimation and multi-agent based simulation of a discrete choice model with network interaction effects. *In*, 2003a.

DUGUNDJI, E. & GULYAS, L. Year. An exploration of the role of global versus local and social versus spatial networks in transportation mode choice behavior in the Netherlands. *In*, 2003b.

DUGUNDJI, E. R. & WALKER, J. L. 2005. Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record: Journal of the Transportation Research Board,* 1921**,** 70-78.

EASLEY, D. & KLEINBERG, J. 2010. *Networks, crowds, and markets: Reasoning about a highly connected world*, Cambridge Univ Pr.

EHSANI, M., GAO, Y. & EMADI, A. 2009. *Modern electric, hybrid electric, and fuel cell vehicles: fundamentals, theory, and design*, CRC.

ENGEL, J., KOLLAT, D. & BLACKWELL, R. 1969. Personality measures and market segmentation: Evidence favors interaction view. *Business Horizons,* 12**,** 61-70.

EPA, U. E. P. A. 2008. *Fuel Economy Guide 2008* [Online]. [Accessed].

EPSTEIN, J. & AXTELL, R. 1996. *Growing artificial societies: social science from the bottom up*, The MIT Press.

ERICKSON, B. (ed.) 1988. *The Relational Basis of Attitudes,* Cambrige, UK: Cambrige University Press.

FACEBOOK. *Statistics* [Online]. http://www.facebook.com/press/info.php?statistics, accessed on Jan. 24, 2012. [Accessed].

FARRELL, W. 2000. *How hits happen: Forecasting predictability in a chaotic marketplace*, Harper Business.

FAUST, K. & WASSERMAN, S. 1994. Social network analysis: Methods and applications. Cambridge University Press New York.

FESTINGER, L. 1954. *A theory of social comparison processes*, Bobbs-Merrill.

FHWA 2009. National Household Travel Survey (NHTS). *version 2.0 datasets.* November 2010 ed.: Federal Highway Administration

FRIEDKIN, N. E. 2006. *A Structural Theory of Social Influence*, Cambridge University Press.

FRISCHKNECHT, B. D., WHITEFOOT, K. & PAPALAMBROS, P. Y. 2010. On the Suitability of Econometric Demand Models in Design for Market Systems. *Journal of Mechanical Design,* 132**,** 121007.

FULK, J. 1993. Social construction of communication technology. *Academy of Management Journal,* 36**,** 921-950.

FULK, J., SCHMITZ, J. & RYU, D. 1995. Cognitive elements in the social construction of communication technology. *Management Communication Quarterly,* 8**,** 259-288.

GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. 2004. *Bayesian Data Analysis,* Boca Raton, FL, Chapman & Hall/CRC.

GOLDBERG, P. K. 1995. Product differentiation and oligopoly in international markets: The case of the US automobile industry. *Econometrica: Journal of the Econometric Society***,** 891-951.

GREEN, M. G., LINSEY, J. S., SEEPERSAD, C. C., WOOD, K. L. & JENSEN, D. J. Year. Frontier Design: A Product Usage Context Method. *In:* Proceedings of the 2006 ASME Design Engineering Technical Conferences and Computers in Engineering Conference, 2006 Philadelphia, PA. DETC/DTM 2006-99608.

GREEN, M. G., PALANI, R. P. K. & WOOD, K. L. Year. Product Usage Context: Improving Customer Needs Gathering and Design Target Setting. *In:* Proceedings of the 2004 ASME Design Engineering Technical Conferences and Computers in Engineering Conference, 2004 Salt Lake City, UT. DETC/DTM2004-57498.

GREEN, M. G., TAN, J., LINSEY, J. S., SEEPERSAD, C. C. & WOOD, K. L. Year. Effects of Product Usage Context on Consumer Product Preferences. *In:* Proceedings of the 2005 ASME Design Engineering Technical Conferences and Computers in Engineering Conference, 2005 Long Beach, CA. DETC2005-85438.

GREEN, P. E. & KRIEGER, A. M. 1995. Alternative Approaches to Cluster-based Market Segmentation. *Journal of the Market Research Society,* 37**,** 221-39.

GREENE, W. 2003. *Econometric analysis*, Prentice Hall Upper Saddle River, NJ.

GU, X., RENAUD, J. E., ASHE, L. M., BATILL, S. M., BUDHIRAJA, A. S. & KRAJEWSKI, L. J. 2002. Decision-Based Collaborative Optimization. *Journal of Mechanical Design,* 124**,** 1-13.

HARTIGAN, J. A. & WONG, M. A. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics),* 28**,** 100-108.

HAUSER, J. R., TOUBIA, O., EVGENIOU, T., BEFURT, R. & DZYABURA, D. 2010. Disjunctions of Conjunctions, Cognitive Simplicity and Consideration Sets. *Journal of Marketing Research,* 47**,** 485-496.

HAUSMAN, J. & MCFADDEN, D. 1984. Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society***,** 1219-1240.

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

HE, L. & CHEN, W. 2011. Usage Context-based Choice Modeling for Hybrid Electric Vehicles. *International Conference on Engineering Design.* Copenhagen, Denmark.

HE, L., CHEN, W. & CONZELMANN, G. 2011a. On Usage Context of Hybrid Electric Vehicle in Choice Studies. *Proceedings of the 2011 ASME Design Engineering Technical Conferences and Computers in Engineering Conference.* Washington, DC.

HE, L., HOYLE, C. & CHEN, W. 2010. A Framework For Choice Modeling In Usage Context-Based Design. *Proceedings of the 2010 ASME Design Engineering Technical Conferences and Computers in Engineering Conference.* Montreal, Canada.

HE, L., HOYLE, C. & CHEN, W. 2011b. Examination of customer satisfaction surveys in choice modelling to support engineering design. *Journal of Engineering Design,* 22**,** 669-687.

HEDEKER, D. & GIBBONS, R. D. 1994. A Random-Effects Ordinal Regression Model for Multilevel Analysis. *Biometrics,* 50**,** 933-944.

HERRMANN, J. W. & SCHMIDT, L. C. Year. Viewing Product Development as a Decision Production System. *In:* Proceedings of the 2002 ASME Design Engineering Technical Conference, September 2002 Montreal, Canada. ASME, New York.

HOYLE, C. & CHEN, W. 2009. Product Attribute Function Deployment (PAFD) for Decision-Based Conceptual Design. *IEEE Transactions on Engineering Management,* 56**,** 271-284.

HOYLE, C., CHEN, W., ANKENMAN, B. & WANG, N. 2009. Optimal Experimental Design of Human Appraisals for Modeling Consumer Preferences in Engineering Design. *Journal of Mechanical Design,* 131**,** 98-107.

HOYLE, C., CHEN, W., WANG, N. & GOMEZ-LEVI, G. 2011. Understanding and modelling heterogeneity of human preferences for engineering design. *Journal of Engineering Design,* 22**,** 583-601.

HOYLE, C., CHEN, W., WANG, N. X. & KOPPELMAN, F. S. 2010. Integrated Bayesian Hierarchical Choice Modeling to Capture Heterogeneous Consumer Preferences in Engineering Design. *Journal of Mechanical Design,* 132.

JOHNSON, R. A. & WICHERN, D. W. 2002. *Applied Multivariate Statistical Analysis,* Upper Saddle River, NJ, Prentice Hall.

KANDEL, D. B. 1978. Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology,* 84**,** 427-436.

KANTER, R. M. 2006. Innovation: The Classic Traps. *Harvard Business Review,* 84**,** 72-83.

KOPPELMAN, F. & BHAT, C. 2006. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. U.S. Department of Transportation.

KOPPELMAN, F. S. & SETHI, V. 2000. Closed Form Discrete Choice Models. *Handbook of Transport Modelling,* 1**,** 211–222.

KOSSINETS, G. & WATTS, D. J. 2006. Empirical Analysis of an Evolving Social Network. *Science,* 311**,** 88-90.

KRISHNAN, V. & ULRICH, K. T. 2001. Product Development Decisions: A Review of the Literature. *Management Science,* 47**,** 1-21.

KROES, E. P. & SHELDON, R. J. 1988. Stated preference methods: an introduction. *Journal of Transport Economics and Policy,* 22**,** 11-25.

KRUGMAN, P. 1990. Increasing returns and economic geography. National Bureau of Economic Research.

KUMAR, D., CHEN, W. & SIMPSON, T. 2009a. A Market-Driven Approach to Product Family Design. *International Journal of Production Research,* 47**,** 71-104.

KUMAR, D., HOYLE, C., CHEN, W., WANG, N., GOMEZ-LEVI, G. & KOPPELMAN, F. 2009b. A Hierarchical Choice Modelling Approach for Incorporating Customer Preferences in Vehicle Package Design. *International Journal of Product Development,* 8**,** 228-251.

KÜPPER, A. 2005. *Location-based services*, Wiley.

LAVIDGE, R. 1966. The cotton candy concept: Intra-individual variability. *Lee Adler & Irving Crespi. Attitude Research at Sea***,** 39-50.

LEENDERS, R. T. A. J. 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social networks,* 24**,** 21-47.

LI, H. & AZARM, S. 2000. Product Design Selection under Uncertainty and with Competitive Advantage. *Journal of Mechanical Design,* 122**,** 411-418.

LOUVIERE, J. J., HENSHER, D. A. & SWAIT, J. D. 2000. *Stated Choice Methods: Analysis and Application,* New York, Cambridge University Press.

LU, M. 1999. Determinants of Residential Satisfaction: Ordered Logit vs. Regression Models. *Growth and Change,* 30**,** 264-287.

LUCE, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis*, Wiley.

LUNN, D., THOMAS, A., BEST, N. & SPIEGELHALTER, D. 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing,* 10**,** 325-337.

MACDONALD, E. F., GONZALEZ, R. & PAPALAMBROS, P. Y. 2009. Preference inconsistency in multidisciplinary design decision making. *Journal of Mechanical Design,* 131**,** 031009.

MAGUIRE, M. 2001. Context of use within usability activities. *International Journal of Human-Computer Studies,* 55**,** 453-483.

MANSKI, C. F. 1995. *Identification problems in the social sciences*, Harvard Univ Pr.

MARSTON, M., ALLEN, J. K. & MISTREE, F. 2000. The Decision Support Problem Technique: Integrating Descriptive and Normative Approaches in Decision Based Design. *Engineering Valuation and Cost Analysis,* 3**,** 107-129.

MASON, R. L., GUNST, R. F. & HESS, J. L. 2003. *Statistical Design and Analysis of Experiments: With Applications to Engineering and Science,* New York, Wiley Series in Probability and Statistics.

MCCULLAGH, P. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological),* 42**,** 109-142.

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

MCFADDEN, D. 1978. *Modelling the choice of residential location*, Institute of Transportation Studies, University of California.

MCFADDEN, D. Year. Sociality, Rationality, and the Ecology of Choice. *In:* HESS, S. & DALY, A., eds. Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference, 2010. Emerald Group Publishing.

MCFADDEN, D. & TRAIN, K. 2000. Mixed MNL Models for Discrete Response. *Journal of Applied Econometrics,* 15**,** 447-470.

MCKELVEY, R. D. & ZAVOINA, W. 1975. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology,* 4**,** 103-120.

MCPHERSON, M., SMITH-LOVIN, L. & COOK, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology***,** 415-444.

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

MICHALEK, J. J., CERYAN, O. & PAPALAMBROS, P. Y. 2006a. Balancing marketing and manufacturing objectives in product line design. *Journal of Mechanical Design,* 128**,** 1196.

MICHALEK, J. J., CERYAN, O., STUDENT, G., PAPALAMBROS, P. Y. & KOREN, Y. 2006b. Balancing Marketing and Manufacturing Objectives in Product Line Design. *Journal of Mechanical Design,* 128**,** 1196.

NORTON, J. A. & BASS, F. M. 1987. A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High-Technology Products. *Management Science,* 33**,** 1069-1086.

ORSBORN, S., CAGAN, J. & BOATWRIGHT, P. 2009. Quantifying aesthetic form preference in a utility function. *Journal of Mechanical Design,* 131**,** 061001.

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

PÁEZ, A., SCOTT, D. M. & VOLZ, E. 2008. A discrete-choice approach to modeling social influence on individual decision making. *Environment and Planning B: Planning and Design,* 35**,** 1055-1069.

PETERSON, R. & WILSON, W. 1992. Measuring customer satisfaction: fact and artifact. *Journal of the Academy of Marketing Science,* 20**,** 61-71.

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

R-PROJECT 2010. R: A Language and Environment for Statistical Computing. http://www.r-project.org/.

RAPOPORT, A. 1953. Spread of information through a population with socio-structural bias: I. Assumption of transitivity. *Bulletin of Mathematical Biology,* 15**,** 523-533.

RATNESHWAR, S. & SHOCKER, A. 1991. Substitution in use and the role of usage context in product category structures. *Journal of Marketing Research,* 28**,** 281-295.

RICE, R., GRANT, A., SCHMITZ, J. & TOROBIN, J. 1990. Individual and network influences on the adoption and perceived outcomes of electronic messaging. *Social networks,* 12**,** 27-55.

ROBERTSON, T. & WARD, S. 1973. Consumer Behavior Research: Promise and Prospects. *Consumer behavior: theoretical sources***,** 3–42.

ROGERS, E. 1995. *Diffusion of innovations,* NY, Free Press.

ROSSI, P., GILULA, Z. & ALLENBY, G. 2001. Overcoming Scale Usage Heterogeneity. *Journal of the American Statistical Association,* 96**,** 20-31.

ROSSI, P. E., ALLENBY, G. M. & MCCULLOCH, R. 2005. *Bayesian Statistics and Marketing,* West Sussex, John Wiley and Sons, Ltd.

ROUSSEAU, A., SHIDORE, N., CARLSON, R. & FREYERMUTH, V. Year. Research on PHEV Battery Requirements and Evaluation of Early Prototypes. *In*, 2007. 16-18.

SAID, L., BOURON, T. & DROGOUL, A. Year. Agent-based interaction analysis of consumer behavior. *In*, 2002. ACM, 190.

SAMUELSON, P. A. 1948. Consumption theory in terms of revealed preference. *Economica,* 15**,** 243-253.

SCHMIDT, A., BEIGL, M. & GELLERSEN, H. 1999. There is more to context than location. *Computers & Graphics,* 23**,** 893-901.

SHAH, J. J. & WRIGHT, P. K. Year. Developing Theoretical Foundations of DfM. *In:* Proceedings of the 2000 ASME Design Engineering Technical Conference, September 2000 Baltimore, MD. ASME, New York.

SHIAU, C.-S. N., KAUSHAL, N., HENDRICKSON, C. T., PETERSON, S. B., WHITACRE, J. F. & MICHALEK, J. J. 2010. Optimal Plug-In Hybrid Electric Vehicle Design and Allocation for Minimum Life Cycle Cost, Petroleum Consumption, and Greenhouse Gas Emissions. *Journal of Mechanical Design,* 132**,** 091013.

SHIAU, C., SAMARAS, C., HAUFFE, R. & MICHALEK, J. 2009a. Impact of battery weight and charging patterns on the economic and environmental benefits of plug-in hybrid vehicles. *Energy Policy,* 37**,** 2653-2663.

SHIAU, C. S. N. & MICHALEK, J. J. 2009. Optimal product design under price competition. *Journal of Mechanical Design,* 131**,** 071003.

SHIAU, C. S. N., SAMARAS, C., HAUFFE, R. & MICHALEK, J. J. 2009b. Impact of battery weight and charging patterns on the economic and environmental benefits of plug-in hybrid vehicles. *Energy Policy,* 37**,** 2653-2663.

SIMONSON, I. & TVERSKY, A. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research,* 29**,** 281-295.

SIMPSON, T. W. 2005. Product Platform Design and Customization: Status and Promise. *AI EDAM,* 18**,** 3-20.

SNIJDERS, T. A. B. 2001. The statistical evaluation of social network dynamics. *Sociological methodology,* 31**,** 361-395.

STATA, C. 1996-2009. StataSE 9.2. College Station, Texas.

STEFFLRE, V. 1971. New Products and New Enterprises: A Report of an Experiment in Applied Social Science. *Irvine, CA: University of California*.

STOPHER, P. & MEYBURG, A. 1979. *Survey sampling and multivariate analysis for social scientists and engineers*, Lexington Books Lexington, MA.

SULLIVAN, E., FERGUSON, S. & DONNDELINGER, J. A. 2011. Exploring Differences in Preference Heterogeneity Representation and Their Influence in Product Family Design. *Proceedings of the 2011 ASME Design Engineering Technical Conferences and Computers in Engineering Conference.* Washington, DC.

SULLIVAN, J., SALMEEN, I. & SIMON, C. 2009. PHEV marketplace penetration: An agent based simulation.

SULLIVAN, J., STEPHAN, C., GOODMAN, B. & EVERSON, M. Year. Market Penetration of More sustainable vehicles: The HEV case. *In*, 2005. 13-15.

TAMHANE, A. C., ANKENMAN, B. E. & YANG, Y. 2002. 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**,** 473-494.

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

THURSTONE, L. L. 1994. A Law of Comparative Judgment. *Psychological Review,* 101**,** p266-70.

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

ULLMAN, D. G. 2002. *The Mechanical Design Process,* Boston, McGraw-Hill.

ULRICH, K. & EPPINGER, S. 2003. *Product design and development*, Irwin Professional Pub.

WALKER, J. L., EHLERS, E., BANERJEE, I. & DUGUNDJI, E. R. 2011. Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice*.

WASSENAAR, H. J. & CHEN, W. 2003. An Approach to Decision-Based Design With Discrete Choice Analysis for Demand Modeling. *Journal of Mechanical Design,* 125**,** 490-497.

WASSENAAR, H. J., CHEN, W., CHENG, J. & SUDJIANTO, A. Year. An Integrated Latent Variable Choice Modeling Approach to Enhancing Product Demand Modeling. *In:* Proceedings of the 2004 ASME Design Engineering Technical Conferences and Computers in Engineering Conference, 2004 Salt Lake City, Utah.

WASSENAAR, H. J., CHEN, W., CHENG, J. & SUDJIANTO, A. 2005. Enhancing Discrete Choice Demand Modeling for Decision-Based Design. *Journal of Mechanical Design,* 127**,** 514-523.

WASSENAAR, H. J., KUMAR, D. & CHEN, W. 2006. Discrete Choice Demand Modeling for Decision-Based Design. *In:* LEWIS, K., CHEN, W. & SCHMIDT, L. (eds.) *Decision Making in Engineering Design.* New York: ASME Press.

WATTS, D. J. 2004. *Six degrees: The science of a connected age*, WW Norton & Company.

WATTS, D. J. & STROGATZ, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature,* 393**,** 440-442.

WEN, C. & KOPPELMAN, F. 2001. The generalized nested logit model. *Transportation Research Part B,* 35**,** 627-641.

WILENSKY, U. 1999. NetLogo [computer software]. Retrieved from http://ccl.northwestern.edu/netlogo ed. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

WILENSKY, U. 2005. NetLogo Small Worlds model. http://ccl.northwestern.edu/netlogo/models/SmallWorlds ed. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University.

WILLIAMS, N., AZARM, S. & KANNAN, P. 2008. Engineering product design optimization for retail channel acceptance. *Journal of Mechanical Design,* 130**,** 061402.

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

YANNOU, B., WANG, J., RIANANTSOA, N., HOYLE, C., DRAYER, M., CHEN, W., ALIZON, F. & MATHIEU, J.-P. 2009. Usage Coverage Model For Choice Modeling: Principles And Taxonomy. *2009 ASME Design Engineering Technical Conferences.* San Diego, CA.

YANNOU, B., WANG, J. & YVARS, P. 2010. Computation of the usage contexts coverage of a jigsaw with CSP techniques. *2010 ASME Design Engineering Technical Conferences.* Montreal, Canada.

Appendix A: Sample Try-It-Out Survey Questionnaire for UCBCM (User 1, Usage Scenario 1)

Assume that you are in the market for a new saw. The four choices you have are shown as follows:

Product	**1**	**3**	**4**	**5**
Picture				
Type	Jigsaw	Jigsaw	Jigsaw	Jigsaw
Amperage	5 A	6.4 A	5.5 A	6.5 A
Speed Range High	3000 SPM	2800 SPM	3200 SPM	3100 SPM
Height	10 in.	4.6 in.	13.63 in.	4.5 in.
Weight	5.8 Lbs	10 Lbs	9.39 Lbs	10 Lbs
Price	$39.97	$97.99	$69.00	$119.00
…				

1. **Given the primary usage of <u>cutting soft wood indoor</u>, please try these products out and rate their performance on a scale from 1 to 5 (5 being the highest) in the following table:**

Product	**1**	**3**	**4**	**5**
Advance speed				
Comfort				
…				

2. **Please make a choice among these four products (which product would you like to purchase? You may not make a selection if you are not happy with any of these products).**

   ____

3. **Please tell us a little bit about yourself:**
   - Are you:
     □ male □ female
   - What is your skill level in terms of saw usage?
     □ Beginner □ Intermediate □ Experienced
4. Which one of the following groups best describes your household's total annual income before taxes?

   □ Under $50,000 □ $50,000-59,999 □ $60,000-69,999
   □ $70,000-$79,999 □ $80,000-89,999 □ $90,000-99,999
   □ $100,000-$109,999 □ $110,000-119,999 □ $120,000-129,999
   □ $130,000-$139,999 □ $140,000-149,999 □ $150,000 or more

   …

5. **Please tell us about your saw usage (up to three usages scenarios):**
Usage Scenario 1 (Primary Usage):
- Do you use saws:
  □Outdoor □Indoor
- Do you use saws to cut:
  □Soft wood □Medium wood □Hard wood

…
Usage Scenario 2:
- Do you use saws:
  □Outdoor □Indoor
- Do you use saws to cut:
  □Soft wood □Medium wood □Hard wood

…
Usage Scenario 3:
- Do you use saws:
  □Outdoor □Indoor
- Do you use saws to cut:
  □Soft wood □Medium wood □Hard wood

…
Please tell us about the importance of these three usages in percentage:
Usage Scenario 1:_______%
Usage Scenario 2:_______%
Usage Scenario 3:_______%

Appendix B: Table B.1: Results of Principal Component Analysis Using NHTS

Attributes	**Drivers & Vehicles**	**Mainstream Family**	**Working Class**	**Higher Income & Education**	**Home Owners**	**Young with Child**	**Apartment Dweller**	**African American / Black**	**Asian**	**Hispanic**	**Rural**	**Middle-Aged**	**Long Distance Worker**	**Student**	**College Graduate**	**Some College**	**Educated**	**Sales**
Household income	0.29	0.09	0.28	**0.40**	0.18	-0.01	-0.07	-0.11	0.03	-0.01	-0.16	0.09	0.10	-0.04	0.25	0.11	0.34	-0.01
Household size	**0.67**	0.57	0.11	0.00	0.03	0.29	-0.03	0.04	0.02	0.01	-0.03	0.04	-0.01	0.04	0.00	0.00	-0.06	0.03
# adults	**0.92**	0.03	0.04	-0.04	0.03	0.02	-0.02	0.03	0.03	0.01	-0.03	0.00	0.00	0.15	-0.03	-0.01	-0.04	0.03
# workers	0.47	0.18	**0.57**	0.07	0.02	0.09	-0.01	-0.02	0.02	-0.01	0.00	0.21	0.06	0.21	0.05	0.03	0.11	0.13
# drivers	**0.89**	0.11	0.10	0.00	0.05	-0.01	-0.03	-0.01	0.01	0.00	0.00	0.07	0.00	0.22	0.00	0.00	0.01	0.03
# vehicles	**0.50**	-0.01	0.18	-0.02	0.17	-0.09	-0.08	-0.02	-0.03	0.01	0.21	0.08	0.25	0.09	0.00	0.05	0.13	0.02
MSA size	0.05	0.03	0.03	0.06	-0.04	-0.01	0.06	0.02	0.12	0.03	**-0.77**	0.02	-0.01	-0.02	0.06	0.03	0.02	0.02
Urban/rural	0.03	-0.01	0.01	-0.05	0.10	-0.01	-0.03	-0.05	-0.04	0.00	**0.79**	0.00	0.05	-0.02	-0.04	-0.01	0.00	-0.02
Home ownership	0.08	-0.05	-0.04	0.05	**0.77**	-0.11	-0.07	-0.10	0.03	0.00	0.03	0.00	-0.01	-0.02	0.04	0.00	0.04	-0.01
Home type SFH	0.09	0.02	0.01	0.03	0.50	0.01	**-0.83**	0.00	-0.05	-0.02	-0.05	0.00	0.02	0.02	0.03	0.02	0.09	-0.01
Home type duplex	-0.02	0.00	-0.02	0.00	0.16	0.06	**0.80**	0.09	-0.11	-0.03	-0.19	-0.02	0.03	0.05	0.02	0.02	0.16	-0.03
Home type townhouse	-0.10	-0.04	-0.03	0.02	**-0.89**	-0.02	0.11	0.01	0.03	0.00	-0.09	-0.01	-0.01	0.00	0.02	0.01	0.04	-0.01
Home type apt/condo	-0.02	0.00	0.03	-0.07	0.01	-0.04	**0.52**	-0.10	0.16	0.05	0.37	0.03	-0.06	-0.08	-0.10	-0.06	-0.35	0.06
Gender	-0.08	0.04	0.07	-0.06	0.01	0.05	0.01	0.02	-0.02	0.00	-0.01	0.04	**-0.76**	0.03	-0.02	0.00	0.04	-0.03
Miles past 12 months	0.02	0.08	0.23	0.04	0.06	0.05	0.01	-0.01	-0.03	0.02	0.20	0.13	**0.46**	-0.09	0.05	0.06	0.20	0.23
Race white	-0.03	-0.05	-0.03	0.01	0.05	-0.04	-0.02	**-0.82**	-0.43	-0.22	0.07	0.00	0.00	-0.01	0.01	0.00	0.04	0.00
Race black	-0.02	0.01	0.01	0.00	-0.03	0.00	0.02	**0.95**	-0.11	-0.08	0.00	0.01	-0.02	0.00	-0.01	0.00	-0.01	0.00
Race Asian	0.03	0.02	0.01	0.04	-0.01	0.02	-0.01	0.09	**0.95**	-0.03	-0.07	0.00	0.00	0.01	0.03	0.00	0.03	0.00
Race Hispanic	0.00	0.01	0.00	0.00	-0.01	0.01	0.00	0.05	-0.01	**0.99**	-0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.00

Attributes	**Drivers & Vehicles**	**Mainstream Family**	**Working Class**	**Higher Income & Education**	**Home Owners**	**Young with Child**	**Apartment Dweller**	**African American / Black**	**Asian**	**Hispanic**	**Rural**	**Middle-Aged**	**Long Distance Worker**	**Student**	**College Graduate**	**Some College**	**Educated**	**Sales**
Age 16-24	0.23	0.06	0.02	-0.09	-0.04	-0.05	0.00	-0.01	0.02	0.02	-0.04	-0.14	-0.02	**0.64**	-0.06	0.05	-0.09	0.18
Age 25-34	0.09	-0.12	0.08	0.00	-0.08	**0.79**	0.02	0.03	0.01	0.02	0.00	0.07	-0.04	-0.13	0.03	0.02	-0.04	0.07
Age 35-44	-0.04	**0.76**	0.08	0.03	0.00	0.04	0.03	0.01	0.06	-0.01	0.04	-0.32	0.06	-0.02	0.07	0.01	0.17	-0.05
Age 45-54	0.06	0.02	0.16	0.03	0.02	-0.10	-0.01	0.01	0.00	0.00	-0.01	**0.90**	0.01	0.02	0.02	-0.01	0.02	0.00
Age 55-64	-0.02	-0.37	**0.61**	0.05	0.04	-0.27	-0.03	0.02	-0.04	0.00	0.00	-0.50	-0.03	-0.17	-0.01	0.00	-0.07	0.00
Age 65+	-0.15	-0.20	**-0.84**	-0.06	0.01	-0.10	0.00	-0.04	-0.02	-0.01	0.00	-0.09	0.01	-0.04	-0.05	-0.03	-0.02	-0.07
Less than highschool	0.01	0.00	-0.04	-0.05	0.01	0.01	0.00	0.05	-0.04	-0.02	0.01	-0.02	0.07	0.03	-0.02	0.00	**-0.85**	-0.03
Highschool grads	0.01	-0.02	-0.05	-0.36	-0.01	-0.03	0.03	0.00	-0.04	0.00	0.10	0.02	0.01	0.01	-0.35	**-0.76**	0.23	0.05
Some college	0.00	0.00	0.01	-0.24	-0.01	-0.01	0.01	0.01	-0.02	0.01	0.03	0.01	-0.01	0.03	-0.30	**0.88**	0.15	0.04
College grads	0.00	0.03	0.04	-0.07	0.01	0.03	-0.02	-0.02	0.02	0.00	-0.05	0.01	-0.02	-0.02	**0.97**	-0.07	0.03	-0.01
Graduate degree	-0.03	-0.01	0.02	**0.82**	0.01	0.01	-0.04	-0.02	0.06	0.00	-0.10	-0.03	-0.02	-0.04	-0.29	-0.10	0.02	-0.07
Occupation sales	0.05	0.03	0.12	-0.12	0.00	0.03	0.00	0.00	0.00	-0.01	-0.02	0.01	0.00	0.04	-0.02	0.00	0.02	**0.92**
Occupation clerical	0.05	-0.01	0.27	-0.32	0.00	-0.01	0.01	-0.02	0.04	0.00	-0.01	0.08	**-0.49**	-0.05	-0.08	0.06	0.17	-0.26
Occupation blue collar	0.02	0.05	0.26	-0.38	0.01	0.05	-0.01	-0.02	0.01	-0.01	0.01	0.10	**0.60**	0.01	-0.19	-0.08	-0.08	-0.33
Occupation professional	0.01	0.13	0.27	**0.66**	0.00	0.06	0.01	0.02	0.00	0.00	0.04	0.14	0.01	0.05	0.26	0.05	0.11	-0.21
Retire	-0.07	-0.39	**-0.65**	-0.07	0.02	-0.21	0.00	-0.01	-0.03	0.00	0.01	-0.28	-0.06	-0.19	-0.04	-0.02	-0.05	-0.10
Single	**-0.78**	-0.02	0.00	-0.05	-0.11	-0.05	0.04	0.04	-0.01	0.01	-0.01	0.04	-0.05	0.25	-0.04	0.02	-0.05	0.02
No Child	-0.28	**-0.74**	-0.18	-0.04	-0.02	-0.29	0.02	-0.04	-0.01	-0.01	0.02	-0.15	0.02	-0.27	-0.01	-0.01	0.05	-0.03
Youngest child under 5	0.06	0.28	0.06	0.04	0.02	**0.80**	0.01	0.01	0.04	0.00	0.00	-0.17	0.02	0.02	0.03	-0.01	0.03	-0.03
Youngest child 6-15	0.14	**0.74**	0.11	0.01	0.00	-0.25	-0.04	0.03	-0.02	0.01	-0.05	0.24	-0.04	-0.24	-0.01	0.02	-0.10	0.07
Youngest child 16-21	0.24	-0.01	0.11	0.02	0.01	-0.06	0.00	0.01	0.01	0.00	0.02	0.14	-0.01	**0.79**	-0.01	0.00	0.01	-0.02

Appendix C: Engineering Design Model for PHEV

The structure of a PHEV is similar to that of an ordinary hybrid electric vehicle (HEV), except that the PHEV carries a larger battery pack and offers plug-charging capability. PHEVs store energy from the electricity grid to partially offset gasoline use for propulsion. The hybrid drivetrain has several advantages in terms of improving vehicle efficiency. First, the electric motor enables the engine to operate at its most efficient load most of the time, utilizing the batteries to smooth out spikes in power demand. Second, having an additional source of power in the form of an electric motor enables designers to select smaller engine design with higher fuel efficiency and lower torque capabilities. Third, HEV and PHEV powertrains enable energy that is otherwise lost in braking to be captured to charge the battery and enable the engine to be shut off rather than idling when the vehicle is at rest.

The storage battery of a PHEV, which can be recharged using conventional electrical outlets, would allow the vehicle to drive for a limited range using energy from the electricity grid. A fully charged PHEV operates in charge-depleting mode (CD-mode) until the battery is depleted to a target state of charge (SOC), at which point the vehicle switches to charge-sustaining mode (CS-mode), using the engine to maintain the target SOC. A PHEV can be further categorized as (1) range-extended or (2) blended, depending on its energy management strategy in the charge-depleting state. A rang-extended PHEV functions as a pure electric vehicle (EV) in charge-depleting mode, using only electrical energy from the battery for propulsion and disabling any engine operation. Blended PHEVs invoke a strategy where the motor provides primary power in charge-depleting mode, but the engine is used as needed to provide additional power. In the charge-sustaining state, all PHEVs operate similarly to a standard HEV, using the engine to maintain the target battery SOC. Since the performance of blended configurations can vary widely based on a broad range of control strategy parameters, for simplicity and fair comparisons the range-extended PHEV that runs entirely on electrical power in the charge-depleting range and switch to operate like a HEV in the charge-sustaining range is considered in this work. Figure C.1 shows a typical pattern for a range-extended PHEV with a higher initial SOC for charging depleting mode and a lower SOC sustaining target. The ability to operate entirely on electricity in the charge-depleting range is advantageous for range-extended PHEVs because they are capable of operating for a time entirely on cheaper energy from the electricity grid.

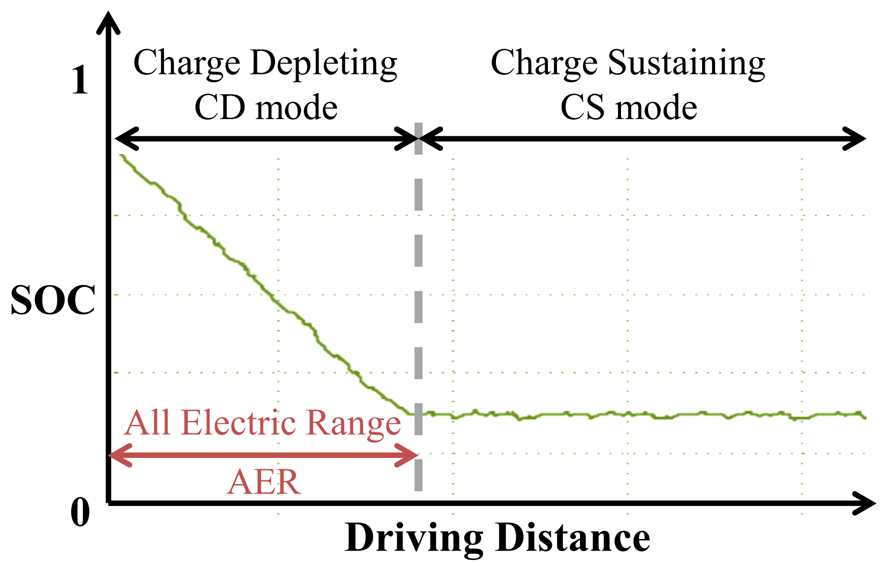


Figure C.1: Typical SOC of A PHEV

The two current dominant battery technologies considered likely candidate for the PHEV applications are nickel-metal hybride (NiMH) and lithium-ion(Li-ion) batteries. NiMH batteries have performed well and have proven reliable in existing hybrid vehicles. However, their relatively low energy density (Wh/L) and specific energy (Wh/kg) implies large, heavy batteries for extended electric travel. Li-ion batteries have higher energy density and specific energy and are benefiting from increased technological advancement, but concern remain regarding calendar life and safety (internal corrosion and high environment temperatures could cause Li-ion batteries to combust). In spite of the technical difficulties to be overcome, Li-ion batteries have been widely evaluated for their great potential as PHEV energy storage devices, thus Li-ion batteries are used in this study.

In (Shiau et al., 2009b), a vehicle engineering design model is built using data from the US Department of Energy Powertrain System Analysis Toolkit (PSAT) vehicle physics simulator to model and examine design tradeoffs between battery capacity and PHEV benefits. configurations. The effects of increasing All Electric Range (AER) on efficiency, operation, operation cost, and operations-associated GHG emissions is fairly linear in this range, as shown in the following equations.

$$\eta_{CD} = -0.010 d_{AER} + 5.67 \tag{C.1}$$

$$\eta_{CS} = -0.068 d_{AER} + 51.7 \tag{C.2}$$

$$c_{OP-CD} = 0.004 d_{AER} + 2.20 \tag{C.3}$$

$$c_{OP-CS} = 0.008 d_{AER} + 5.79 \tag{C.4}$$

where d_{AER} is AER in miles, η_{CD} and η_{CS} are the CD-mode and CS-mode efficiencies in unites of miles per kWh and miles per gallon, respectively, c_{OP-CD} and c_{OP-CS} are the operation costs per 100 miles under CD- and CS-mode, respectively.

For a distance d traveled between charges in a vehicle with an all electric range of d_{AER}, the distance traveled in CD-mode d_{CD} and the distance traveled in CS-mode d_{CS} are calculated as:

$$d_{CD} = \begin{cases} d, & \text{if } d \le d_{AER} \\ d_{AER}, & \text{if } d > d_{AER} \end{cases} \tag{C.5}$$

$$d_{CS} = \begin{cases} 0, & \text{if } d \le d_{AER} \\ d - d_{AER}, & \text{if } d > d_{AER} \end{cases} \tag{C.6}$$

The average fuel consumption per mile g is calculated in Eqn. (C.7), where η_{CS} is the fuel efficiency in CS-mode. The second performance characteristic is average operation cost, which represents the average consumer expense per mile associated with recharging cost and fuel expense, as shown in Eqn. (C.8), where $\eta_C = 88\%$ is the charging efficiency, $c_{ELEC} = \$0.11/\text{kWh}$ is the cost of electricity, and c_{GAS}=\$3.00/gal is gasoline cost.

$$g = \frac{1}{d}\left(\frac{d_{CS}}{\eta_{CS}}\right) \tag{C.7}$$

$$c_{OP} = \frac{1}{d}\left(\frac{d_{CD}}{\eta_{CD}}\frac{c_{ELEC}}{\eta_C} + \frac{d_{CS}}{\eta_{CS}}c_{GAS}\right) \tag{C.8}$$

For further evaluating the net cost implications over the vehicle lifetime, the total cost is calculated by taking into account the vehicle base cost, battery purchase price, and net present value of operation costs, and cost imposed by a potential tax on CO_2. The equation for the net present value of lifetime cost per miles is given by:

$$c_{TOT} = \frac{1}{d_{LIFE}}\left(\left(c_{VEH} + c_{BAT}\kappa\right) + \sum_{n=1}^{N}\frac{c_{OP}d_{ANUL}}{\left(1+r\right)^n}\right). \tag{C.9}$$

It is assumed that the annual vehicle miles traveled $d_{ANUL} = 12{,}500$ miles (ref EPA 2005), the vehicle lifetime $N = 12$ years, and thus vehicle lifetime mileage $d_{LIFE} = 150{,}000$ miles. Vehicle purchase cost includes the vehicle base cost (excluding the battery) c_{VEH} plus total battery capacity cost $c_{BAT} = \$1000/\text{kWh}$ multiplied by battery capacity κ, in kWh. The second term in

Eqn. (C.9) is net present value of operation costs c_{OP} (Eqn. (C.8)). The net present value of annual operational costs are calculated using a discount rate $r = 5\%$. No battery replacement is considered in this case, as many vehicle manufacturers offer batter warranty up to over 100,000 miles. Potential carbon tax charges related to GHG emission are not included as part of the life-cycle cost.

Appendix D: Small-World Network Simulation Model in NetLogo

The Small-World Network simulation model of 41,330 California respondents from NHTS is adapted from the NetLogo Small-World Network model (Wilensky, 2005) using NetLogo (Wilensky, 1999). The interface of NetLogo model is shown in Figure D.1.

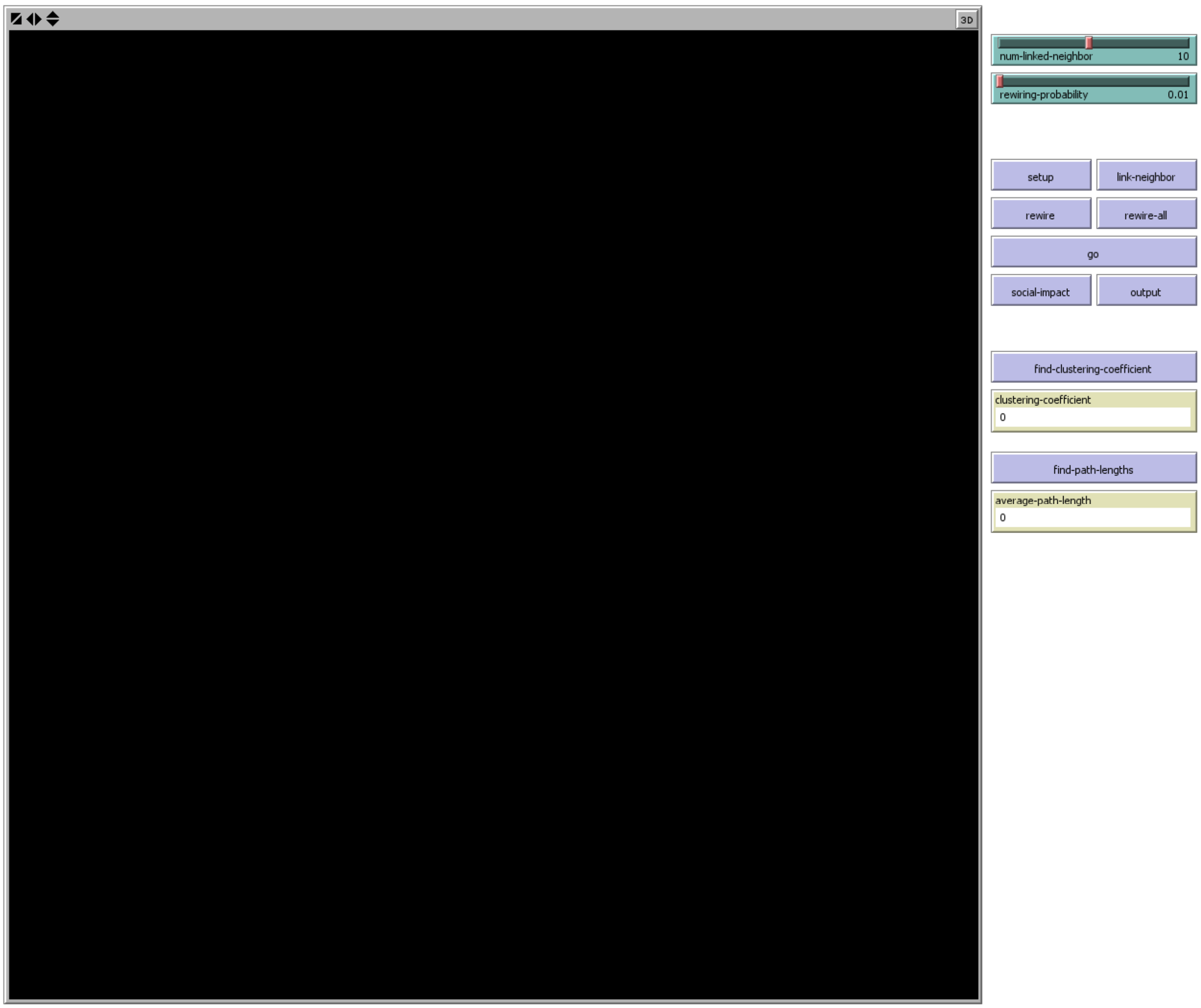


Figure D.1: NetLogo Interface of Small-World Network Simulation Model of California Respondents from NHTS

The information page of the NetLogo model is included as follows:

WHAT IS IT?

This model creates a Small-World Network with geographic data (longitude and latitude mapped from zipcode) of California sample population extracted from National Household Travel Survey 2009 (http://nhts.ornl.gov/).

HOW IT WORKS

This model is an adaptation of a model proposed by Duncan Watts and Steve Strogatz (1998). It begins with an empty world of California state. The SETUP button creates a network where each person (or "node") is mapped onto the map based their latitude and longitude location imported from CA_loc.txt file in the current folder.

The LINK-NEIGHBOR button creates connections between each person and his/her N nearest neighbors, N being the number of linked neighbors controlled by the NUM-LINKED-NEIGHBOR slider.

The REWIRE button picks a random connection (or "edge") and rewires it with the rewiring probability p controlled by the REWIRING-PROBABILITY slider. By rewiring, we mean changing one end of a connected pair of nodes, and keeping the other end the same.

The REWIRE-ALL button creates the network and then visits all edges and tries to rewire them. The REWIRING-PROBABILITY slider determines the probability that an edge will get rewired. Running REWIRE-ALL at multiple probabilities produces a range of possible networks with varying average path lengths and clustering coefficients.

The SOCIAL-IMPACT button calculates the average friend effect for each year from 2002 to 2009. The Average Friend Effect is defined as percentage of linked neighbors owning hybrid electric vehicles, as shown in the following equation:

Average Friend Effect = number of linked neighbors who own hybrid electric vehicle at time t / number of linked neighbors

To identify small worlds, the "clustering coefficient" (abbreviated "cc") and "average path length" (abbreviated "apl") of the network are calculated after the FIND-CLUSTERING COEFFICIENT or FIND-PATH-LENGTHS buttons are pressed. Networks with short average path lengths and high clustering coefficients are considered small world networks.

Average Path Length: Average path length is calculated by finding the shortest path between all pairs of nodes, adding them up, and then dividing by the total number of pairs. This shows us, on average, the number of steps it takes to get from one member of the network to another.

Clustering Coefficient: Another property of small world networks is that from one person's perspective it seems unlikely that they could be only a few steps away from anybody else in the world. This is because their friends more or less know all the same people they do. The clustering coefficient is a measure of this "all-my-friends-know-each-other" property. This is sometimes described as the friends of my friends are my friends. More precisely, the clustering coefficient of a node is the ratio of existing links connecting a node's neighbors to each other to the maximum possible number of such links. You can see this is if you press the HIGHLIGHT button and click a node, that will display all of the neighbors in blue and the edges connecting those neighbors in yellow. The more yellow links, the higher the clustering coefficient for the node you are examining (the one in pink) will be. The clustering coefficient for the entire network is the average of the clustering coefficients of all the nodes. A high clustering coefficient for a network is another indication of a small world.

HOW TO USE IT

The NUM-LINKED-NEIGHBOR slider controls the number of linked neighbor each person has. Choose a size and press SETUP.

Pressing the REWIRE button picks one edge at random, rewires it with rewiring probability.

Pressing the REWIRE-ALL button rewires all the edges with the current rewiring probability, then plots the resulting network properties on the rewire-all plot. Changing the REWIRING-PROBABILITY slider changes the fraction of links rewired after each run.

The GO button completes LINK-NEIGHBOR, REWIRE-ALL, and FIND-CLUSTERING-COEFFICIENT in a single click.

The OUTPUT button exports the average friend effect for each person into a new txt file named by the user.

The AVERAGE-PATH-LENGTH and CLUSTERING-COEFFICIENT monitors display the values for the entire network.

THINGS TO NOTICE

Note that for certain ranges of the fraction of nodes, the average path length decreases faster than the clustering coefficient. In fact, there is a range of values for which the average path length is much smaller than clustering coefficient. (Note that the values for average path length and clustering coefficient have been normalized, so that they are more directly comparable.) Networks in that range are considered small worlds.

THINGS TO TRY

Try plotting the values for different rewiring probabilities and observe the trends of the values for average path length and clustering coefficient. What is the relationship between rewiring probability and fraction of nodes? In other words, what is the relationship between the rewire-one plot and the rewire-all plot?

Set NUM-LINKED-NEIGHBOR to 10 and then press SETUP. Go to BehaviorSpace and run the VARY-REWIRING-PROBABILITY experiment. Try running the experiment multiple times. What range of rewiring probabilities result in small world networks?

EXTENDING THE MODEL

Try to see if you can introduce the scale-free property into the current small-world network by using the preferential attachment mechanism.

In a precursor to this model, Watts and Strogatz created an "alpha" model where the rewiring was not based on a global rewiring probability. Instead, the probability that a node got connected to another node depended on how many mutual connections the two nodes had. The extent to which mutual connections mattered was determined by the parameter "alpha." Create the "alpha" model and see if it also can result in small world formation.

NETWORK CONCEPTS

In this model we need to find the shortest paths between all pairs of nodes. This is accomplished through the use of a standard dynamic programming algorithm called the Floyd Warshall algorithm. You may have noticed that the model runs slowly for large number of nodes. That is because the time it takes for the Floyd Warshall algorithm (or other "all-pairs-shortest-path" algorithm) to run grows polynomially with the number of nodes. For more information on the Floyd Warshall algorithm please consult: http://en.wikipedia.org/wiki/Floyd-Warshall_algorithm

NETLOGO FEATURES

The various network/link features (introduced in NetLogo 4.0) are used extensively in this model.

Lists are used heavily in the procedures that calculates shortest paths.

RELATED MODELS

See other models in the Networks section of the Models Library, such as Giant Component and Preferential Attachment.

CREDITS AND REFERENCES

This model is adapted from:

Wilensky, U. (2005). NetLogo Small Worlds model. http://ccl.northwestern.edu/netlogo/models/SmallWorlds. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

The Small World network described here was originally published in: DJ Watts and SH Strogatz. Collective dynamics of 'small-world' networks, Nature,393:440-442 (1998)

For more information please see Watts' website: http://smallworld.columbia.edu/index.html

The small worlds idea was first made popular by Stanley Milgram's famous experiment (1967) which found that two random US citizens where on average connected by six acquaintances (giving rise to the popular "six degrees of separation" expression): Stanley Milgram. The Small World Problem, Psychology Today, 2: 60-67 (1967).

This experiment was popularized into a game called "six degrees of Kevin Bacon" which you can find more information about here: http://www.cs.virginia.edu/oracle/

HOW TO CITE

If you mention this model in an academic publication, we ask that you include this citation for the NetLogo software:

```
   - Copyright 2005 Uri Wilensky. All rights reserved. See
http://ccl.northwestern.edu/netlogo/models/SmallWorlds for terms
of use.

   COPYRIGHT NOTICE
   Copyright 2005 Uri Wilensky. All rights reserved.

   Permission to use, modify or redistribute this model is hereby
granted, provided that both of the following requirements are
followed:
   a) this copyright notice is included.
   b) this model will not be redistributed for profit without
permission from Uri Wilensky. Contact Uri Wilensky for
appropriate licenses for redistribution for profit.
```

The NetLogo software is available to download at the following link:

http://ccl.northwestern.edu/netlogo/download.shtml.

The NetLogo model file is available to public at the following link:

https://docs.google.com/open?id=0Bz-d9_dIxQRAcWhpS2ZWcFVKQzQ

The CA_loc.txt file needed as input for NetLogo model is available to public at the lowing link:

https://docs.google.com/open?id=0Bz-d9_dIxQRAS3lJZjBEemdrY2s

Vita

NAME

Lin He

PLACE OF BIRTH

Shanghai, P.R. CHINA

EDUCATION

NORTHWESTERN UNIVERSITY, Evanston, Illinois, USA — August 2012

__Ph.D.__ in Mechanical Engineering

TSINGHUA UNIVERSITY, Beijing, P.R. CHINA — July 2007

__Bachelor of Engineering__ in Vehicle Engineering, Minor in English

EMPLOYMENT

PALO ALTO RESEARCH CENTER — June - September 2011

Summer Research Intern in Automation for Engineered Systems

TRAVELERS INSURANCE COMPANIES — June - August 2010

Summer Research Intern in Predictive Modeling

TEACHING EXPERIENCES

NORTHWESTERN UNIVERSITY

Guest Lecturer, Design Research Class — Fall 2011
Trainee, Graduate Teaching Certificate Program — Fall 2010 - Summer 2011
Teaching Assistant, Computational Methods for Engineering Design Class — Fall 2008, 2009, 2010
Grader, Reliability Engineering Class — Winter 2008, 2009, 2010, 2011

HONORS & AWARDS

NORTHWESTERN UNIVERSITY
Terminal Year Fellowship — 2011 – 2012
Crown Family Fellowship — 2010 – 2011
Altair Fellowship — 2009 – 2011
Graduate Travel Grant — 2009 – 2011
Walter P. Murphy Fellowship — 2007 – 2008

BEIJING EDUCATION DEPARTMENT
Beijing Summa Cum Laude Award — 2007

TSINGHUA UNIVERSTIY
Tsinghua Summa Cum Laude Award — 2007
Excellent Academic Scholarship — 2004 – 2006
Freshman Fellowship — 2003

JOURNAL ARTICLES

He, L., Hoyle, C., and Chen, W., *Choice Modeling for Usage Context-Based Design*. Journal of Mechanical Design, 2012, **134**(3): 031007.

He, L., Chen, W., and Conzelmann, G., *Impact of Vehicle Usage on Consumer Choice of Hybrid Electric Vehicles*. Journal of Transportation Research Part D: Transport and Environment, 2012, **17**(3): p. 208-214.

He, L., Hoyle, C., and Chen, W., *Examination of Customer Satisfaction Surveys in Choice Modeling to Support Engineering Design*. Journal of Engineering Design, 2011, **22**(10): p. 669-687.

REFERRED CONFERENCE PAPERS (PEERED REVIEWED)

He, L., and Chen, W., (2012). *Incorporating Social Impact On New Product Adoption In Choice Modeling: A Case Study In Green Vehicles*. Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE), DETC 2012-71123, August 12-15, 2012, Chicago, IL.

He, L., Chen, W., and Conzelmann, G., (2011). *On Usage Context of Hybrid Electric Vehicle and Its Applicability to Choice Study*. Proceedings of the ASME 2011 IDETC/CIE, DETC 2011-48385, August 28-31, 2011, Washington, DC.

He, L., and Chen, W., (2011). *Usage Context-based Choice Modeling for Hybrid Electric Vehicles*. Proceedings of the 2011 International Conference on Engineering Design, August 15-18, 2011, Copenhagen, Denmark.

He, L., Hoyle, C., Chen, W., Wang, J., and Yannou, B., (2010). *A Framework for Choice Modeling in Usage Context-Based Design*. Proceedings of the ASME 2010 IDETC/CIE, DETC 2010-28490, August 15-18, 2010, Montreal, Quebec, Canada.

He, L., Hoyle, C., and Chen, W., (2009). *A Mixed Logit Choice Modeling Approach using Customer Satisfaction Surveys to Support Engineering Design*. Proceedings of the ASME 2009 IDETC/CIE, DETC 2009-87052, August 30-September 2, 2009, San Diego, CA.

INVITED PRESENTATIONS

- "*Influence of Usage Context and Social Network on Consumers' Choice of Hybrid Electric Vehicles*", Argonne National Laboratory, Argonne, IL, February 2012.
- "*On Usage Context of Hybrid Electric Vehicle in Choice Studies*", ASME IDETC/CIE, Washington, DC, August 2011.
- "*Usage Context-based Choice Modeling for Hybrid Electric Vehicles*", International Conference on Engineering Design, Copenhagen, Denmark, August 2011.
- "*A Framework for Choice Modeling in Usage Context-based Design*", ASME IDETC/CIE, Montreal, Quebec, Canada, August 2010.
- "*Teaching Computational Methods for Engineering Design using HyperWorks*", 2010 Americas HyperWorks Technology Conference, Novi, Michigan, April 2010.
- "*A Mixed Logit Choice Modeling Approach using Customer Satisfaction Surveys to Support Engineering Design*", ASME IDETC/CIE, San Diego, CA, August 2009.