Ph. D. Dissertation in Engineering

Modeling Consumers' New Product Adoption Behavior with Choice Set Formation Stage

선택집합 구성단계를 고려한 소비자의 신제품 채택 모형 개발

August 2017

HyungBin Moon

Technology Management, Economics, and Policy Program

Graduate School of Seoul National University

Modeling Consumers' New Product Adoption Behavior with Choice Set Formation Stage

지도교수 이 종 수

이 논문을 공학박사학위 논문으로 제출함
2017 년 8 월

서울대학교 대학원
협동과정 기술경영경제정책 전공
문 형 빈

문형빈의 공학박사학위 논문을 인준함
2017 년 6 월

위 원 장 ______이 정 동______(인)

부위원장 ______이 종 수______(인)

위 원 ______조 영 상______(인)

위 원 ______구 윤 모______(인)

위 원 ______신 정 우______(인)

Abstract

Modeling Consumers' New Product Adoption Behavior with Choice Set Formation Stage

HyungBin Moon

Technology Management, Economics, and Policy Program

The Graduate School

Seoul National University

In the present world, technological developments of unique innovations are introducing various products into the market, making the decision process of consumers more complicated than ever. As a result, consumers with limited information processing capacities are required to adopt a selection behavior that distinguishes certain alternatives for evaluating and comparing attributes of them among all possible alternatives. In other words, before the consumer makes a decision, they should formulate their own choice set with the potential alternatives. Such a change in consumer behavior is affecting the diffusion of innovative products, more specifically when the innovative product is not included or considered in the choice set of the consumer, the diffusion rate of the new product is expected to fail to meet expectations. Therefore, in order to accelerate the

diffusion of new technology, it is essential for the marketers or policy makers to recognize consumers' choice set formation behavior, and construct the marketing strategy or policies accordingly. In addition, understanding the nature of consumers' choice set formation behavior will allow more precise estimation of consumers' preferences for new technologies.

In this dissertation, I propose a new model that incorporates the consumers' choice set formation stage into product adoption behavior. Considering the model specification, I divide the consumers' decision making process into three stages: choice set formation stage, product choice stage, and usage determination stage. The proposed model can directly analyze the effect of consumers' choice set formation behavior on their preferences toward different alternatives. Furthermore, the proposed model can estimate how technological development and the changes in choice set formation behavior affect the market.

In addition, this dissertation applies the proposed model on the automobile market to demonstrate the importance of considering consumers' choice set formation behavior in market forecasts. The empirical study consists of three parts: 1) I examine how estimation results can vary according to the consumer's choice set assumption when utilizing revealed preference data. 2) Through the application of the proposed model to analyze the automobile market, I derived multiple factors that affect the consumers' choice set formation behavior, and how they change consumers' preference for different types of automobiles. In addition,

I examined how personal characteristics and the type of automobile affect the distance traveled for different consumers. 3) Based on the estimated results, I examine the proposed model's strength as a forecasting framework by analyzing the changes in the automobile market and its ripple effects in various scenarios. Considering the analysis, I assume different scenarios that represent performance improvement of automobiles' product innovation and changes in consumers' choice set formation behavior through perception changes.

Keywords: Discrete-continuous choice model; Consumer preference; Heuristics in decision making process; Choice set formation behavior; Diffusion of innovation

Student Number: 2014-30287

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background

As of 2017, “the fourth industrial revolution” is the most discussed expression that describes the future industry. The fourth industrial revolution came into focus when it was first announced in the 2016 World Economic Forum (WEF). The first industrial revolution took place with the emergence of the steam engine and the second industrial revolution allowed the development of the mass production system. Subsequently, the development of the information and communication technology led to the third industrial revolution, and finally, the fourth industrial revolution emerged, which narrowed the boundaries between physical, digital, and biological dimensions and lead the convergence of technologies in recent years.

Currently, the fourth industrial revolution is beginning to weaken the boundaries between traditional industries, giving birth to new convergence industries. This convergence has generated new innovative products. In addition, as firms accumulate the innovative capabilities, they are able to shorten the development phase of new products by increasing the efficiency and reducing costs in the production process (Gupta and Souder, 1998; Sherman, Souder, and Jenssen, 2000; Yang, 2004). Under this industrial environment, in addition to an increase in the number of products being released into the market, the product life cycle has become shorter than before.

The rapid production of more diversified and complex products involving high technologies are complicating the consumer decision-making processes[1]. While purchasing a product, consumers search for possible alternatives that suit the purpose of their purchase and evaluate them prior to making the best decision. While evaluating the alternatives, consumers identify the attributes of individual alternatives and consider the trade-offs among them, in order to consider the best alternative that fits the purpose of the purchase. However, consumers generally have limited capacity for processing information (Simon, 1955; Bettman, Luce, and Payne, 1998; Jun, Vogt, and MacKay, 2010; Liu and Dukes, 2016), constraining the number of alternatives and trade-offs to consider (Hensher, 2006)[2]. As a result, with more products in the market, it is challenging for consumers to evaluate all the possible alternatives.

As the decision making process becomes more complicated, consumers gain a tendency to simplify the process by employing heuristics. Generally, heuristics are used in psychology and behavioral economics to describe the decision making rule of a consumer in simplifying the decision making process (Tversky and Kahneman, 1973; Berkeley and Humphreys, 1982; Shanteau, 1989; Gigerenzer and Gaissmaier, 2011). One of the most significant examples of

[1] The decision making process of consumers consists of five steps: problem recognition, information search, alternative evaluation, choice, and outcomes (Dewey, 1910; Engel, Blackwell, and Kollat, 1978; Olshavsky and Granbois, 1979).
[2] In other words, consumers have bounded rationality rather than perfect rationality in reality (Simon, 1955; 1956; Kahneman and Tversky, 1979).

consumer heuristics is when the consumer arbitrarily constructs a choice set without considering all the alternatives available at the stage of comparing and evaluating the alternatives (Payne, 1976; Olshavsky, 1979; Bettman and Park, 1980; Ursic and Helgeson, 1990). This type of behavior is most common when there are too many alternatives available and limited information processing capacities challenge the consumers' evaluation. In this case, the consumer constructs a viable choice set by selecting the potential alternatives among all the possible alternatives, based on pre-existing knowledge and information seeking costs. The consumer then evaluates and compares the alternatives within the viable choice set to choose the best product that meets the purpose of the purchase.

Understanding consumer heuristics associated with choice set formation behavior is crucial in designing policies and marketing strategies for the diffusion of new technologies[3]. In complex market situations where a variety of products are released in a short period of time, the diffusion of innovation is not expected to take place if it is not considered within the consumers' choice set. Therefore, an assisting policy and marketing strategy is required to encourage the new products to be considered within the consumers' choice set at the choice set formation stage. In order to realize this, a thorough analysis on consumer preference in relation to the choice set formation stage is required. Specifically, a researcher should

[3] The diffusion process of innovation can be divided into five stage as shown in Figure 1 (Rogers, 1962). If an innovative product is not considered by the consumers in their choice set, the innovation cannot be diffused and could be captured in the chasm (Moore, 1991).

analyze how individual characteristics, such as demographics, policy recognition, and lifestyle affect the choice set formation behavior in order to construct a strategy that allows consumers to consider new products in their choice set.

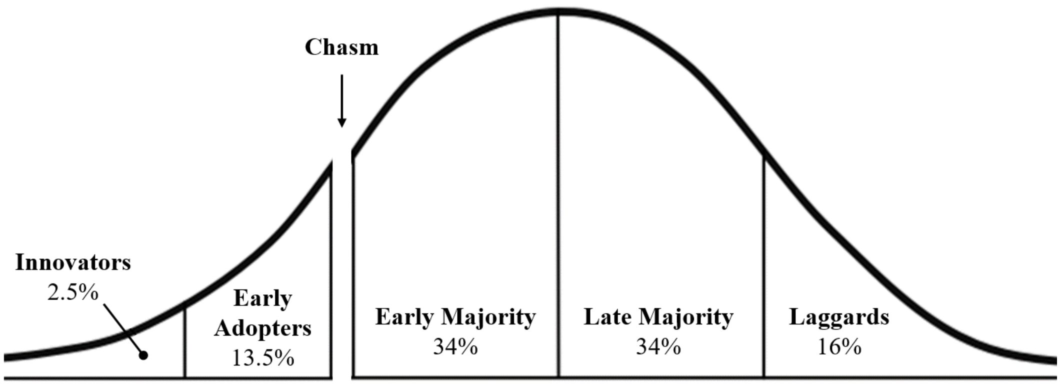


Figure 1. Diffusion process of innovation on the basis of consumer categories

In order to analyze the preferences considering the consumers' choice set formation behavior, it is necessary to construct an analysis model that can more realistically describe the consumers' decision making process. Through the construction of a choice model which considers that in the decision making process, consumers make a choice set before choosing a product, a more realistic preference analysis is possible, while the traditional models focus only on consumers' product choice. Specifically, a model considering the choice set formation behavior should be able to analyze how consumers' choice set formation behavior varies according to their individual characteristics, such as demographics, policy recognition, and lifestyle; and identify the type of consumers who tend to adopt an innovative product. Considering this, polices and marketing strategies can be designed from the consumers' viewpoint for the diffusion of an innovation. In addition, a researcher can estimate the changes in the adoption rate of the new product when consumers' tendency to include the product in the choice set increases, by constructing a model that can identify the relationship between consumer preference for the new product and the choice set formation behavior. Such a model should also be able to evaluate the effectiveness of policies and marketing strategies that assist the new innovations to be included in the consumers' choice set in terms of diffusion of innovation.

This dissertation proposes a new consumer product adoption model that reflects consumers' choice set formation behavior. The main purpose of this dissertation is to comprehend how individual characteristics affect the choice set formation behavior and how it affects consumers' preferences. Furthermore, the new model analyzes the direct and ripple effects of technological development, and the consumer behavior changes in the market by incorporating the consumers' usage determination for the chosen product. In conclusion, the proposed model can predict how consumers' product choice and usage changes when there are technological improvements and behavioral changes in consumers' choice set formation. Applying such a model is expected to provide various implications in the fields of technology management and policy.

1.2 Research Objectives

This dissertation has two main research objectives. The first objective is to propose a methodology that models consumer behavior in relation to the choice set formation, and links them with product choice and usage determination. In order to realize this, it is assumed that the new product adoption process of the consumers consists of three stages. The first stage is the choice set formation stage where consumers construct their choice set with certain alternatives among all the possible alternatives. The main focus of the model in the choice set formation stage is to analyze how individual characteristics, such as demographics, policy recognition, and lifestyle affect consumers' choice set formation behavior. A multivariate model is used to denote the consumers' choice set formation stage, which allows for multiple choices. The second stage is the product choice stage. The main focus of the model in the product choice stage is to analyze consumer preference for each alternative. In order to realize this, it is assumed that consumers' choice set formation behavior affects the preference of each alternative. More specifically, the model assumes that the baseline utility of each alternative, expressed as an alternative specific constant, is affected by the choice set formation probability, where the choice set formation probability denotes the probability of a consumer to consider a particular alternative in their choice set. The whole assumption is based on the credence that consumers' choice set is an fuzzy set, which means that the behavior of consumers considering a particular alternative in the choice set is probabilistic, where the alternatives possess different degrees of membership in a choice set (e.g., Cascetta and Papola, 2001; Chen-Yi, Ke-Ting, and Gwo-Hshiung, 2007). Considering that a choice set formation probability is estimated differently for each consumer, the model in the product choice stage is expected to identify heterogeneity in consumers' preferences. The third stage is the usage determination stage. In this stage, the model analyzes consumers' decision on how much to use the product chosen in the product choice stage. As Dubin and McFadden (1984) pointed out, there is an interconnection between the unobserved factors that affect consumers' adoption choice and usage. Considering this, I construct the model for usage determination behavior that relieves the endogeniety problem by applying the choice set formation probability derived by the model for the product choice stage as an instrumental variable in the model in the usage determination stage.

The proposed model can also be utilized as a tool to estimate the ripple effect of changes in consumers' choice set formation behavior and technological innovation in the market. This dissertation also proposes a framework for analyzing the ripple effect using the proposed model. The framework for analyzing the ripple effects is summarized as follows. First, I construct a scenario consisting of the changes in consumer perceptions and attitudes, along with improvements in product performance due to technological innovation. Further, I predict the changes in consumers' choice set formation behavior under each scenario, and analyze how it affects consumers' product choice and usage.

Another purpose of this dissertation is to derive the technology management and policy implications by applying the proposed model. As mentioned earlier, the key to designing a strategy to increase the adoption rate of a new product or innovation is to identify the characteristics of a consumer that have a high tendency to include the innovative product in their choice set. If the market environment is too complex with several alternatives, there is a high probability that a consumer with limited information processing capacity may not allow the new product to be included in their choice set. As a result, the consumer does not properly evaluate the new product, leading to a lesser chance for the product to be chosen. Therefore, analyzing the relationship between consumers' choice set formation behavior and preference for alternatives can serve as a guidance in constructing the diffusion strategy of the new product.

In this dissertation, the proposed model is applied to the automobile market. Prior to estimating the proposed model, I examine how the results of the preference analysis using revealed preference (RP) data of automobile purchase change according to the researchers' assumptions on consumers' choice set. Considering the analysis, I assume that there are four choice sets: 1) choice set consisting of all the possible alternatives, 2) choice set of randomly extracted alternatives, 3) choice set consisting of the alternatives that have characteristics similar to the actual product chosen by the consumers, and 4) choice set that consist of alternatives actually considered by the consumer before the purchase. Further, I compare the results of the preference analysis from each choice set to empirically verify previous theoretical arguments related to choice set formation. The purpose of this empirical study is to confirm the importance of preference analysis considering consumers' choice set formation behavior.

Subsequently, I apply the proposed model to estimate consumers' preferences in the automobile market. More specifically, I identify the factors that influence consumers' choice set formation behavior when purchasing an automobile and how the preference of each type of vehicle changes according to the choice set formation behavior. In parallel, I empirically prove that consumers with different choice set formation behaviors have different preferences for alternatives. In addition, I confirm how the distance traveled or the usage of the product, varies depending on the individual characteristics and vehicle attributes. Through this empirical study, I analyze how consumers' choice set formation behavior affect consumers' product choice behavior, and prove that choice set formation behavior can have a significant influence on the diffusion of new products.

Finally, based on the proposed model, I apply the framework of analyzing the ripple effect to the automobile market. In order to realize this, I assume different scenarios of changes in vehicle attributes due to technological innovation and consumer perceptions. I utilize the estimation result of the proposed model to analyze the changes in consumers' choice set formation behavior and consumers' preferences under each scenario. By analyzing the changes in the automobile market, this empirical study provides evidence for the advantages of the proposed model as a framework for forecasting and analyzing the ripple effects on the market. In particular, I aim to identify the effectiveness of technological innovation and government policies in relation to energy savings and greenhouse gas emission reductions in the automobile industry under various scenarios. Based on the estimation results, I derive valuable technology management and policy implications to serve as a guideline for policy makers and technological development.

1.3 Research Outline

The dissertation comprises of five chapters. Chapter 2 reviews previous studies related to the research topic of this dissertation. First, the studies related to the proposed model are reviewed. Considering that the proposed model is based on a multivariate choice model and discrete-continuous choice model, previous studies related to the development of these two models are reviewed. Second, previous studies providing a theoretical background of consumers' choice set formation behavior in the decision making process is reviewed. Further, I identify the limitations of previous studies, and discuss the improvements and implications of the model proposed in this dissertation.

Chapter 3 proposes the new model that incorporates the consumers' choice set formation stage. The proposed model integrates the choice set formation behavior that is influenced by consumers' individual characteristics, within the product choice and usage determination stages of the consumers' decision making process. The proposed model can be utilized to derive valuable implications for policy and marketing strategies for the diffusion of new products. More specifically, the adoption of a new product is divided into three stages: choice set formation, product choice, and usage determination. The three stages are modeled using multivariate and discrete-continuous choice models. In addition, I propose a new framework to identify the effect of technological innovation and changes in consumer perceptions on the choice set formation behavior, and analyze how these affect consumers' product choice and usage determination.

Chapter 4 consists of empirical studies where the proposed model is applied to the automobile market in order to derive implications of technology management and policy. Prior to the analysis, a survey was conducted to collect RP data on consumers' automobile purchase. Using this RP data, I empirically identify the factors that influence the three stages within the consumers' decision making process, and analyze how the stages are related to one another. In addition, in order to estimate the ripple effects of technological innovation and consumer perception change on the market, various scenario analyses are undertaken. The results of the scenario analysis provide evidences of the strengths of the proposed model as a framework to undertake market forecasting more precisely.

Finally, chapter 5 provides a summary of the proposed model and the empirical studies. In addition to the description of this dissertation's contribution, the chapter suggests the limitations of this research and recommends research topics for future researchers.

Chapter 2. Literature Review

2.1 Discrete Choice Models

2.1.1 Discrete Choice Models with Single Choice

2.1.1.1 Multinomial Logit Model

The multinomial logit model utilizes the random utility framework to analyze consumers' homogenous preference (McFadden, 1974b; Train, 2003). More specifically, the multinomial logit model is best suited to analyze consumers' utility maximization decision making behavior when there are more than two alternatives. In the multinomial logit model, the utility of consumer n choosing alternative j in choice situation t is expressed as equation (2.1). Consumer utility U_{njt} consists of a deterministic term V_{njt} and stochastic term ε_{njt}. The deterministic term is denoted by the product of the vector of preference parameters β and the vector of observable attribute x_{njt}. The stochastic term ε_{njt} in equation (2.2) is assumed to follow the independently and identically distributed (iid) type I extreme value distribution.

$$U_{njt} = V_{njt} + \varepsilon_{njt} = \beta' x_{njt} + \varepsilon_{njt} \quad \text{Eq. (2.1)}$$

$$f(\varepsilon_{njt}) = e^{-\varepsilon_{njt}} e^{-e^{-\varepsilon_{njt}}} \quad \text{Eq. (2.2)}$$

Under the assumption that consumers always seek to maximize utility, the choice probability P_{nit} in the multinomial logit model is denoted by equation (2.3). P_{nit} denotes the probability of consumer n choosing alternative i in choice situation t. Incorporating the stochastic term ε_{njt} with its distributional assumption, the choice probability P_{nit} takes the closed form of equation (2.4) (McFadden, 1974b).

$$\begin{aligned} P_{nit} &= P(U_{nit} > U_{njt},\ \forall j \neq i) \\ &= P(V_{nit} + \varepsilon_{nit} > V_{njt} + \varepsilon_{njt},\ \forall j \neq i) \\ &= P(\varepsilon_{njt} < \varepsilon_{nit} + V_{nit} - V_{njt},\ \forall j \neq i) \end{aligned} \quad \cdots\cdots \text{Eq. (2.3)}$$

$$\begin{aligned} P_{nit} &= \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{nit} + V_{nit} - V_{njt})}} \right) e^{-\varepsilon_{nit}} e^{-e^{-\varepsilon_{nit}}} d\varepsilon_{nit} \\ &= \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}} = \frac{e^{\beta' x_{nit}}}{\sum_j e^{\beta' x_{njt}}} \end{aligned} \quad \cdots\cdots \text{Eq. (2.4)}$$

The likelihood of consumer n and the likelihood for the sample sample are expressed as equation (2.5) and (2.6) respectively, where each consumer's choice is independent from the other consumers' choices. Here, y_{nit} equals one if consumer n choses alternative i in choice situation t and zero otherwise.

$$P_n = \prod_t \prod_i (P_{nit})^{y_{nit}} \quad \cdots\cdots \text{Eq. (2.5)}$$

$$Likelihood = \prod_{n=1}^{N} P_n = \prod_{n=1}^{N} \prod_{t} \prod_{i} (P_{nit})^{y_{nit}} \qquad \cdots\cdots\cdots \text{Eq. (2.6)}$$

Although the choice probability is expressed in a closed form, which provide some advantages in the estimation process, the multinomial logit model has two limitations. Firstly, the model assumes that the consumers have homogeneous preference when estimating the preference parameters. Such an assumption does not allow the researchers to comprehend the heterogeneous preferences of the consumers. Secondly, the model assumes independence from irrelevant alternatives (IIA), which limits the flexibility in describing the relationship between the alternatives. In order to overcome these limitations, many researchers have introduced other types of models, such as the mixed logit model, hierarchical multinomial logit model, and latent class model.

2.1.1.2 Mixed Logit Model

The mixed logit model expresses the heterogeneity of each consumer preference by assuming a distribution on the estimated parameters for the attributes of the product. Although the distribution of the parameters can be assumed to be normal, log-normal, truncated normal, or censored normal, the normal distribution is widely used. However, for attributes such as price, where the direction of the preference is clear (consumers prefer lower price), the parameter is often assumed to follow a log-normal distribution (Train and Sonnier, 2005). The mixed logit model that assumes the directionality for a particular attribute parameter is called the theory-constrained mixed logit model (T-MIXL) (Keane and Wasi, 2013).

Equation (2.7) denotes the utility of consumer n choosing alternative j in choice situation t, where the preference parameter β_n follows the normal distribution with mean b and variance W (McFadden and Train, 2000). The utility U_{njt} consists of the deterministic term V_{njt} and stochastic term ε_{njt}, which is assumed to follow the type I extreme value distribution.

$$\begin{aligned} U_{njt} &= V_{njt} + \varepsilon_{njt} = \beta_n' x_{njt} + \varepsilon_{njt} \\ \beta_n &\sim N(b, W) \end{aligned} \quad \cdots\cdots \text{Eq. (2.7)}$$

The choice probability of the mixed logit model, which takes the integral form of the multinomial logit probability $L_{nit}(\beta_n)$ is expressed as equation (2.8). Here, β_n is assumed to follow the distribution of the density function $f(\beta_n \mid b, W)$.

$$\begin{aligned} P_{nit} &= \int L_{nit}(\beta_n) f(\beta_n \mid b, W) d\beta_n \\ L_{nit}(\beta_n) &= \frac{e^{V_{nit}}}{\sum_j e^{V_{njt}}} = \frac{e^{\beta_n' x_{nit}}}{\sum_j e^{\beta_n' x_{njt}}} \end{aligned} \quad \cdots\cdots \text{Eq. (2.8)}$$

The likelihood of consumer n and the likelihood for the sample are expressed as equation (2.9) and (2.10) respectively, where y_{nit} equals one if consumer n choses alternative i in choice situation t and zero otherwise.

$$P_n = \int \prod_t \prod_i \{L_{nit}(\beta_n)\}^{y_{nit}} f(\beta_n \mid b, W) d\beta_n \quad \cdots\cdots \text{Eq. (2.9)}$$

$$Likelihood = \prod_{n=1}^{N} P_n = \int \prod_{n=1}^{N} \prod_t \prod_i \{L_{nit}(\beta_n)\}^{y_{nit}} f(\beta_n \mid b, W) d\beta_n \quad \cdots\cdots \text{Eq. (2.10)}$$

2.1.1.3 Hierarchical Bayesian Multinomial Logit Model

The researcher can more flexibly comprehend the heterogeneity of consumer preference by utilizing the hierarchical Bayesian estimation method in the mixed logit model estimation. The hierarchical Bayesian multinomial logit model was first introduced by Allenby and Rossi (2006). The hierarchical Bayesian multinomial logit model analyzes the heterogeneity of consumer preference assuming that the preference parameter β_n is expressed as a function of covariates z_n and denoted as equation (2.11).

$$\beta_n = \Gamma' z_n + \zeta_n, \qquad \zeta_n \sim N(0, \Sigma) \quad \cdots\cdots \text{Eq. (2.11)}$$

where Γ is the coefficient explaining the relationship between the covariate z_n and preference parameter β_n. ζ_n denotes all the unobserved heterogeneity not captured by z_n. A consumer's socio-demographic characteristics, perception, and attitude is generally utilized for the covariate z_n.

The Bayesian estimation method used in estimating the hierarchical Bayesian multinomial logit model is based on Bayes' rule, denoted in equation (2.12).

$$P(\theta \mid Y) \propto P(\theta) \times P(Y \mid \theta) \quad \cdots\cdots \text{Eq. (2.12)}$$

where Y and θ represent the observed data and the parameter, respectively. $P(\theta)$ is the prior distribution of the parameter and $P(Y \mid \theta)$ represents the likelihood function which is the distribution of the data conditional on the parameters θ. $P(\theta \mid Y)$ represents the posterior distribution that is the prior distribution updated by the likelihood.

The most commonly used Bayesian estimation procedure is the Markov chain Monte Carlo (MCMC) Gibbs sampler, which comprises of three steps as shown in equation (2.13).

$$\begin{array}{l} \Gamma \mid \Sigma, \beta_n \\ \Sigma \mid \beta_n, \Gamma \\ \beta_n \mid \Gamma, \Sigma \end{array} \quad \cdots\cdots \text{Eq. (2.13)}$$

When the prior distribution of Γ is assumed to be normal and the distribution of Σ is assumed to be inverse-Wishart, the conditional distribution of each variable is denoted by equation (2.14) and (2.15).

$$\Gamma \mid \Sigma, \beta_n, Z \;\; \forall n = \gamma \mid \Sigma, \beta, Z \;\; \forall n \sim Normal(\gamma^*, S) \quad \cdots\cdots\cdots \text{Eq. (2.14)}$$

where

$$\beta = \left(\beta_1', \beta_2', ..., \beta_n', ..., \beta_N'\right)'$$

$$\gamma^* = S\left(Z^{*\prime}(I \otimes \Sigma^{-1})\beta\right)$$

$$S = \left(Z^{*\prime}(I \otimes \Sigma^{-1})Z^*\right)^{-1}$$

$$Z^* = \left(Z_{l=1} \otimes I, Z_{l=2} \otimes I, ..., Z_{l=n} \otimes I\right)$$

$$\Sigma \mid \beta_n, \Gamma \;\; \forall n \sim Invert\ Wishart\left(K + N, (KI + N\bar{S}) / (K + N)\right) \quad \cdots\cdots\cdots \text{Eq. (2.15)}$$

where

$$\bar{S} = (1/N)\sum_n \left(\beta_n - \Gamma z_n\right)\left(\beta_n - \Gamma z_n\right)'$$

K = the number of random variables

2.1.1.4 Latent Class Model

The latent class model is another model that can reflect consumer heterogeneity. In the latent class model, it is assumed that consumers' preferences are divided into several segments, where consumers belonging to the same segment have the same preference.

When it is assumed that the consumers' preferences are divided into Q segments, and the utility that consumer n belonging to segment q obtains from alternative j in choice situation t is denoted by equation (2.16) (Greene and Hensher, 2003). Similar to the multinomial logit model, the utility $U_{njt|q}$ consists of deterministic term $V_{njt|q}$ and stochastic term $\varepsilon_{njt|q}$, where $V_{njt|q}$ is denoted as a function of the alternative's attributes.

$$U_{njt|q} = V_{njt|q} + \varepsilon_{njt|q} = \beta_q' x_{njt} + \varepsilon_{njt|q} \quad \text{Eq. (2.16)}$$

When the stochastic term $\varepsilon_{njt|q}$ in the latent class model is assumed to follow the type I extreme value distribution, the choice probability of consumer n from segment q choosing alternative i in choice situation t is denoted by equation (2.17).

$$P_{nit|q} = \frac{e^{V_{nit|q}}}{\sum_j e^{V_{njt|q}}} = \frac{e^{\beta_q' x_{nit}}}{\sum_j e^{\beta_q' x_{njt}}} \quad \text{Eq. (2.17)}$$

The likelihood function of consumer n belonging to segment q is expressed as equation (2.18), where $y_{nit} = 1$ if the consumer n belonging to segment q choses alternative i in a choice situation t and zero otherwise.

$$P_{n|q} = \prod_t \prod_i \left(P_{nit|q}\right)^{y_{nit}} \qquad \text{Eq. (2.18)}$$

Further, in order to identify the segment that the consumer n belongs to, the probability of the consumer n belonging to segment q is defined using the multinomial logit model as represented in equation (2.19). Here, it is assumed that the Qth segment is the baseline and the consumer n's individual characteristics z_n affect the consumer's segment assignment.

$$H_{nq} = \frac{\exp(\lambda_q' z_n)}{1 + \sum_{q=1}^{Q-1} \exp(\lambda_q' z_n)} \qquad \text{Eq. (2.19)}$$

Under these assumptions, the likelihood of consumer n is denoted as equation (2.20) and the likelihood for the sample is expressed as equation (2.21).

$$P_n = \sum_q H_{nq} P_{n|q} \qquad \text{Eq. (2.20)}$$

$$Likelihood = \prod_{n=1}^{N} P_n = \prod_{n=1}^{N} \sum_q H_{nq} P_{n|q} \qquad \text{Eq. (2.21)}$$

2.1.2 Discrete Choice Models with Multiple Choice

The multivariate probit model is the most representative of the discrete choice models that describes consumers' multiple choice behaviors (Chib and Greenberg, 1998). The multivariate probit model is applied to a situation where the consumer can select multiple alternatives among J alternatives. The consumer n's utility U_{nj} from choosing alternative j is assumed to be in the form of equation (2.22), where β is the parameter that represents consumer preference and x_{nj} is the explanatory variables that affect the preference for each alternative.

$$U_{nj} = V_{nj} + \varepsilon_{nj} = \beta' x_{nj} + \varepsilon_{nj} \qquad \text{Eq. (2.22)}$$

In the multivariate probit model, it is assumed that consumer n chooses alternative j ($y_{nj} = 1$) if the utility that consumer n obtains from alternative j is more than zero and does not choose alternative j ($y_{nj} = 0$) otherwise. The relationship between the discrete outcome and the latent variable (utility) is expressed as equation (2.23).

$$y_{nj} = \begin{cases} 1, & \text{if } U_{nj} > 0 \\ 0, & \text{if } U_{nj} \le 0 \end{cases} \qquad \text{Eq. (2.23)}$$

Assuming that the stochastic term $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2}, ..., \varepsilon_{nJ})$ of equation (2.23) follows a multivariate normal distribution with mean 0 and variance Σ as in equation (2.24), the probability P_n that the consumer n chooses outcomes $y_n = (y_{n1}, y_{n2}, ..., y_{nJ})$ is expressed as equation (2.25).

$$\varepsilon_n \sim MVN[0, \Sigma] \quad \cdots\cdots \text{Eq. (2.24)}$$

$$P_n = \int_{A_1} \cdots \int_{A_J} \phi_J(\varepsilon_n \mid 0, \Sigma) d\varepsilon_n \quad \cdots\cdots \text{Eq. (2.25)}$$

where

$$\phi_J(\varepsilon_n \mid 0, \Sigma) = \frac{1}{(2\pi)^{J/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}\varepsilon_n{}' \Sigma^{-1} \varepsilon_n}$$

$$A_j = \begin{cases} (-\infty, \beta' x_{nj}), & \text{if } y_{nj} = 0 \\ (\beta' x_{nj}, \infty), & \text{if } y_{nj} = 1 \end{cases}$$

2.2 Discrete-Continuous Choice Models

The discrete-continuous model, which expresses the consumers' ownership and utilization decisions, has evolved into various forms. In particular, different models have evolved based on how the jointness between the discrete stage expressing ownership and continuous stages expressing utilization was constructed. The jointness is important in the discrete-continuous model as it treats the endogeniety between the ownership and utilization stages. These problems have been actively discussed subsequent to Heckman (1976, 1978), and

Dubin and McFadden (1984). Prvious studies have suggested that endogeniety arises from the correlation between the unobserved factors influencing the ownership and utilization in real life. For example, the fact that the house is not well ventilated is an unobserved factor that increases the consumer’s utility of an air conditioner. Such an action can increase both, the probability of choice and intensity of use of the appliance (Dubin and McFadden, 1984). The existence of the correlation between unobserved factors make the way to deal with the jointness important in the discrete-continuous model. Subsequent to reviewing previous studies, the discrete-continuous model can be divided into the two categories.

2.2.1 Sequential Estimation Technique-based Model

One class of studies initiated by Heckman (1976, 1978) and Hausman (1979) attempted to correct the endogeneity problem in the discrete-continuous models by using sequential estimation techniques (e.g., Greene, 1981; Dubin and McFadden, 1984; Mannering and Winston, 1985; Train, 1986; Goldberg, 1998; West, 2004). Although there is a difference between the studies, the sequential estimation technique generally consists of the following steps. First, the model assumes the distribution of the stochastic term in the discrete choice equation expressing the consumers’ ownership choice. Second, based on the estimation, the correlation term is estimated. The correlation term can be constructed using the estimation result of the discrete choice equation, and the term means the expected value of stochastic term in the continuous equation. The correlation term is then added into the continuous equation, which resolves the endogeneity between the two equations.

Among the sequential estimation technique-based models, Dubin and McFadden (1984)'s model is one of the most representative that became the basis for several succeeding researches. They applied the discrete-continuous model in analyzing consumers' choice and utilization of residential electric appliances. The Dubin and McFadden (1984)'s model incorporated Hausman (1979, 1981)'s method of deriving the indirect utility functions from econometric partial demand systems into the McFadden (1974a)'s discrete choice model.

In order to construct a generalized discrete-continuous model, it is assumed that the consumer's income is y and portfolio i's price (annualized total life cycle cost) is r_i in the Dubin and McFadden (1984)'s model, where the consumer has m number of mutually exclusive and exhaustive appliance portfolios. As denoted in equation (2.26), they considered the unit electricity consumption (UEC) as a linear function of income.

$$x_1 = \beta_i (y - r_i) + m^i (p_1, p_2) + \nu_{1i} \quad \cdots\cdots \text{Eq. (2.26)}$$

where p_1 is the price of electricity, p_2 is the price of an alternative energy source other than electricity, m^i is the linear function in parameters, and v_{1i} is a term that is affected by the discrete choice i and follows a particular distribution. The general solution of the indirect utility function, which is derived from the demand equation (UEC), is expressed as equation (2.27). Here, ψ is the increasing function.

$$u = \psi([M^i(p_1, p_2) + y - r_i + v_{1i} / \beta_i]e^{-\beta_i p_1}, p_2, v_{2i}) \quad \cdots\cdots \text{Eq. (2.27)}$$

where $M^i(p_1, p_2) = \int_{p_1}^{0} m^i(t, p_2)e^{\beta_i(p_1 - t)}dt$

Under these settings, the demand for the alternative energy x_2 is denoted by equation (2.28).

$$x_2 = -M_2^i(p_1, p_2) - e^{\beta_i p_1}\psi_2 / \psi_1 \quad \cdots\cdots \text{Eq. (2.28)}$$

where $M_2^i = \partial M^i / \partial p_2$ and $\psi_2 / \psi_1 = (\partial \psi / \partial p_2) / (\partial \psi / \partial p_1)$.

Equation (2.27) can also be expressed as equation (2.29) by assuming $v_{1i} = \eta$ in the generalized indirect utility.

$$u = [M^i(p_1, p_2) + y - r_i + \eta / \beta_i] e^{-\beta_i p_1} + v_{2i} \quad \cdots\cdots \text{Eq. (2.29)}$$

Here, the discrete choice probability satisfies equation (2.30).

$$\begin{aligned} P_i = \mathrm{Prob}\{ & [M^i(p_1, p_2) + y - r_i + \eta / \beta_i] e^{-\beta_i p_1} \\ & > [M^j(p_1, p_2) + y - r_j + \eta / \beta_j] e^{-\beta_j p_1} \text{ for } j \neq i \} \end{aligned} \quad \cdots\cdots \text{Eq. (2.30)}$$

Expressing the indirect utility function in a simple functional form by assuming a specific function for m^i provides equation (2.31). In this form, $\beta_i = \beta$ is assumed for all the alternatives.

$$u = \left(\alpha_0^i + \frac{\alpha_1^i}{\beta} + \alpha_1^i p_1 + \alpha_2^i p_2 + \beta(y - r_i) + \eta \right) e^{-\beta p_1} - \alpha_5 \ln p_2 + v_{2i} \quad \cdots\cdots \text{Eq. (2.31)}$$

Here, the UEC equation and the choice probability is expressed as equation (2.32) and (2.33), respectively.

$$x_1 = \alpha_0^i + \alpha_1^i p_1 + \alpha_2^i p_2 + \beta(y - r_i) + \eta \quad \cdots\cdots \text{Eq. (2.32)}$$

$$P_i = \mathrm{Prob}(v_{2j} - v_{2i} < W_i - W_j \text{ for } j \neq i) \quad \cdots\cdots \text{Eq. (2.33)}$$

where $W_i = V_i e^{-\beta p_1} = \left(\alpha_0^i + \frac{\alpha_1^i}{\beta} + \alpha_1^i p_1 + \alpha_2^i p_2 - \beta r_i \right) e^{-\beta p_1}$

Equations (2.32) and (2.33) are the discrete-continuous model suggested by Dubin and McFadden (1984) and can be arbitrarily estimated under the assumption that the stochastic terms of the appliance portfolio choice and the UEC are independent of each other. However, they pointed out that such an estimation would not describe the real world choice. In reality, unobserved factors that increase the intensity of use are likely to increase the choice probability of the product, thus making it challenging to state that product choice and usage determination are independent of each other. As a result, Dubin and McFadden (1984) proposed a new method to express the relationship between the product choice and usage determination.

When it is assumed that individual characteristics w can also affect the indirect utility function, equation (2.31) can be rewrited in a simple form as in equation (2.34).

$$u = \left(\alpha_0^i + \frac{\alpha_1^i}{\beta} + \alpha_1^i p_1 + \alpha_2^i p_2 + w'\gamma + \beta(y - r_i) + \eta \right) e^{-\beta p_1} + \varepsilon_i \quad \cdots\cdots\cdots\cdots \text{Eq. (2.34)}$$

Here, it is assumed that the annualized total life cycle cost can be written in the same form as equation (2.35).

$$r_i = \sum_{j=1}^{m} p_j q_{ji} + \rho r_{ki} \quad \cdots\cdots \text{Eq. (2.35)}$$

where $\rho = \rho_0 + \rho_1 y$ and $q_{ji} = \tilde{q}_j + \tilde{q}_{ji}$

In the above equation, r_{ki} is the capital cost of portfolio i, ρ is the discount rate, and q_{ji} is the annual fuel consumption of the fuel j within the given portfolio i. Here, it is assumed that q_{ji} is the sum of $\tilde{q}_j$ and $\tilde{q}_{ji}$, where $\tilde{q}_j$ is annual consumption of fuel j which is independent of portfolio choice and $\tilde{q}_{ji}$ is annual comsumption of fuel j by portfolio i.

The stochastic term (unobserved factor) ε_i in equation (2.34) is assumed to follow the independent extreme value distribution as in equation (2.36).

$$f(\varepsilon_i) = e^{-\varepsilon_i} e^{-e^{-\varepsilon_i}} \quad \cdots\cdots \text{Eq. (2.36)}$$

The portfolio choice probabilities are represented in a nonlinear multinomial logit form as in equation (2.37).

$$
\begin{aligned}
P_i &= \text{Prob}[u_i > u_j \text{ for } j \neq i] \\
&= \text{Prob}[\varepsilon_i - \varepsilon_i < (\alpha_0^i - \alpha_0^j - \beta(p_1(q_{1i} - q_{1j}) + p_2(q_{2i} - q_{2j})) \\
&\qquad\qquad - \beta\rho(r_{ki} - r_{kj}))e^{-\beta p_1} \text{ for } j \neq i] \qquad \cdots\cdots\cdots \text{Eq. (2.37)} \\
&= \frac{\exp[(\alpha_0^i - \beta(p_1 q_{1i} + p_2 q_{2i}) - \beta\rho r_{ki})e^{-\beta p_1}]}{\sum_{j=1}^{m} \exp[(\alpha_0^j - \beta(p_1 q_{1j} + p_2 q_{2j}) - \beta\rho r_{kj})e^{-\beta p_1}]}
\end{aligned}
$$

Next, when Roy's identity is applied to equation (2.34), the demand function can be expressed as equation (2.38). Further, two methods are applied to consider the dependence between the product choice and usage determination.

$$
\begin{aligned}
x - q_{1i} &= \sum_{j=1}^{m} \alpha_0^i \delta_{ji} + \alpha_1^i p_1 + \alpha_2^i p_2 + w'\gamma + \beta(y - \sum_{j=1}^{m} PIOP_j \cdot \delta_{ji}) \\
&\quad - \beta\rho \sum_{j=1}^{m} PICP_j \cdot \delta_{ji} + \eta \qquad \cdots\cdots\cdots \text{Eq. (2.38)}
\end{aligned}
$$

where δ_{ji} is the dummy variable that has a value of 1 when $i = j$, $PIOP_j$ is the operating cost, and $PICP_j$ is the capital cost.

.

Reduced form method

By using the estimated choice probability $\hat{P}_j$ from equation (2.37) as a proxy for δ_{ji}, the dependency between the residual η of the usage function and the choice dummy δ_{ji} can be considered in the model.

$$x - q_{1i} = \sum_{j=1}^{m} \alpha_0^i \hat{P}_j + \alpha_1^i p_1 + \alpha_2^i p_2 + w'\gamma + \beta(y - \sum_{j=1}^{m} PIOP_j \cdot \hat{P}_j) - \beta\rho \sum_{j=1}^{m} PICP_j \cdot \hat{P}_j + \xi_1 \quad \cdots\cdots\cdots\cdots \text{Eq. (2.39)}$$

Conditional expectation correlation method

Another method to handle the endogeniety between the product choice and usage determination is to include a new term that permits a consistent estimate of $E(\eta \mid i)$.

$$x - q_{1i} = \sum_{j=1}^{m} \alpha_0^i \delta_{ji} + \alpha_1^i p_1 + \alpha_2^i p_2 + w'\gamma + \beta(y - \sum_{j=1}^{m} PIOP_j \cdot \delta_{ji}) - \beta\rho \sum_{j=1}^{m} PICP_j \cdot \delta_{ji} + \sum_{j \neq i}^{m} \gamma_j \left[\frac{\hat{P}_j \ln \hat{P}_j}{1 - \hat{P}_j} + \ln \hat{P}_i \right] + \xi_2 \quad \cdots\cdots\cdots \text{Eq. (2.40)}$$

2.2.2 Multiple Discrete-Continuous Extreme Value Model

The multiple discrete-continuous extreme value (MDCEV) model proposed by Bhat (2005, 2008) is another model that analyzes the product choice and usage determination. The MDCEV model is useful in analyzing how consumers determine their product portfolio and usage when using multiple goods. In addition, the model combines the product choice and usage determination stages into one estimation stage, handling for the endogeneity problem in the previous models (Spissu, Pinjari, Pendyala, and Bhat, 2009).

Several studies that utilized the MDCEV model primarily analyzed the consumers' choice of vehicle (product choice) and travel distance (usage determination) in the automobile market (e.g., Bhat and Sen, 2006; Ahn, Jeong, and Kim, 2008; Fang, 2008; Bhat, Sen, and Eluru, 2009). These studies estimated the baseline utility for each vehicle alternative by using the vehicle choice and mileage information of the consumer. Further, they observed the tendency of diminishing marginal utility for each alternative. In addition, the MDCEV model was also used to investigate media usage behavior (e.g., Woo, Choi, Shin, and Lee, 2014), fuel choice behavior for heating (e.g., Jeong, Kim, and Lee, 2011), and spending behavior (e.g., Ferdous, Pinjari, Bhat, and Pendyala, 2010).

According to the MDCEV specifications, when the consumer n chooses J number of alternatives among K, and decides to use m_j amount for alternative j, the utility that consumer n obtains from alternative j can be expressed as equation (2.41).

$$U_n(m_1,...,m_J,0,...,0) = \sum_{j=1}^{K} \Psi(x_j)(m_j+\gamma)^{\alpha_j} \quad \cdots\cdots \text{Eq. (2.41)}$$

where $\Psi(x_j)$ is the baseline utility from choosing j, and γ is the translation parameter that decides whether there is an interior corner solution. $\gamma \neq 0$ denotes that there is a corner solution which allows the possibility of alternative j not being used at all. $\gamma = 0$ specifies the existence of the interior solution, referring to the usage being greater than zero for all the alternatives (Kim, Allenby, and Rossi, 2002; Bhat, 2005; 2008). α_j defined by $1/(1+\exp(-\delta_j))$ is the satiation parameter indicating the degree of diminishing marginal utility and has a value between 0 and 1. However, in the MDCEV specification such as equation (2.41), it is challenging to simultaneously estimate γ and α_j due to the identification problem (Bhat, 2008). Therefore, in order to solve the identification problem, several previous studies estimated the value of γ and α_j after assuming that γ is the same for all the alternatives.

In a general MDCEV model, it is assumed that the baseline utility $\Psi(x_j)$ has always positive value and is expressed in random utility form as in equation (2.42).

$$\Psi(x_j, \varepsilon_j) = \Psi(x_j) e^{\varepsilon_j} = \exp(\beta' x_j + \varepsilon_j) \quad \cdots\cdots \text{Eq. (2.42)}$$

Thus, equation (2.41) can be expressed as equation (2.43).

$$U = \sum_{j=1}^{K} [\exp(\beta' x_j + \varepsilon_j)](m_j + \gamma)^{\alpha_j} \quad \text{Eq. (2.43)}$$

In addition, the model is expected to consider the budget constraints of the consumers as consumers constantly choose and use the alternatives that maximize their utility within their budget constraints. Thus, the budget constraint can be expressed as equation (2.44).

$$\sum_{j=1}^{K} m_j = M \quad \text{Eq. (2.44)}$$

where, m_j represents the usage of alternative j and M represents the total possible usage (budget constraint).

As a result, the utility maximization problem of equation (2.43) can be solved by applying the Lagrangian method and the Kuhn-Tucker condition within the budget constraint as equation (2.44). Equation (2.45) depicts the probability of using m_j of an alternative J among a total of K alternatives.

$$P(m_2^*, m_3^*, ..., m_j^*, 0, 0, ..., 0) = \left[\prod_{i=1}^{J} c_i\right]\left[\prod_{i=1}^{J} \frac{1}{c_i}\right]\left[\prod_{i=1}^{J} e^{V_i} \Big/ \left(\sum_{j=1}^{K} e^{V_i}\right)^J\right](J-1)! \quad \text{Eq. (2.45)}$$

where $c_i = (1-\alpha_i / m_i^* + \gamma)$ and $V_j = \beta' x_j + \ln\alpha_j + (\alpha_j - 1)\ln(m_j^* + \gamma)$. The MDCEV model has evolved into a mixed-MDCEV, multiple discrete-continuous nested extreme value (MDCNEV), and multiple discrete-choice probit (MDCP) models, while relieving the assumptions on independence between alternatives and homogeneity in consumers' preferences.

2.3 Discrete Choice Models considering Choice Set Formation

2.3.1 Heuristics in the Decision Process

Every consumer evaluates the available alternatives in order to choose the specific goods or services that maximize their utility. However, individual consumer's decision process differ based on various factors such as the individual's information processing capabilities and choice environment. Such differences exist owing to the fact that in reality, consumers have bounded rationality rather than perfect rationality (Simon, 1955; 1956; Kahneman and Tversky, 1979). Therefore, if the choice situation is complicated, heuristics are used to simplify the decision process (Mueller and de Haan, 2009).

In the decision process, two types of heuristics are often used by the consumers. One is to select several alternatives from all the possible alternatives and formate choice set (Figure 2). The other heuristic is to evaluate the alternatives in a choice set based on specific attributes (Figure 3).

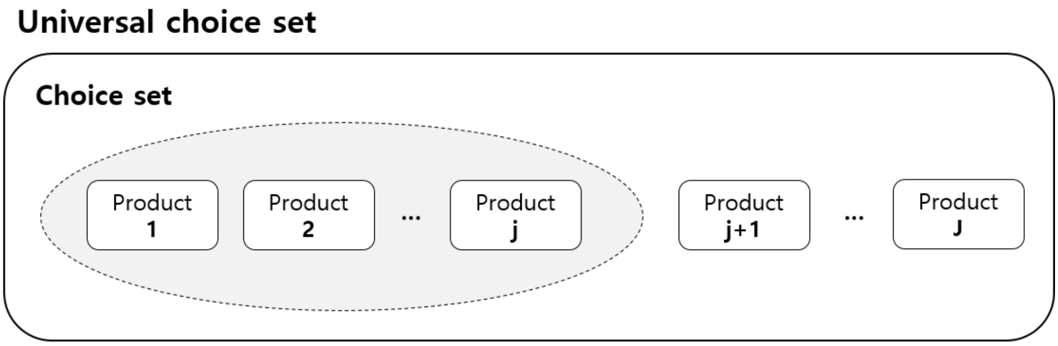


Figure 2. Construction of the choice set

Figure 3. Attribute non-attendance

Consumers use the heuristics that construct their own choice set when there are too many alternatives to evaluate with limited information processing capacity, or when there is not enough information about the alternatives. When the consumer applies the heuristics that construct their choice set, then the decision process consists of two stages (Payne, 1976; Olshavsky, 1979; Bettman and Park, 1980; Ursic and Helgeson, 1990). The first stage is the screening stage, where the consumer removes several alternatives by applying a simple non-compensatory rule. In the second stage, the consumer evaluates the remaining alternatives by applying a compensatory multi-attribute weighting rule, then identifies the best alternative.

Another heuristics introduced earlier is the “attribute non-attendance” behavior. In other words, consumers does not take into account all attributes of alternatives when evaluating the alternatives (Lockwood, 1996; Rekola 2003; Hensher, Rose, and Greene, 2005; Puckett and Hensher, 2008; Scarpa, Gilbride, Campbell, and Hensher, 2009). An attribute non-attendance behavior primarily occurs when the consumer lacks information about the attributes, when it is challenging to evaluate the difference of attributes between the alternatives, and when the consumer considers that certain attributes are not important. As a result, the consumer evaluates the alternatives based on a few or one of the easily distinguishable attribute.

2.3.2 Semi-compensatory Discrete Choice Models

Consumers’ choice set formation behavior in the decision process has been analyzed in the fields of consumer preference analysis and behavioral economics since the 1970s. Previous studies assume that, when there are too many alternatives in the choice situation, consumers follow the two-stage choice process, which consists of the non-compensatory heuristics and compensatory evaluation process (Payne, 1976; Lussier and Olshavsky, 1979). More specifically, in the first stage, the consumer uses non-compensatory heuristics to reduce the size of the choice set and simplify the choice task. Further, in the second stage, the consumer performs a compensatory evaluation process that evaluates alternatives in their choice set and chooses the best alternative. In this respect, the discrete choice model considering the consumers' choice set formation stage is called the semi-compensatory discrete choice model.

There are two main approaches to model the consumers' choice set formation behavior. The first is the explicit approach, which was initiated from the two-stage choice model of Manski (1977) (e.g., Swait and Ben-Akiva, 1986, 1987; Roberts and Lattin, 1991; Andrews and Srinivasan, 1995; Ben-Akiva and Boccara, 1995; Horowitz and Louviere, 1995; Chiang, Chib, and Narasimhan, 1999; Swait, 2001; Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftn, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). The studies model the consumers' choice set formation behavior and assume that these behaviors have a direct effect on the choice probability of the alternative. These studies are characterized by the assumption that the consumers' choice set follows the nature of the crisp set considered in the set theory. In other words, it is an approach that considers a particular alternative to be included or not included in the choice set. The studies of this type are elaborated below.

The second approach is the implicit approach, which was first introduced by Cascetta and Papola (2001) (e.g., Quattrone and Vitetta, 2011; Chen-Yi, Ke-Ting, and Gwo-Hshiung, 2007). These studies assume that the availability and perception of a particular alternative affects the preference of the alternative (preference adjustment effect). In this approach, it is assumed that the consumers' choice set follows the properties of the fuzzy set[4]. In other words, the possibility of a certain alternative to be included in the consumers' choice set is represented by a probability. With the application of this approach, Cascetta and Papola (2001) proposed the implicit availability/perception random utility (IAPRU) model. As shown in equations (2.46) ~ (2.48), the IAPRU model assumes that the degree of membership $\mu_C^n(i)$ representing the tendency of consumer n considering alternative i in choice set C affects the preference of a particular alternative ($0 \le \mu_C^n(i) \le 1$).

$$U_{ni} = V_{ni} + \ln \mu_C^n(i) + \varepsilon_{ni} \quad \cdots \text{Eq. (2.46)}$$

$$\mu_C^n(i) = \ln \bar{\mu}_C^n(i) - \frac{\left(1 - \bar{\mu}_C^n(i)\right)}{2\bar{\mu}_C^n(i)} + \eta_{ni} \quad \cdots \text{Eq. (2.47)}$$

[4] In the classical set theory, it is assumed that the elements are included or not included in a set (on/off). However, the elements have a degree of membership to being included in a set in the fuzzy set theory (Zadeh, 1965; Zimmermann, 2011). A fuzzy set theory is generally employed in describing the fuzziness of human behavior.

$$\bar{\mu}_C^n(i) = \frac{1}{1+\exp\left(\sum_k -\gamma_k Y_{nik}\right)} \quad \text{Eq. (2.48)}$$

where V_{ni} is the deterministic term of the utility that the consumer n obtains form alternative i, ε_{ni} is the stochastic temr of the utility, and η_{ni} is the error occurring during the Taylor expansion. In addition, Y_{nik} is the term that represents the alternative attributes and individual characterisitcs that affect alternative i's degree of membership

Previous studies that deal with the subject of consumers' choice set formation have considered the consumers' choice set formation behavior in the model for the following two purposes. One class of studies sought to improve the analysis results of consumers' preferences by considering the choice set formation behaviors in the decision making process (Swait and Ben-Akiva, 1986; Roberts and Lattin, 1991; Andrews and Srinivasan, 1995; Ben-Akiva and Boccara, 1995; Horowitz and Louviere, 1995; Chiang, Chib, and Narasimhan, 1999; Swait, 2001; Cascetta and Papola, 2001). Rather than observing the consumers' choice set formation behavior, this class of studies assume a choice set formation rule in the model. Another group of studies attempted to determine the factors that influence the consumers' choice set formation behaviors and analyze the magnitude of their effects (Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftn, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013).

The most representative study among the first class of studies is Swait and Ben-Akiva (1986), which theoretically pointed out the potential problems that could arise when researchers misspecify a choice set in analyzing consumers' preferences. The study pointed out the following two main problems that occur when the consumers' choice set is assumed to be equal to the universal choice set: (1) The alternative specific constants are downward biased for the alternatives which tend to be considered in the choice set alone, not with other alternatives. (2) By assuming that the alternatives that were not actually considered are included in the choice set, the marginal utility for the product attributes is distorted. Subsequent to finding that considering the choice set formation is important for sophisticated estimation in the analysis of consumer preferences, several studies attempted to model the choice set formation behavior.

The models developed in the first class of studies assume the two-stage approach proposed by Manski (1977) as the basic model. In the two-stage approach of Manski (1977), it is assumed that the probability of the consumer n choosing an alternative i is expressed as equation (2.49).

$$P_{ni} = \sum_{C_n \in G} P(i \mid C_n) Q_n(C_n) \quad \text{Eq. (2.49)}$$

where C_n is the choice set formed by the consumer n, G is the set of the nonempty subset of universal choice set M , $P(i\,|\,C_n)$ is the probability of consumer n choosing alternative i from choice set C_n, and $Q_n(C_n)$ is the probability of consumer n constructing choice set C_n.

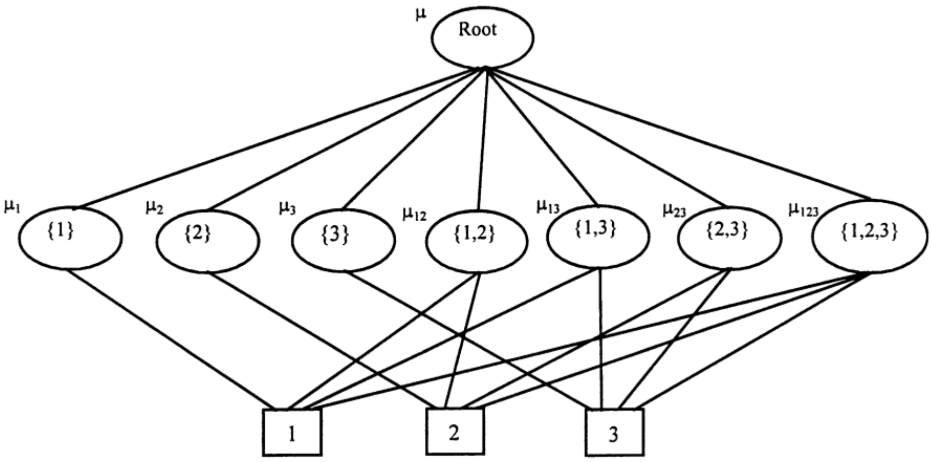
One of the most widely known expansion model, the Swati and Ben-Akiva (1987) model attempted to incorporate the consumers' choice set formation behavior $Q_n(C_n)$ into the model by assuming equation (2.50). This model assumed consumer n's behavior to include the alternative i into the choice set (A_{ni}) as a binary decision behavior ($A_{ni}=1$ if alternative i is included in the choice set C_n and zero otherwise), and that the product choice behavior given the choice set is expressed as the multinomial logit form. Here, as shown in equation (2.51), D_i represents the choice set formation probability, which denotes the probability of consumer n including alternative i within the choice set.

$$Q_n(C_n)=\frac{\prod_{i\in C_n} D_i \prod_{j\in M-C_n}(1-D_j)}{1-\prod_{j\in M}(1-D_j)} \quad \text{Eq. (2.50)}$$

$$D_i=\Pr(A_{ni}=1),\ \forall i\in M \quad \text{Eq. (2.51)}$$

Ben-Akiva and Boccara (1995), the expansion model along with Swait and Ben-Akiva (1987)'s model, assumed that the choice set formation probability is expressed in a recursion form, where they proposed a model in which the choice set formation probability D_i is expressed in the form of equation (2.52). While the model proposed by Swait and Ben-Akiva (1987) assumed that the probability of each alternative being included in the choice set is independent from each other, the model of Ben-Akiva and Boccara (1995) assumed that there is a relationship among the alternatives.

$$D_i = \Pr(C_i) = \Pr(A_{ni} = 1 \mid C_{i-1}) \Pr(C_{i-1}) \quad \text{for } i = 2, ..., J \qquad \text{Eq. (2.52)}$$



Source: Swait (2001)

Figure 4. GenL model suggested by Swait (2001)

Another representative extension model is the choice set generation logit (GenL) model proposed by Swait (2001). The model applied the discrete choice model of the generalized extreme value (GEV) family to reflect the consumers' choice set formation behavior. GenL model assumes that the consumers' choice set formation behavior and the product choice behaviors are expressed as shown in Figure 4 when the consumer has three alternatives to choose from.

The choice probability of the consumer under the GenL model follows the GEV model as shown in equations (2.53) ~ (2.56). Here, C_k represents the choice set including the alternatives $i \in \{1,...,J\}$ and $K_i = \{k \mid 1 \le k \le K,\ i \in C_k\}$ is assumed to represent all the possible C_k.

$$P_i = \sum_{k \in K_i} P(i \mid C_k) Q(C_k) \quad \text{Eq. (2.53)}$$

$$P(i \mid C_k) = \frac{\exp(\mu_k V_i)}{\sum_{j \in C_k} \exp(\mu_k V_j)} \quad \forall i \in C_k,\ k = 1,...,K \quad \text{Eq. (2.54)}$$

$$Q(C_k) = \frac{\exp(\mu I_k)}{\sum_{r=1}^{K} \exp(\mu I_r)},\ k = 1,...,K \quad \text{Eq. (2.55)}$$

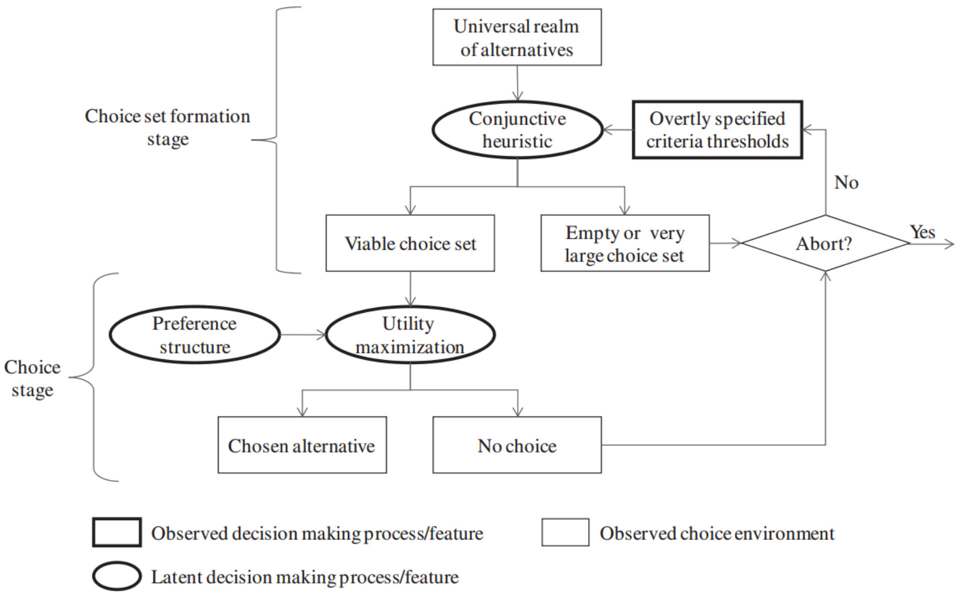
$$I_k = \frac{1}{\mu_k} \ln\left(\sum_{j \in C_k} \exp(\mu_k V_j) \right),\ k = 1,...,\mathrm{K} \quad \text{Eq. (2.56)}$$

As metioned above, the other class of studies on choice set formation behavior attempted to determine the factors that influence the consumers' choice set formation behaviors and analyze the magnitude of their effects (Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftn, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). Kaplan, Shiftn, and Bekhor (2012) observed the actual choice set information of the consumers, in order to identify the heuristics applied by the consumers in constructing the choice set. More specifically, the study focuses on identifying various criteria that consumers consider when selecting potential alternatives in constructing a choice set, and how consumers' thresholds are distributed for each criterion. Considering this, Kaplan, Shiftn, and Bekhor (2012) show that the decision-making process of consumers is divided into a choice set formation stage and choice stage as shown in Figure 5.

The second class of studies analyzed the pattern of choice set formation behaviors based on the fact that consumers' choice set formation stages consist of a combination of criteria thresholds. These studies assumed the probability of the consumers' choice set formation behavior $Q_n(C_n)$ to be in the form of equation (2.57). Here, it is assumed that there are H number of criterions that affect the decision of the choice set formation.

$$Q_n(C_n) = P(t_{1n}^* \cap t_{2n}^* \cap ... \cap t_{Hn}^*) = P(t_{1n}^*) \times P(t_{2n}^*) \times ... \times P(t_{Hn}^*) \quad \cdots\cdots\cdots\cdots \text{Eq. (2.57)}$$

where $P(t_{hn}^*)$ represents the probability of the consumer n choosing threshold t_{hn}^* for a criterion h.



Source: Kaplan, Shiftn, and Bekhor (2012)

Figure 5. Two-stage decision process in Kaplan, Shiftn, and Bekhor (2012)

Previous studies, as shown in equation (2.58), assumed consumer n's threshold t^*_{hn} for criterion h as a function of personal attribute Z_{hn}.

$$t^*_{hn} = \alpha'_h Z_{hn} + \varepsilon_{hn} \quad \text{Eq. (2.58)}$$

Further, consumers' choice set formation behavior was analyzed by applying an ordered-response model. The threshold t^*_h is defined by M_h

number of threshold categories ($m_h = 1, 2, ..., M_h$) and θ_h is defined by the set of constraints ($\theta_1, \theta_2, ..., \theta_h, ..., \theta_H$) that distinguish the threshold categories satisfying $\theta_1 < \theta_2 < ... < \theta_{m_h} < ... < \theta_{M_h}$. In other words, if threshold t_h^* is included in the threshold category m , then it has a value between $\theta_{(m-1)_h}$ and θ_{m_h} as in equation (2.59).

$$\theta_{(m-1)_k} < t_h^* < \theta_{m_k} \quad \cdots\cdots \text{Eq. (2.59)}$$

Assuming that the unobserved factor ε_{hn} of criterion h is independently and identically distributed (iid) normal across individuals, $Q_n(C_n)$ can be expressed as equation (2.60).

$$\begin{aligned} Q_n(C_n) &= \prod_{h=1}^{H} \prod_{m_h=1}^{M_h} \left[\Pr\left(\theta_{(m-1)_h} < t_{hn}^* < \theta_{m_h} \right) \right]^{d_{m_{hn}}} \\ &= \prod_{h=1}^{H} \prod_{m_h=1}^{M_h} \left[\Phi\left(\theta_{(m-1)_h} - \alpha_h' Z_{hn} \right) - \Phi\left(\theta_{m_h} - \alpha_h' Z_{hn} \right) \right]^{d_{m_{hn}}} \end{aligned} \quad \cdots\cdots \text{Eq. (2.60)}$$

2.4 Limitations of Previous Literature and Research Motivation

An analysis of consumers' preferences for new technologies using consumers' choice information is significant as it provides various implications in technology management and policy aspects. Previous studies have used two methods to analyze consumers' preferences. One is the stated preference (SP) method which uses the information selected by consumers in a virtual choice situation, and the other is the RP method, which uses the actual purchase data of the consumers in the market. Each methodology has different strengths and weaknesses, and has been independently developed in the field of consumer preference research. Recently, a combined methodology that uses both the methods simultaneously has also been proposed (Brownstone, Bunch, and Train, 2000; Bhat and Castelar, 2002; Börjesson, 2008 ; Axsen, Mountain, and Jaccard, 2009).

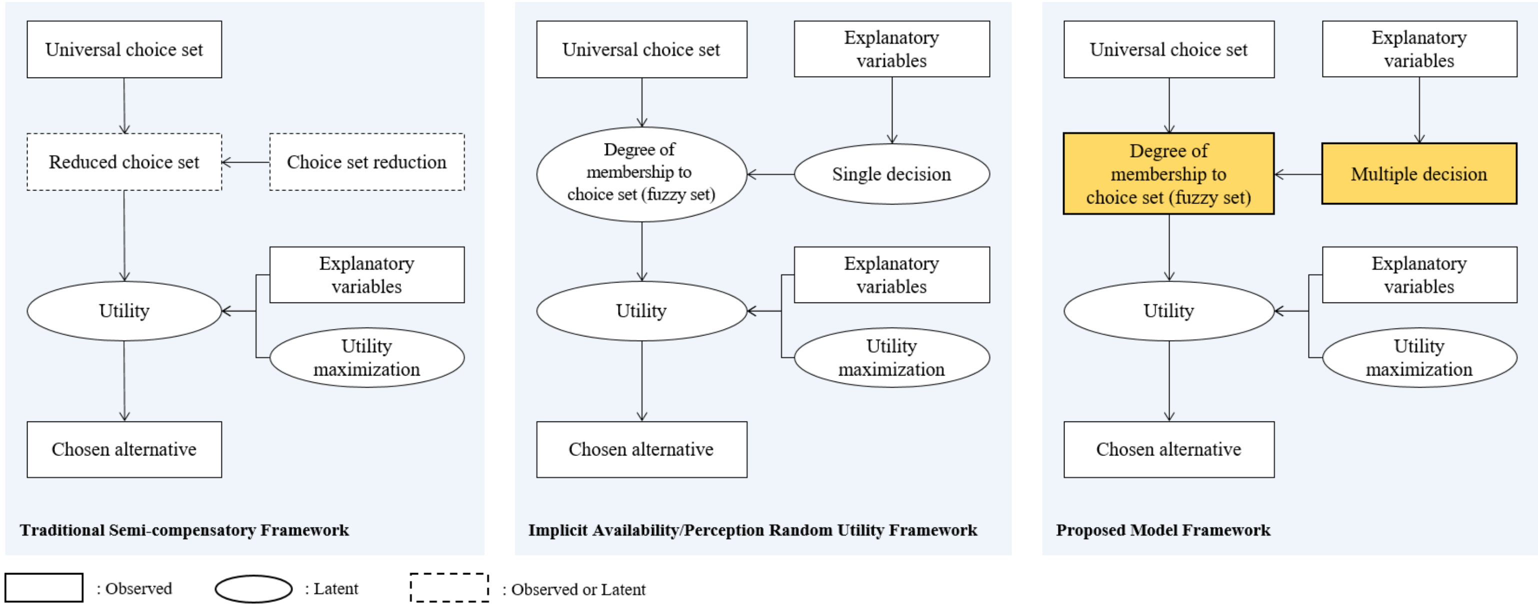
The SP method is useful in analyzing consumers' preferences for new products or services that have not been introduced to the market or have not been diffused sufficiently, making it challenging for researchers to obtain actual choice information. However, it has been pointed out that consumers' choice in the hypothetical choice situation assumed in the SP method is different from consumers' behavior in the real market. Such a limitation serves as a restriction on the application of the SP method (Brownstone, Bunch, and Train, 2000). More specifically, there is a possibility that a consumer makes a choice in a hypothetical situation without fully understanding the attributes of a new product or service, thus the biased analysis result can be obtained. In addition, the consumer may choose a politically correct alternative in the hypothetical case, but may not be willing to pay for the alternative in an actual purchase situation. For example, a consumer who does not intend to pay an additional cost for an environmentally positive product (e.g., electric vehicle or eco-friendly detergent) can respond that they would make the "right choice" in the hypothetical choice situation (Brownstone, Bunch and Train, 2000).

The RP method is primarily used when the data of the actual market purchase of the consumer can be obtained. The advantage of using the RP method is that it analyzes the preference through the actual choice of the consumer. However, considering that the analysis result significantly depends on how the consumers' choice set is formed, the researchers are expected to pay close attention to assuming and constructing the consumers' choice set. As mentioned earlier, since the 1970s, previous studies which incorporate consumers' choice set behaviors into analytical models using RP data has been steadily developed in the field of behavioral economics. However, prvious studies still have two limitations.

First, previous studies may not clearly express how consumers' preference changes or is affected by the consumers' choice set formation behavior. Previous studies have attempted to improve the fitness of a model by modeling and incorporating consumers' choice set formation behavior into the existing discrete choice models (e.g., Swait and Ben-Akiva, 1987; Swait, 2001). Other researches sought to identify consumers' decision criteria in the process of the choice set formation (e.g., Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftn, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). However, these studies seem to fail to explicitly show how consumers' preferences for alternatives can change according to their choice set formation patterns. Considering Cascetta and Papola (2001) and the follow-up studies, all only pointed out that preferences can differ according to the probability that a particular alternative is considered in the choice set. These studies assumed that the consumers' choice set follows the fuzzy set properties and the probability of a particular alternative to be included in the choice set affects the preference of that alternative. In other words, previous studies considered consumers' choice set formation behavior as a single choice decision. In addition, the probability was assumed to be a latent variable, which is a function of individual characteristics. However, considering the fact that the choice set formation behavior has a multiple choice form, the previous studies seem unable to elaborately observe the consumers' choice set formation behavior. Considering that the researchers cannot observe the consumers' actual choice set formation behavior, it appears challenging for them to obtain any realistic results.

Second, the previous studies only identified the choice set formation stage and the decision stage, and did not consider the consumption level of goods. In particular, in the case of goods such as home appliances, analyzing the usage determination behavior as well as the product choice behavior has an important implication in terms of evaluating the economic and environmental effects of consumers' choices.

Therefore, this study proposes a methodology to analyze consumers' decision making process of three stages: choice set formation stage, product choice stage, and usage determination stage. Figure 6 shows the comparison between the previous studies and the proposed model. The proposed model is the expansion of the previous models and is a series of IAPRU framework, which assumes that the probability of the consumers' choice set formation affects the preference of the alternative. However, unlike the previous studies, the proposed model reflects the multiple choice behavior in the choice set formation stage and models the consumers' choice set formation behavior based on the observed data. Utilizing the proposed model could allow researchers to analyze the impact of consumers' choice set formation behavior on consumers' choice of product, usage determination, and its economic and social ripple effects.



Source: Kaplan, Beckhor, and Shiftan (2011)

Figure 6. Proposed model framework versus previous model frameworks

Chapter 3. Methodology

This chapter introduces a new discrete-continuous choice model, which incorporates consumers' choice set formation behavior. The proposed model assumes that choice set formation behavior is an independent step in the decision making process, and affects product choice and usage determination. Section 3.1 introduces the framework of the overall research methodology and suggests the implications of the proposed model. Following this, section 3.2 presents a model for consumers' choice set formation stage and section 3.3 presents a model for the product choice stage in the proposed model. Finally, section 3.4 describes a model for the usage determination stage in the proposed model.

3.1 Methodological Framework

In this dissertation, I propose a new methodology for analyzing the linkage between consumers' choice set formation, product choice, and usage determination in the decision making process. The framework of the overall research methodology is presented in Figure 7.

First, consumers' choice set formation behavior is described by a multivariate probit model considering the characteristics of multiple choice. Further, the product choice behavior is modeled using a multinomial logit model, where it is assumed that the choice set formation behavior affects consumers' utility of alternatives. Using the estimated parameters in the choice set formation stage, the predicted probability that a particular alternative is included in the choice set (i.e., choice set formation probability) is obtained. Next, the choice set formation behavior and product choice behavior is related by constructing the model such that the choice set formation probability modifies the utility of alternatives. Finally, the usage determination behavior that is influenced by consumers' product choice is modeled. This constitutes a three-stage model that continuously represents the choice set formation, product choice, and usage determination behavior.

In the proposed model, consumers' decision making process comprises of a sequence of choice set formation, product choice, and usage determination based on the results of previous studies. As mentioned earlier, previous studies suggest that the consumer follows a semi-compensatory choice process in a complex choice situation, and assume that consumers' choice set formation behavior occurs prior to choosing a product (Payne, Bettman, and Johnson, 1993; Cascetta and Papola, 2001; Cantillo and de Dios Ortúzar, 2005; Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftan, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). In the proposed model, it is assumed that the consumers' choice behavior has a semi-compensatory characteristic, similar to a previous study, and consumers consider some of all the possible alternatives probabilistically within the choice set depending on their recognition and accessibility of the alternatives.

Next, consumers are assumed to choose a product after evaluating the trade-off between the alternatives' attributes in the choice set. Previous studies also argue that there is a dependency between the product choice and usage determination, which cause an endogeneity problem. In order to solve this problem, a previous study suggests that the two stages (product choice and usage determination) should be jointly considered in the model (Greene, 1981; Lee, 1982; 1983; Dubin and McFadden, 1984; Mannering and Winston, 1985; Train, 1986; Goldberg, 1998; West, 2004). One of the methods to solve the endogeneity problem of considering the two stages jointly in the model is by using the estimated choice set formation probability as an instrumental variable in the usage determination stage. Therefore, such a sequential estimation method is employed in the proposed model. However, the fact that product choice and usage determination are sequentially estimated in the model does not indicate that the two behaviors occur separately and consecutively. The proposed model assumed that the product choice and usage determination behavior are simultaneous events, which is similar to previous literatures.

While analyzing consumer preferences, it is possible to use RP or SP data. Despite many advantages of SP data, this dissertation suggests a methodology that utilizes RP data, considering that consumers can make distorted choices in a hypothetical choice situation. This dissertation particularly constructs a model that considers the choice set formation and usage determination behavior. Thus, RP data which can describe consumers' actual behavior in a more detailed manner, is more appropriate for the proposed model.

The proposed methodology has two advantages when compared to the previous studies. First, the proposed model can evidently observe how the choice set formation behaviors affect consumers' preferences when purchasing products and identifying various factors that influence consumers' choice set formation behavior. As mentioned in previous studies, while analyzing consumer preferences using RP data, the result of such analysis can be different depending on how a researcher constructs' consumers' choice set (Swait and Ben-Akiva 1986; Swait, 2001). Previous studies that recognized this problem have proposed a discrete choice model that considers the choice set formation behavior, but they only show the tendency of consumers' choice set formation behavior or their decision criterion. Therefore, this study is meaningful as it not only identifies the factors that affect consumers' choice set formation behaviors, but also evidently shows a change in preference for alternatives by the consumers' choice set formation behavior.

Second, while previous studies are based on a two-stage model for analyzing the choice set formation and product choice behavior (Payne, 1976; Lussier and Olshavsky, 1979; Swait and Ben-Akiva, 1986; 1987; Roberts and Lattin, 1991; Andrews and Srinivasan, 1995; Ben-Akiva and Boccara, 1995; Horowitz and Louviere, 1995; Chiang, Chib, and Narasimhan, 1999; Swait, 2001), the proposed model consists of three stages (choice set formation, product choice, and usage determination) and analyzes how the choice set formation behavior influences consumers' usage determination as well as product choice. In other words, the strength of the proposed model is a sophisticated methodology that can analyze the economic and environmental ripple effects of consumers' behavioral changes in the decision making process by analyzing product choice and usage determination behavior, in addition to the choice set formation behavior.

When the proposed model is empirically applied, the analysis results offer the following management and policy implications. First, it is expected to provide strategic implications for the diffusion of technological innovation by analyzing the effect of consumers' choice set formation behavior on preference for a new product. An example of applying the proposed model to the automotive market is as follows. Suppose that consumers are more likely to purchase eco-friendly vehicles, such as hybrid and electric vehicles, when compared to other alternatives considered in their choice set. In this case, if the consumers do not tend to include eco-friendly vehicles in their choice set since they do not have sufficient knowledge on the policies related to eco-friendly vehicles or environmental advantages of eco-friendly vehicles, this behavior can be a barrier to the diffusion of such innovative products. The proposed model of this dissertation enables finding the factors that influence the diffusion of a new technology in consumers' choice set formation behavior as shown in the example.

Second, by using the proposed model, it is possible to evaluate the economic and environmental ripple effects of the diffusion of new technology considering the various behavior change scenarios of consumers in the decision making process. In other words, the effects of the diffusion of a new technology can be forecasted by observing the effect of changes in consumers' choice set formation behavior, which is affected by individuals' attitude and perception of product choice and usage determination.

Discrete-Continuous Choice Models considering Consumers' Choice Set Formation Behavior

[Stage 1: Choice Set Formation]

Multivariate Probit Model

$$U_{ni}^1 = \delta_i' Z_n^1 + \varepsilon_{ni}^1, \quad \varepsilon_n^1 \sim MVN[0,\Sigma] \quad \longrightarrow \quad y_{ni}^1 = \begin{cases} 1, & \text{if } U_{ni}^1 > 0 \\ 0, & \text{if } U_{ni}^1 \le 0 \end{cases}$$

$$\Rightarrow \quad P_n^1 = \int_{A_1} \cdots \int_{A_J} \phi_J(\varepsilon_n^1 \mid 0,\Sigma)\, d\varepsilon_n^1 \quad \longrightarrow \quad \hat{P}_{ni}^1 = \Pr(y_{ni}^1 = 1 \mid Z_n^1, \hat{\delta}_i') = \Phi(\hat{\delta}_i' Z_n^1)$$

where

$$\phi_J(\varepsilon_n^1 \mid 0,\Sigma) = \frac{1}{(2\pi)^{J/2} \mid \Sigma \mid^{1/2}} e^{-\frac{1}{2}\varepsilon_n^{1\prime}\Sigma^{-1}\varepsilon_n^1} \quad \& \quad A_i = \begin{cases} (-\infty, \delta_i' Z_n^1), & \text{if } y_{ni}^1 = 0 \\ (\delta_i' Z_n^1, \infty), & \text{if } y_{ni}^1 = 1 \end{cases}$$

[Stage 2: Product Choice]

Multinomial Logit Model

$$U_{ni}^2 = \alpha_{ni} + \beta' X_i^2 + \varepsilon_{ni}^2, \quad \varepsilon_{ni}^2 \sim \text{iid extreme value} \quad \longrightarrow \quad y_{ni}^2 = \begin{cases} 1, & \text{if } U_{ni} > U_{nj}, \forall j \ne i \\ 0, & \text{otherwise} \end{cases}$$

where $\alpha_{ni} = \bar{\alpha}_i + \alpha_i^{choiceset} \hat{P}_{ni}^1$

$$\Rightarrow \quad P_{ni}^2 = \frac{\exp(\alpha_{ni} + \beta' X_i^2)}{\sum_{j=1}^J \exp(\alpha_{nj} + \beta' X_j^2)}$$

[Stage 3: Usage Determination]

Linear Model (Dubin and McFadden, 1984)

$$m_n = \sum_{j=1}^J \left[\left(\gamma' X_j^3 + \rho' Z_n^3 \right) \hat{P}_{nj}^2 \right] + \varepsilon_n^3$$

Simulation Analysis Procedure:

1) Construct scenarios that affect choice set formation behavior
2) Predict choice set formation behavior on each scenarios
3) Predict consumers' preference for alternatives considering choice set formation
4) Predict usage determination behavior on each scenarios
5) Analyze economic or environmental ripple effects on each scenario

Expected Results :

- Influence of individual characteristics on choice set formation behavior
- Quantified preference in choosing and consuming products/services
- Relationship between choice set formation behavior and preference for alternatives
- Effect of changes in individual characteristics on choice set formation
- Effect of changes in choice set formation behavior on consumption behavior

Implications :

- Consumers' choice set formation behavior can affect preference and market shares if there are a lot of alternatives in a market
- Consumers' choice set formation behavior should be observed when the impact of technological improvement on market and society is evaluated

Figure 7. Schematic diagram of the proposed methodology of this dissertation

3.2 Model for Choice Set Formation Stage

3.2.1 Overview of the Model

As mentioned earlier, several previous studies stated that consumers in complex market conditions follow a semi-compensatory choice process (Payne, Bettman, and Johnson, 1993; Cascetta and Papola, 2001; Cantillo and de Dios Ortúzar, 2005; Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftan, and Bekhor, 2012; Zolfaghari, Sivakumar and Polak, 2013), where the consumers' purchase process consists of a non-compensatory choice set formation stage and a compensatory product choice stage. More specifically, before choosing a product, a consumer chooses certain alternatives and constructs their choice set based on factors, such as perception and accessibility to alternatives, rather than evaluating all the possible alternatives (non-compensatory process). Subsequently, the consumer evaluates the alternatives in the choice set considering the trade-offs between the attributes of the alternatives and makes a final choice (compensatory process). Such behavior occurs as consumers possess limited information processing capacities (Simon, 1955; Bettman, Luce, and Payne, 1998; Jun, Vogt, and MacKay, 2010; Liu and Dukes, 2016). In this case, there are limited alternatives that can be evaluated by a consumer considering the trade-offs between the attributes.

In the previous studies, consumers' choice set formation behavior has been viewed from two perspectives. The first is that consumer' choice set has the property of a crisp set, which is usually considered in set theory. Considering this point of view, a certain alternative is included or not included in the choice set (e.g., Kaplan, Bekhor, and Shiftan, 2011; Zolfaghari, Sivakumar, and Polak, 2013). Another perspective assumes that the alternatives are considered probabilistically within the choice set and consumers' choice set has the property of a fuzzy set (e.g., Cascetta and Papola, 2001; Chen-Yi, Ke-Ting, and Gwo-Hshiung, 2007).

The proposed model assumes that the consumers' choice set has the property of a fuzzy set and consumers' behavior considering a particular alternative in the choice set is probabilistic (Figure 8). In order to do this, I use a multivariate model to describe the stage of constructing the choice set.

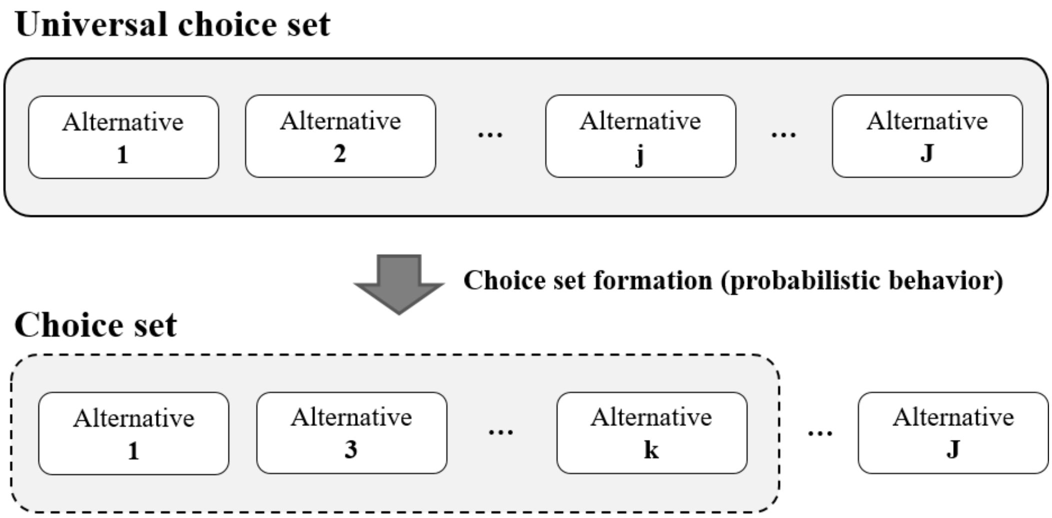


Figure 8. Consumers' choice set formation behavior

It is reasonable to assume that the behavior of consumers deciding to consider a particular alternative in the choice set is related to the behavior for another alternative, rather than being independent. For example, in an automobile purchasing situation, those who prefer gasoline vehicles could more likely consider compact or mid-size gasoline vehicles in the choice set and may not consider mid-size hybrid or electric vehicles in the choice set. In order to reflect these behavioral tendencies, the proposed model uses the multivariate probit model to describe the consumers' choice set formation stage.

3.2.2 Model Specification

First, the utility that consumer n obtains from considering alternative i in the choice set is assumed to be divided into observable term V_{ni}^1 and unobservable term ε_{ni}^1. When the observable term is assumed to be a function of individual characteristic variables Z_n^1, the utility U_{ni}^1 can be expressed as in equation (3.1).

$$U_{ni}^1 = V_{ni}^1 + \varepsilon_{ni}^1 = \delta_i' Z_n^1 + \varepsilon_{ni}^1 \quad \cdots\cdots \text{Eq. (3.1)}$$

Under the multivariate probit model specification for describing the choice set formation, it is assumed that utility U_{ni}^1 is greater than zero if consumer n considers alternative i in the choice set ($y_{ni}^1 = 1$), otherwise ($y_{ni}^1 = 0$) U_{ni}^1 is less than or equal to zero. The relationship between a discrete outcome and a latent variable can be expressed as equation (3.2).

$$y_{ni}^{1} = \begin{cases} 1, \text{ if } U_{ni}^{1} > 0 \\ 0, \text{ if } U_{ni}^{1} \le 0 \end{cases} \quad \cdots\cdots \text{Eq. (3.2)}$$

If the unobservable term $\varepsilon_n^1 = (\varepsilon_{n1}^1, \varepsilon_{n2}^1, ..., \varepsilon_{nJ}^1)$ in equation (3.1) is assumed to follow a multivariate normal distribution with a mean of zero and variance Σ, which is shown as equation (3.3), the probability P_n^1 that consumer n chooses the outcome $y_n^1 = (y_{n1}^1, y_{n2}^1, ..., y_{nJ}^1)$ can be expressed as equation (3.4).

$$\varepsilon_n^1 \sim MVN[0, \Sigma] \quad \cdots\cdots \text{Eq. (3.3)}$$

$$P_n^1 = \int_{A_1} \cdots \int_{A_J} \phi_J(\varepsilon_n^1 \mid 0, \Sigma) d\varepsilon_n^1 \quad \cdots\cdots \text{Eq. (3.4)}$$

where

$$\phi_J(\varepsilon_n^1 \mid 0, \Sigma) = \frac{1}{(2\pi)^{J/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}\varepsilon_n^{1\prime}\Sigma^{-1}\varepsilon_n^1}$$

$$A_i = \begin{cases} (-\infty, \delta_i' Z_n^1), \text{ if } y_{ni}^1 = 0 \\ (\delta_i' Z_n^1, \infty), \quad \text{if } y_{ni}^1 = 1 \end{cases}$$

3.2.3 Estimation Procedure

The likelihood function of the multivariate probit model consists of multiple integrations, which are not represented in a closed form, thus it is challenging to estimate using the classical maximum likelihood estimation method. Therefore, the multivariate probit model is generally estimated using the simulated maximum likelihood estimation method or the Bayesian estimation method. In this dissertation, I use a Bayesian estimation method to estimate a multivariate probit model that expresses consumers' choice set formation behavior (Rossi, Allenby, and McCulloch, 2005; Koop, Poirier, and Tobias, 2007). First, it is assumed that the preference parameter δ and variance Σ follow a normal distribution and inverse Wishart distribution respectively, as shown in equations (3.5) and (3.6).

$$\delta \sim N(\mu_\delta, V_\delta) \quad \text{Eq. (3.5)}$$

$$\Sigma^{-1} \sim Wishart(v, \Omega^{-1}) \quad \text{Eq. (3.6)}$$

Under such a distributional assumption, the parameters can be estimated using the succession of conditional draws as expressed in equation (3.7). The parameters are estimated by drawing from the posterior distribution using the Gibbs sampling method, which is one of the sampling methods based in the MCMC technique.

$$\begin{aligned}&\delta \mid \Sigma^{-1}, W^{*}, y \\ &\Sigma^{-1} \mid W^{*}, \delta, y \\ &W^{*} \mid \delta, \Sigma^{-1}, y\end{aligned} \qquad \text{Eq. (3.7)}$$

where W^{*} and y indicate the latent utility and observed outcome (choice set formation behavior), respectively.

3.3 Model for Product Choice Stage

3.3.1 Overview of the Model

The second stage of the proposed model represents consumers' product choice behavior. The key point of the model for consumers' product choice behavior is to assume that consumers' choice behavior is affected by the choice set formation behavior. Figure 9 shows the relationship between the choice set formation and product choice behavior, which is assumed in the proposed model.

In the proposed model, the alternative specific constant α_{nk} of the alternative k for consumer n is assumed to consist of $\bar{\alpha}_k$ and $\alpha_k^{choiceset}\hat{P}_{nk}^{1}$. The first term $\bar{\alpha}_k$ is a constant term which is not different across individuals and the second term $\alpha_k^{choiceset}\hat{P}_{nk}^{1}$ is a term which is affected by the estimated choice set formation probability in the choice set formation stage. Here, the estimated probability for consumer n to include alternative k within the choice set can be expressed as equation (3.8).

$$\hat{P}_{nk}^{1} = \Pr(y_{nk}^{1} = 1 \mid Z_{n}^{1}, \hat{\delta}_{k}') = \Phi(\hat{\delta}_{k}' Z_{n}^{1}) \quad \text{Eq. (3.8)}$$

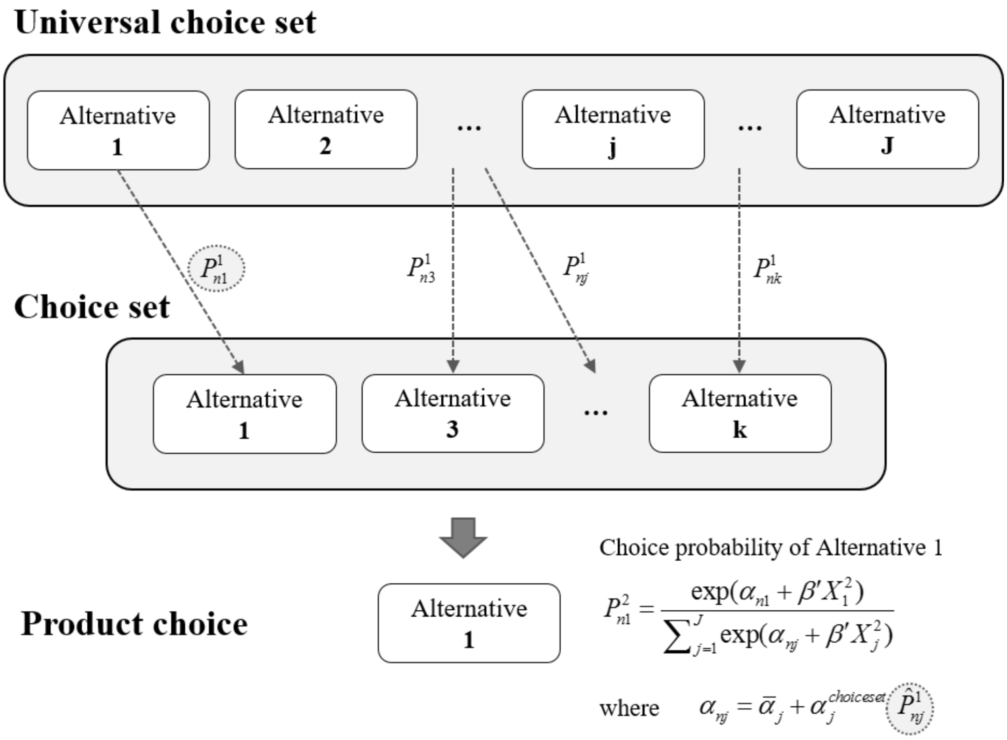


Figure 9. Relationship between choice set formation and choice

3.3.2 Model Specification

As mentioned earlier, the proposed model assumes that the consumers' choice set formation behavior plays a role in adjusting the alternative specific constant of each alternative in the product choice stage. In order to reflect this assumption, consumers' product choice behavior is expressed using a multinomial logit model. In the product choice stage, the utility that consumer n obtains from choosing alternative i can be expressed as equation (3.9). In addition, the unobservable term ε_{ni}^2 is assumed to follow type I extreme value distribution as shown in equation (3.10).

$$U_{ni}^2 = V_{ni}^2 + \varepsilon_{ni}^2 = \alpha_{ni} + \beta' X_i^2 + \varepsilon_{ni}^2 \quad \text{Eq. (3.9)}$$

$$f(\varepsilon_{ni}^2) = e^{-\varepsilon_{ni}^2} e^{-e^{-\varepsilon_{ni}^2}} \quad \text{Eq. (3.10)}$$

In order to express that consumers' choice set formation behavior affects the utility of alternatives at the product choice stage, the alternative specific constant of alternative i for consumer n is assumed to consist of $\bar{\alpha}_i$ and $\alpha_i^{choiceset} \hat{P}_{ni}^1$, which are a constant term and a term affected by estimated choice set formation probability $\hat{P}_{n1}^1$ respectively, as shown in equation (3.11).

$$\alpha_{ni} = \bar{\alpha}_i + \alpha_i^{choiceset} \hat{P}_{ni}^1 \quad \text{Eq. (3.11)}$$

In equation (3.11), $\alpha_i^{choiceset}$ indicates the parameter that represents how much the alternative specific constant of alternative i can be changed as the choice set formation probability $\hat{P}_{ni}^{1}$ changes by one unit. In other words, it can be observed how a consumer's preference for alternatives can change depending on the tendency to include alternative i in the choice set.

Based on these model specifications, the choice probability of alternative i for consumer n in the product choice stage can be expressed as equation (3.12).

$$\begin{aligned} P_{ni}^{2} &= P(U_{ni}^{2} > U_{nj}^{2},\ \forall j \neq i) \\ &= P(V_{ni}^{2} + \varepsilon_{ni}^{2} > V_{nj}^{2} + \varepsilon_{nj}^{2},\ \forall j \neq i) \\ &= P(\varepsilon_{ni}^{2} > \varepsilon_{nj}^{2} + V_{nj}^{2} - V_{ni}^{2},\ \forall j \neq i) \end{aligned} \quad \cdots\cdots \text{Eq. (3.12)}$$

Considering the distributional assumption of unobservable term ε_{ni}^{2}, the choice probability can be rewritten as equations (3.13) and (3.14).

$$P_{ni}^{2} = \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{ni}^{2} + V_{ni}^{2} - V_{nj}^{2})}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \quad \cdots\cdots \text{Eq. (3.13)}$$

$$P_{ni}^{2} = \frac{\exp(V_{ni}^{2})}{\sum_{j=1}^{J} \exp(V_{nj}^{2})} = \frac{\exp(\alpha_{ni} + \beta' X_{i}^{2})}{\sum_{j=1}^{J} \exp(\alpha_{nj} + \beta' X_{j}^{2})} \quad \cdots\cdots \text{Eq. (3.14)}$$

If it is assumed that $y_{ni} = 1$ when consumer n chooses alternative i and

$y_{ni} = 0$ when consumer n does not choose alternative i, the likelihood of consumer n can be expressed as equation (3.16).

$$P_n^2 = \prod_{i=1}^{J} \left(P_{ni}^2\right)^{y_{ni}} = \prod_{i=1}^{J} \left(\frac{\exp(\alpha_{ni} + \beta' X_i^2)}{\sum_{j=1}^{J} \exp(\alpha_{nj} + \beta' X_j^2)} \right)^{y_{ni}} \quad \text{Eq. (3.15)}$$

$$Likelihood = \prod_{n=1}^{N} P_n^2 = \prod_{n=1}^{N} \prod_{i=1}^{J} \left(P_{ni}^2\right)^{y_{ni}} \quad \text{Eq. (3.16)}$$

Based on the multinomial logit framework explained above, the relationship between the choice set formation and product choice behavior can be identified, where the extent to which the choice set formation behavior affects consumers' preferences in the product choice stage can be analyzed.

3.3.3 Estimation Procedure

The likelihood function of a multinomial logit model is expressed in a closed form. Therefore, it can be easily estimated using the classical maximum likelihood estimation method. The proposed model estimates the preference parameters $(\bar{\alpha}_{i|q}, \alpha_{i|q}^{choiceset}, \beta_q)$ in the product choice stage using the classical maximum likelihood estimation method, which estimates the parameters maximizing the likelihood.

3.4 Model for Usage Determination Stage

3.4.1 Overview of the Model

The last stage of the proposed model is a usage determination stage for the alternative which was chosen in the product choice stage. While choosing a product or service that consumes energy (e.g., appliance, vehicle, or transit), a consumer generally makes a discrete choice of what to buy and a continuous choice of how much to use the product. As mentioned earlier, the proposed model aims to analyze how such discrete-continuous choice behavior is affected by the choice set formation. In order to do this, the linkage between the product choice and usage determination is modeled in the usage determination stage, and the proposed model of the three stages is ultimately constructed.

As mentioned in chapter 2, there are sequential estimation technique-based models (e.g., Heckman, 1976; 1978; Hausman, 1979; Greene, 1981; Lee, 1982; 1983; Dubin and McFadden, 1984; Mannering and Winston, 1985; Train, 1986; Goldberg, 1998; West, 2004) and MDCEV-based models (e.g., Bhat, 2005; 2008; Pinjari and Bhat, 2010; Bhat, Castro, and Khan, 2013) to represent consumers' discrete-continuous behavior. The proposed model follows the sequential estimation technique-based model framework proposed by Dubin and McFadden (1984) to evidently observe the influence of choice set formation behavior[5].

---

[5] The details of the methodologies proposed by the previous studies are reviewed in chapter 2.

3.4.2 Model specification

In order to link the product choice and usage determination stages, the predicted choice probability $\hat{P}_{ni}^{2}$ obtained by using the estimated parameters in the product choice stage is utilized as an instrumental variable (ψ_{ni}) in the usage determination stage, where the usage (m_n) of a consumer n is expressed as a function of $\hat{P}_{ni}^{2}$ as equation (3.17).

$$\begin{aligned} m_n &= \sum_{j=1}^{J}\left[\left(\gamma' X_j^3 + \rho' Z_n^3\right)\psi_{nj}\right] + \varepsilon_n^3 \\ &= \sum_{j=1}^{J}\left[\left(\gamma' X_j^3 + \rho' Z_n^3\right)\hat{P}_{nj}^{2}\right] + \varepsilon_n^3 \end{aligned} \qquad \text{Eq. (3.17)}$$

In this manner, the following two effects can be obtained by using $\hat{P}_{ni}^{2}$ as an instrumental variable in the equation for expressing consumers' usage for the product in the usage determination stage. First, the use of an instrumental variable solves the endogeneity problem that exists in the model for consumers' usage determination behavior and provides an unbiased estimator. In general, there is a correlation between the unobserved factors that affect product choice and usage determination (Dubin and McFadden, 1984). Without taking this into account, estimating the two stages independently yields biased estimators. Utilizing the choice probabilities as instrumental variables in the usage determination stage play a role in solving the endogeneity problem. Second, by using an instrumental variable and linking the product choice stage with the usage determination stage, the manner in which the choice set formation behavior affects the usage determination behavior can be observed. This is significant as it constitutes a methodology for analyzing the economic and environmental ripple effects of consumers' behavior changes.

3.4.3 Estimation Procedure

The proposed model considers only the error between the observations as it uses a pooled cross-sectional data that does not consider time dimension. In addition, the proposed model assumes that each observation is independent of each other. Accordingly, equation (3.17), which represents consumers' usage determination behavior, is estimated by using an ordinary least square estimation method.

Chapter 4. Empirical Study: Consumers' Adoption Behavior for New Products in the Automobile Market

4.1 Overview of the Empirical Study

4.1.1 Research Background

An automobile market is a typical market where consumers tend to simplify the decision making process as there are several products that consumers can purchase (Shocker, Ben-Akiva, Boccara, and Nedungadi, 1991; Lapersonne, Laurent, and Le Goff, 1995; Punj and Brookes, 2002; Chen-Yi, Ke-Ting, and Gwo-Hshiung, 2007). Considering the fact that there are numerous vehicle models in the market, and that several automobile attributes need to be compared and evaluated, consumers with limited information processing capacity find it virtually impossible to evaluate all the possible alternatives. Therefore, a consumer generally constructs a choice set with certain alternatives, evaluates attributes of the alternatives within the choice set, and makes a product choice. Considering this point, the automobile market is suitable to apply the proposed model in this dissertation.

In this chapter, the proposed model is applied to the automobile market and the following empirical studies are conducted. First, I emphasize the importance of considering consumers' actual choice sets by showing how the analysis results can differ depending on the assumption of consumers’ choice sets. Further, I apply the three-stage choice model to the automobile market, considering the choice set formation behavior to observe how the changes in the choice set formation stage affect consumers’ choice behavior for a new automobile. In addition, the changes in the automobile market and its ripple effect are analyzed when automobile technologies are advanced and consumers’ choice set formation behavior changes.

4.1.2 Structure of the Empirical Study

Previous studies have highlighted that consumers’ preferences cannot be analyzed accurately if consumers’ choice sets are misconfigured when researchers analyze the consumers’ preferences using RP data (Swait and Ben-Akiva, 1986; Swait, 2001; Kaplan, Shiftan, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). Accordingly, an attempt was made to consider consumers’ choice set formation behavior in the discrete choice models.

In this chapter, I empirically examine how the results of consumers’ preference analysis vary according to the assumption of consumers’ choice sets. Further, I investigate the effect of the choice set formation behavior on consumers' adoption of new technology by applying the proposed model. In other words, the empirical study analyzes the pattern of the choice set formation behavior according to individual characteristics in purchasing a new automobile, and identifies the effect of including a specific vehicle type within the choice set on the preference and usage of the vehicle. It also analyzes the ripple effects of technological improvements and consumers' awareness changes. Ultimately, the empirical study aims to identify the factors that influence the diffusion of new technology in consumers' decision making process.

The structure of the empirical study is shown in Figure 10. First, empirical study 1 observes how the results of consumers' preference analysis vary according to the manner in which a researcher reconstructs the consumers' choice set, while analyzing the preferences using RP data related to automobile purchases. Swait and Ben-Akiva (1986) noted that if a researcher misconfigures the consumers' choice set, the following two problems could arise: (1) The alternative specific constants are downward biased for the alternatives which tend to be considered in the choice set alone, not with other alternatives. (2) By assuming that the alternatives that were not actually considered are included in the choice set, the marginal utility for the product attributes is distorted. Empirical study 1 empirically confirms the theoretical arguments of previous studies for the automobile market. Considering this purpose, the following four cases are taken into account for the choice set assumption and the results of the preference analyses are compared:

1) Choice set 1: A set of all possible alternatives (universal choice set).

2) Choice set 2: A set of alternatives that are randomly extracted.
3) Choice set 3: A set of alternatives that have an engine displacement size similar to the automobile purchased.
4) Choice set 4: A set of alternatives that are actually considered by a consumer in a choice situation.

In empirical study 2, the proposed model of this dissertation is applied to the automobile market, in order to analyze the relationship between the consumers' choice set formation, product choice, and usage determination behavior. More specifically, I investigate the consumers' choice set formation behavior according to individual characteristics in purchasing a new automobile, and analyze how the choice set formation behavior affects consumers' decision of vehicle choice and mileage. The choice set formation, product choice, and usage determination stages are estimated by applying a multivariate probit model, a multinomial logit model, and a linear model, respectively, and a discussion of the results of each stage is included.

Finally, in empirical study 3, I construct scenarios consisting of changes in consumer perceptions affecting the choice set formation and changes in automobile performance due to technological improvements. Further, I forecast the changes in the automobile market for each of the scenarios and examine the environmental impacts of the changes.

Structure of the Empirical Study

Observing Consumers' Preference based on the Different Types of Reconstructed Choice Set	Identifying a Relation between the Choice Set Formation, Product Choice, and Usage Determination Behavior	*Ex-ante* Impact Evaluation of Consumers' Perception Change and Technological Improvement on Automobile Market and Environment
Contents : • Using RP data, I analyzed consumers' preference for adopting a new automobile by assuming the four types of reconstructed choice set ✓ **Choice set 1**: A set of all possible alternatives (universal choice set) ✓ **Choice set 2**: A set of alternatives that are randomly extracted ✓ **Choice set 3**: A set of alternatives that have an engine displacement size similar to the automobile purchased ✓ **Choice set 4**: A set of alternatives that are actually considered by a consumer in a choice situation	**Contents :** • I applied the model that consists of three stages to automobile market, and analyzed general behavioral connection between choice set formation, product choice and usage determination • The study identified consumers' choice set formation behavior and factors that affect to choice set formation	**Contents :** • I constructed four scenarios with consumers' perception change and technological improvement in automobile industry ✓ **Perception** 1) knowledge of energy policy 2) knowledge of externality ✓ **Technological improvement** 1) Fuel efficiency of gasoline vehicles 2) Sales price of hybrid/electric vehicles • Market change and its ripple effect were analyzed in this study
Expected results : • The study shows a different results by choice set assumption in revealed preference method • The theoretical findings of Swait and Ben-Akiva (1986) can be empirically demonstrated in this study	**Expected results :** • The influence of choice set formation behavior to following decision stages (product choice and usage determination) can be observed in this study	**Expected results :** • The influence of consumers' perception change and technological improvement on market and environment can be observed in this study

Figure 10. Structure of the empirical study

4.1.3 Data

4.1.3.1 Overview of the Survey

Data used in the empirical studies were collected through an online survey. The survey was conducted by a professional survey company (Gallup Korea) from April 21 to 28, 2017. The respondents included 809 adults aged between 20 and 59 years located in Korea. A sample was drawn using the purposive quota-sampling method on the basis of the respondents' age, gender, and geographical region, in order to maintain a representation of the actual population. Table 1 summarizes the demographic characteristics of the respondents.

First, the respondents were questioned on their automobile purchase history, including the vehicle that they currently owned. In addition, the survey collected information on the two vehicle models that were the potential alternatives, but were not actually purchased when the respondent purchased a new automobile. Considering this information, it is possible to identify the consumers' choice set in their choice situations[6]. As described in chapter 3, the proposed model assumes that the consumers' choice set has the characteristic of a fuzzy set. Assuming that the consumers' choice set is a fuzzy set, consumers' choice set formation behavior probabilistically considers a particular vehicle type within the choice set. In this case, although a researcher may not fully observe the consumers' choice set

[6] In the questionnaire, the respondents were questioned on the two vehicle models that were most likely to be purchased, except for the vehicle that they actually purchased.

formation behavior, their behavior where they include a particular vehicle type in the choice set can be analyzed by identifying the types of vehicles that are mainly considered in the consumers' choice set. Therefore, it is expected that a meaningful result can be obtained through the questionnaire used in this empirical study.

Table 1. Demographic characteristics of the respondents

Category		**Samples (%)**
Total		809 (100.0)
Gender	Male	415 (51.3)
	Female	394 (48.7)
Age	20 – 29	171 (21.1)
	30 – 39	244 (30.2)
	40 – 49	243 (30.0)
	50 - 59	151 (18.7)
Geographic region	Captical area	465 (57.5)
	Chungcheong area	69 (8.5)
	Gyeongsang area	196 (24.2)
	Jeolla area	56 (6.9)
	Gangwon area	19 (2.3)
	Jeju area	4 (0.5)
Household monthly income	Less than KRW 3 million	169 (20.9)
	KRW 3 million – 4 million	144 (17.8)
	KRW 4 million – 5 million	157 (19.4)
	KRW 5 million – 7 million	180 (22.2)
	More than KRW 7 million	159 (19.7)

In addition, the respondents were questioned on their demographics (age, gender, household income, etc.), lifestyle, knowledge of energy policies related to the transport sector, and perception of eco-friendly vehicles.

In the empirical study, some outliers were excluded from the raw data obtained through the survey to ensure reliability of the estimation results. First, considering that the responses depend on the memory of the respondents, purchase data subsequent to 2010 were used in the analysis to ensure the accuracy of the response. Second, only the data of purchasing a new automobile was used in the analysis. In the case of buying a used car, the purchase price differs based on the purchase method, model year, and vehicle condition. Therefore, considering that it is challenging to identify the types of vehicles a consumer considers when a consumer purchases a used car, the cases where a used car was purchased were excluded from the analysis. Third, the cases where a foreign car was either considered within the consumers' choice set or actually purchased by the consumer, were excluded from the analysis. It is challenging to obtain meaningful results when foreign cars are included in the analysis data as there is a large difference in the performance and price based on the manufacturer as well as a significant gap in the sales price between domestic and foreign cars. It is expected that the representativeness of the analysis results will be retained if only domestic cars were included in the analysis as there are not many cases where a consumer considered a foreign car within the choice set or purchased a foreign car[7]. According to the Ministry of Land, Infrastructure and Transport (2017), among the new passenger cars registered between 2010 and 2016, the proportion of foreign cars was approximately 13.0%. Therefore, the analysis results of consumer preference and ripple effects from technological improvements, and perception changes in this empirical study can be observed as a result of considering only 87.0% of the entire market. Fourth, liquefied petroleum gas (LPG) vehicles were excluded from the analysis data. It is not appropriate to include LPG vehicles in analyzing general consumers' preferences as only special customers, such as disabled individuals, taxi drives, and car rental companies, can purchase LPG vehicles in Korea. Finally, 411 samples of 337 respondents were used in the analysis.

4.1.3.2 Automobile specification data

In the empirical study, consumers' preferences were analyzed using RP data. Therefore, in order to reconstruct the consumers' choice set, information of the automobiles in the market at the time of purchasing a new automobile by the consumer is required. Accordingly, I collected information of the automobiles sold by domestic manufacturers (Hyundai, Kia, Renault Samsung, Chevrolet, and Ssangyong) in Korea since 2010. First, the automobiles are classified into 10

[7] Among 701 samples of 499 respondents who purchased a new automobile after 2010, only 11.1% purchased a foreign car.

categories based on their engine size (sports utility vehicle (SUV) is classified as a distinct type) and fuel type as shown in Table 2. The specifications of the automobiles, which are sold from 2010 to 2017 in the Korean automobile market, such as the manufacturer, fuel type, purchase price, and fuel efficiency were collected from an online automobile information database[8].

Table 2. Vehicle categories

Engine size	Fuel type	Examples
Economy /Subcompact	Gasoline	Hyundai Accent, Kia Morning, Kia Pride, Chevrolet Spark
Compact	Gasoline	Hyundai Avante, Kia K3, Renault Samsung SM3, Chevrolet Cruze
	Diesel	Hyundai Avante, Kia K3, Renault Samsung SM3, Chevrolet Cruze
	Hybrid/electric	Hyundai Ionic, Kia Soul EV, Renault Samsung SM3
Mid-size	Gasoline	Hyundai Sonata, Kia K5, Renault Samsung SM5, Chevrolet Malibu
	Diesel	Hyundai Sonata, Kia K5, Renault Samsung SM5, Chevrolet Malibu
	Hybrid/electric	Hyundai Sonata, Kia K5
Full-size	Gasoline	Hyundai Grandeur, Hyundai Genesis, Kia K7, Kia K9, Renault Samsung SM7, Chevrolet Alpheon, SsangYong Chairman
SUV	Gasoline	Hyundai Tucson, Kia Sportage, Kia Sorento, Renault Samsung QM5, Chevrolet Trax, SsangYong Tivoli
	Diesel	Hyundai Tucson, Hyundai SantaFe, Kia Sportage, Kia Sorento, Renault Samsung QM3, Renault Samsung QM6, Chevrolet Trax, SsangYong Tivoli

[8] Daum Automobile (auto.daum.net) and Naver Automobile (auto.naver.com) offer the specifications of the domestic automobiles in Korea.

4.1.3.3 Individual characteristic data

In order to analyze the effect of individual characteristics on consumers' behavior, individual characteristics are collected using a survey. The individual characteristics include demographics, such as age, gender, household income, and the number of family members, and general knowledge of the transport sector, such as the knowledge of environmental externality of the sector. In addition, the perception of eco-friendly vehicles and fuel type of the vehicle that a respondent owned in the past are also considered as individual characteristics. The individual characteristics used in the empirical study are shown in Table 3.

Table 3. Individual characteristics data

Type	Variable	Question
Demographics	Age	Age
	Gender	Gender
	Household monthly income	Household monthly income (seven-point likert scale)
	The number of family member	The number of family member
General knowledge of transport sector	Knowledge of environmental externalities in transport sector	How much do you know about environmental externality caused by driving a car? (five-point likert scale)
	Knowledge of energy policies in transport sector	How much do you know about energy policies (corporate average fuel economy, energy labelin, etc.) in the transport sector? (five-point likert scale)
Perception about eco-friendly vehicles	Perception about environmental positive effect of eco-friendly vehicles	Do you expect that the spread of eco-friendly vehicles will have a positive impact on energy saving and environmental improvement? (five-point likert scale)
Information of a vehicle owned in the past	Fuel type of the vehicle previously owned	What is fuel type of the vehicle you have previously owned?

4.2 Observing Consumers' Preferences based on the Different Types of Reconstructed Choice Set

4.2.1 Introduction

In this section, I examine how the analysis results can vary based on the choice set assumptions when a researcher analyzes consumers' preferences by using RP data. In order to realize this, I assume four types of choice sets as shown in Table 4 and analyze consumers' preferences accordingly.

Table 4. Different types of reconstructed choice sets

Type	Assumption
Choice set 1	A set of all possible alternatives (universal choice set)
Choice set 2	A set of alternatives that are randomly drawn
Choice set 3	A set of alternatives that have similar engine displacement size to the automobile purchased
Choice set 4	A set of alternatives that are actually considered by a consumer in a choice situation

Choice set 1: A set of all possible alternatives (universal choice set)

Choice set 1 is the most common assumption, which assumes that a consumer considers all the available automobile alternatives. All the 10 vehicle types introduced in Table 2 are included in choice set 1, and vehicle models manufactured by different manufacturers were considered separately for each vehicle type as shown in Table 5 (one representative vehicle was included as an alternative for each manufacturer).

Choice set 2: A set of alternatives that are randomly drawn

It is also possible to reconstruct a consumers' choice set by randomly drawing alternatives from the universal choice set. In particular, previous studies found that when analyzing consumers' preferences by using a model that has independence from irrelevant alternatives (IIA) property, such as a multinomial logit model, consistent estimation can be obtained if the alternatives are randomly drawn from the universal choice set and included in a consumers' choice set (Brownstone, Bunch, and Train, 2000). In order to construct a consumers' choice set by randomly drawing alternatives from the universal choice set, a random variable d that follows a uniform distribution over the interval [0.5, 10.5] is employed as shown in equation (4.1). This is a method to construct a choice set by randomly drawing vehicle alternatives from 10 vehicle categories defined above.

$$d \sim uniform(0.5, 10.5) \quad \text{Eq. (4.1)}$$

The values drawn randomly from the uniform distribution is rounded off to obtain integers from 1 to 10, and the draw is repeated until two different types, out of the nine types excluding the vehicle type actually purchased, are obtained.

Table 5. Information of choice set 1 (universal choice set)

Year	Economy /subcompact	Compact			Mid-size			Full-size	SUV		Total
	Gasoline	Gasoline	Diesel	Hybrid /Electric	Gasoline	Diesel	Hybrid /Electric	Gasoline	Gasoline	Diesel	
2010	H,K,C	H,K,R,C	H,K,C	K	H,K,R,C	H,C	H,K	H,K,R,C,S	H,K,R,C	H,K,R,C,S	33
2011	H,K,C	H,K,R,C	H,K,C	K	H,K,R,C	H,K,C	H,K	H,K,R,C,S	H,K,R	H,K,R,C,S	33
2012	H,K,C	H,K,R,C	H,K,C	K	H,K,R,C	H,R,C	H,K	H,K,R,C,S	H,K,R,C	H,K,R,C,S	34
2013	H,K,C	H,K,R,C	H,K,C	H	H,K,R,C	H,R,	H,K	H,K,R,C,S	H,K,R,C	H,K,R,C,S	33
2014	H,K,C	H,K,R,C	H,K,R,C	H,K,R,	H,K,R,C	H,R,C	H,K	H,K,R,C,S	H,K,R,C	H,K,R,C,S	37
2015	H,K,C	H,K,R,C	H,K,R,C	H,K,R	H,K,R,C	H,K,R,C	H,K,	H,K,R,C,S	H,K,R,C,S	H,K,R,C,S	39
2016	H,K,C	H,K,R,C	H,K,R,C	H,K,R,	H,K,R,C	H,K,R,C	H,K	H,K,R,C,S	H,K,R,C,S	H,K,R,C,S	39
2017	H,K,C	H,K,R,C	H,K,R,C	H,K,R	H,K,R,C	H,K,R,C	H,K,C	H,K,R,C,S	H,K,R,C,S	H,K,R,C,S	40

Note: H = Hyundai, K = KIA, R = Renault Samsung, C = Chevrolet, S = SsangYong.

This allows three different vehicle types to be randomly included in each consumer's choice set. Table 6 shows the result of the randomly reconstructed choice set (choice set 2) for 411 observations.

Table 6. Information of choice set 2

No.	Size	Fuel type	Automobile owend	Alternatives in choice set (randomly sampled)
1	Economy /Subcompact	Gasoline	64 (15.6%)	69 (8.4%)
2	Compact	Gasoline	96 (23.4%)	83 (10.1%)
3	Compact	Diesel	10 (2.4%)	85 (10.3%)
4	Compact	Hybrid/Electric	4 (1.0%)	90 (10.9%)
5	Mid-size	Gasoline	87 (21.2%)	76 (9.2%)
6	Mid-size	Diesel	10 (2.4%)	81 (9.9%)
7	Mid-size	Hybrid/Electric	10 (2.4%)	102 (12.4%)
8	Full-size	Gasoline	42 (10.2%)	86 (10.5%)
9	SUV	Gasoline	8 (1.9%)	87 (10.6%)
10	SUV	Diesel	80 (19.5%)	63 (7.7%)
Total			411 (100.0%)	822 (100.0%)

Choice set 3: A set of alternatives that have an engine displacement size similar to the automobile purchased

Choice set 3 assumes that the consumers include alternatives within the choice set by applying a certain consistent criteria. Here, it is assumed that a consumer constructs a choice set based on the engine displacement size, as assumed by Woo, Moon, Lee, Jo, and Lee (2017), where a consumer includes alternatives that have similar engine displacement size within the choice set and choose the most preferred one in the choice set.

Choice set 3 consists of alternatives that have an engine displacement size similar to the vehicle that the consumer actually purchased, as shown in Table 7. In this case, it is assumed that the consumers who purchased an SUV in Korea, considered compact and mid-size sedans within their choice set as SUVs have engine displacement sizes similar to those of compact and mid-size sedans.

Table 7. Information of choice set 3

Automobile owned	Reconstructed choice set
Economy/subcompact	Economy/subcompact, Compact
Compact	Compact, Mid-size, SUV
Mid-size	Compact, Mid-size, SUV
Full-size	Mid-size, Full-size
SUV	Compact, Mid-size, SUV

Choice set 4: A set of alternatives that are actually considered by a consumer in a choice situation

Choice set 4 consists of alternatives that the consumers actually considered while purchasing a car. A survey was used to identify the alternatives that consumers actually considered in their choice set. In other words, the consumers were questioned on which vehicle models were considered with the purchased model. Table 8 shows the vehicle types considered by the consumers who purchased a certain type of vehicle.

Table 8. Information of choice set 4

		Alternatives in choice set					
		Economy /subcompact	Compact	Mid-size	Full-size	SUV	Total
Automobile purchased	Economy /subcompact	**78 (60.9%)**	40 (31.3%)	5 (3.9%)	1 (0.8%)	4 (3.1%)	128 (100.0%)
	Compact	19 (8.6%)	**132 (60.0%)**	39 (17.7%)	7 (3.2%)	23 (10.5%)	220 (100.0%)
	Mid-size	0 (0.0%)	29 (13.6%)	**127 (59.3%)**	18 (8.4%)	40 (18.7%)	214 (100.0%)
	Full-size	2 (2.4%)	3 (3.6%)	19 (22.6%)	**50 (59.5%)**	10 (11.9%)	84 (100.0%)
	SUV	3 (1.7%)	15 (8.5%)	22 (12.5%)	8 (4.5%)	**128 (72.7%)**	176 (100.0%)

The consumers' choice set formation behaviors represented in Table 8 show that consumers generally consider alternatives that have engine displacement sizes similar to the vehicle purchased within the choice set. More specifically, consumers purchasing an economy/subcompact sedan tend to include economy/subcompact and compact sedans in their choice set, and the consumers purchasing a compact sedan tend to consider compact and mid-size sedans, and SUVs within their choice set. In addition, the consumers who purchase a mid-size sedan tend to include compact and mid-size sedans, and SUVs in their choice set, and consumers purchasing a full-size sedan generally include mid-size and full-size sedans, and SUVs within their choice set. Lastly, consumers who purchase a SUV tend to consider other SUVs in their choice set rather than other types of vehicles.

The significant observation is that the probability to be included in the choice set (choice set formation probability) of each vehicle type varies based on the type of vehicle a consumer purchases. Considering economy/subcompact and full-size sedans, the probabilities to be included in a consumer's choice set are relatively low (3.2% and 4.2% on an average, respectively) when a consumer purchases other types of vehicles other than an economy/subcompact or full-size sedan. On the other hand, the choice set formation probabilities of compact and mid-size sedans are 14.2% on an average, which is a relatively high probability when a consumer purchases other types of vehicles other than compact and mid-size sedans[9].

[9] Considering SUVs, the probability to be included in a consumer's choice set is 11.0% on an average when a consumer purchases other types of vehicles other than SUV.

4.2.2 Estimation Results and Discussion

Table 9 shows the estimation results of consumers' utility function for purchasing a new automobile, represented in equation (4.2), by applying choice sets 1 to 4.

$$U_{nj} = \alpha_j + \beta_{salesprice} X_{nj,salesprice} + \beta_{fuel\,efficiency} X_{nj,\,fuel\,efficiency} + \varepsilon_{nj} \quad \cdots\cdots\cdots\cdots\cdots \text{Eq. (4.2)}$$

where $X_{nj,salesprice}$ is the sales price of a new automobile (10 million KRW[10]) and $X_{nj,\,fuelefficiency}$ is the fuel efficiency (kilometer (km)/1,000 KRW) which indicates the distance that can be traveled with 1,000 KRW.

In this empirical study, alternative specific constants, which represent the average effect of unobserved factors on the utility of each alternative, indicates how much the alternative is preferred in relation to a diesel SUV, regardless of its observed factors (attributes). The engine displacement is also a significant attribute of an automobile, but it is excluded in the consumer preference analysis to avoid a multicollinearity problem. The engine displacements of vehicles sold in Korea between 2010 and 2017 is significantly correlated with the sales prices[11].

[10] "KRW" refers to the South Korean won. According to the Economic Statistics System (ECOS) of the Bank of Korea, USD 1 = KRW 1133.57 as of April 2017.

[11] The correlation between the sales price and engine displacement of 288 vehicles sold in Korea between 2010 and 2017 was approximately 80.97%.

Table 9. Estimation results for different types of choice sets

Attribute	Choice set 1			Choice set 2			Choice set 3			Choice set 4		
	Estimates	S.E	P>|z|	Estimates	S.E	P>|z|	Estimates	S.E	P>|z|	Estimates	S.E	P>|z|
ASC1	**-0.790**	0.331	0.017	-0.586	0.366	0.110	18.143	1044.112	0.986	**1.576**	0.450	0.000
ASC2	-0.048	0.267	0.856	-0.164	0.306	0.593	0.197	0.265	0.457	**0.929**	0.338	0.006
ASC3	**-3.793**	0.411	0.000	**-3.740**	0.456	0.000	**-3.779**	0.415	0.000	**-2.658**	0.448	0.000
ASC4	**-4.191**	0.653	0.000	**-4.631**	0.741	0.000	**-4.663**	0.691	0.000	**-3.523**	0.700	0.000
ASC5	**1.021**	0.232	0.000	**1.001**	0.284	0.000	**1.174**	0.232	0.000	**1.369**	0.315	0.000
ASC6	**-2.644**	0.359	0.000	**-2.569**	0.408	0.000	**-2.677**	0.361	0.000	**-2.066**	0.405	0.000
ASC7	**-2.108**	0.361	0.000	**-2.326**	0.394	0.000	**-2.328**	0.371	0.000	**-1.726**	0.411	0.000
ASC8	**1.733**	0.317	0.000	**1.700**	0.369	0.000	20.584	1241.924	0.987	**2.313**	0.435	0.000
ASC9	**-1.169**	0.429	0.006	**-1.119**	0.468	0.017	**-0.904**	0.431	0.036	**-1.075**	0.429	0.012
Sales price	**-0.620**	0.168	0.000	**-0.493**	0.163	0.003	**-0.407**	0.164	0.013	**-0.286**	0.146	0.050
Fuel efficiency	**0.501**	0.069	0.000	**0.546**	0.075	0.000	**0.568**	0.073	0.000	**0.445**	0.071	0.000

Note 1: Estimates in bold indicate significance at the 10% level.

Note 2: 'ASC' indicates alternative specific constant.

Table 9 shows that the result of choice sets 1 and 2, which include all the possible alternatives and randomly sampled alternatives respectively, are rather similar. Considering this result, the theoretical findings of Brownstone, Bunch, and Train (2000) that consistent estimates can be obtained from a sampling approach for reconstructing the consumers' choice set when IIA models such as a multinomial model is employed to analyze consumer preference, is empirically confirmed. A chi-square test was conducted to confirm that the results of choice sets 1 and 2 have the same estimates statistically. If the following null hypothesis is accepted, then the estimates of the two choice sets are statistically the same:

$$H_0 : \boldsymbol{\alpha}_{choiceset1} = \boldsymbol{\alpha}_{choiceset2} \ \& \ \boldsymbol{\beta}_{choiceset1} = \boldsymbol{\beta}_{choiceset2}$$

The chi-square test showed that the null hypothesis is accepted with a probability of 95.46% ($\chi^2_{11} = 4.46$). Thus, the results of choice sets 1 and 2 are not statistically different.

The estimation results show that the result of choice set 4 that consists of the alternatives which consumers actually include within their choice set is different from the result of other choice set assumptions. In order to confirm this statistically, I test whether the following null hypothesis is accepted:

$$H_0 : \boldsymbol{\alpha}_{choiceset1} = \boldsymbol{\alpha}_{choiceset4} \ \& \ \boldsymbol{\beta}_{choiceset1} = \boldsymbol{\beta}_{choiceset4}$$

As a result of the chi-square test, the null hypothesis is rejected at 1% significance level ($\chi^2_{11} = 40.14$). Therefore, the result of choice set 4 is statistically different from the results of choice sets 1 and 2.

Considering choice set 4, sensitivity to sales price is relatively low when compared to the results of other choice sets. This result indicates that the consumers' sensitivity to attributes could be overestimated if a researcher does not properly consider the actual choice sets of consumers. This is discussed in more detail in section 4.2.3 through market simulations. As shown in all the results of the choice set assumptions, it is revealed that consumers prefer high fuel efficiency. It was also confirmed that the marginal utility of the fuel efficiency is overestimated in the case of choice sets 1 to 3, when compared to the choice set 4.

In addition, the results of choice sets 1 and 2 show that alternative specific constants for mid-size and full-size gasoline sedans are relatively high, while the alternative specific constants for compact gasoline, mid-size diesel, and mid-size eco-friendly sedans are relatively low. On the other hand, in the result of choice set 4, which consists of alternatives that consumers actually considered while purchasing an automobile, the difference between the alternative specific constants tend to be smaller than those of choice sets 1 and 2. In particular, alternative specific constants for economy/subcompact and compact gasoline sedans are positive in the result of choice set 4, while the alternative specific constant for economy/subcompact gasoline sedan is estimated as a negative value and the alternative specific constant for compact gasoline sedan is not statistically different from that for diesel SUV in the result of choice sets 1 and 2. As observed in Table 8, economy/subcompact gasoline sedans do not tend to be considered within the choice set when a consumer purchases other types of vehicles. A previous study theoretically mentioned that alternative specific constants, which do not tend to be considered within the choice set when a consumer purchases other types of vehicles, are downward biased (Swait and Ben-Akiva, 1986). This study empirically confirms the tendency.

Further, I employ four indicators to compare the accuracy of the estimated parameter and prediction capability of the results of the choice set assumptions. Table 10 shows the results of the model suitability test according to the choice set assumptions. It is found that the analysis result of choice set 4, which consists of alternatives that consumers actually consider when they purchase a new automobile, describes consumer preferences more accurately. Therefore, it is important to consider the consumers' actual choice set in order to accurately estimate the utility function of consumers.

Table 10. Model validation by choice set type

	Choice set 1	Choice set 2	Choice set 3	Choice set 4
Log-likelihood	-1283.35	-846.69	-989.16	**-824.83**
BIC (Bayesian Information Criteria)[12]	2672.35	1786.09	2079.56	**1743.84**
holdout cross validation[13]	21.95%	58.54%	50.00%	**69.51%**
k-fold cross validation[14]	18.72%	58.39%	46.23%	**68.60%**

4.2.3 Market Simulation and Discussion

In this section, a market simulation is undertaken using the estimation results of the choice sets assumed in section 4.2.1. Through simulation analysis, I would like to point out how analysis results can be distorted if a researcher fails to consider the actual choice set of a consumer. More specifically, in the scenarios where sales prices decrease and fuel efficiencies improve for eco-friendly (hybrid and electric) vehicles, how consumers' choice probabilities changes is simulated. Considering this, the choice probability of each alternative is estimated by using the information of automobiles that is sold in the Korean automobile market in 2017 as shown in Table 11.

---

[12] $BIC = -2\ln L + k\ln N$ (L = likelihood, k = the number of parameters, N = sample size)

[13] A holdout cross validation is a validation method that estimates the model using 80% of the samples and verifies how much of the actual outcomes of the remaining 20% samples can be recovered.

[14] A k-fold cross validation is a validation method that divides the entire sample into k segments and verifies how much of the actual outcomes of the remaining samples except k-th segment can be recovered using the estimation result of the k-th segment. The average of hit rates indicates the predictive capability of the model.

Table 11. Specification of doemestic cars sold in Korea (May 2017)

Size	Fuel type	Manufacturer	Model	Sales price	Engine dsplacement	Fuel efficiency
Economy/ Subcompact	Gasoline	Hyundai	Accent	1386	1480	9.64
		Kia	Morning, Pride	1366	1246	9.30
		Chevrolet	Spark, Aveo	1438	1181	9.58
Compact	Gasoline	Hyundai	Avante	1935	1591	8.54
		Kia	K3	1780	1591	9.14
		Renault Samsung	SM3	1785	1598	9.97
		Chevrolet	Cruze	2020	1399	8.97
	Diesel	Hyundai	Avante	2034	1582	13.93
		Kia	K3	2110	1582	14.20
		Renault Samsung	SM3 Neo	2038	1461	13.66
		Chevrolet	Cruze	2306	1598	11.58
	Hybrid/ Electric	Hyundai	Ioniq Hybrid, Electric	3373	1580	16.62
		Kia	Soul EV	4208	1591	15.97
		Renault Samsung	SM3 ZE	4000	1598	14.06
Mid-size	Gasoline	Hyundai	Sonata	2605	1999	8.37
		Kia	K5	2810	1795	7.71
		Renault Samsung	SM5	2195	1998	8.37
		Chevrolet	Malibu	2879	1744	7.91
	Diesel	Hyundai	Sonata	2743	1685	12.73
		Kia	K5	2825	1685	12.23
		Renault Samsung	SM5	2659	1461	12.73
		Chevrolet	Malibu	2858	1956	10.26
	Hybrid/ Electric	Hyundai	Sonata Hybrid	3108	1999	11.93
		Kia	K5 Hybrid	3068	1999	11.46
		Chevrolet	Malibu Hybrid	3264	1796	11.36

Table 11. Specification of doemestic cars sold in Korea (May 2017) (continued)

Size	Fuel type	Manufacturer	Model	Sales price	Engine dsplacement	Fuel efficiency
Full-size	Gasoline	Hyundai	Grandeur, Genesis	4726	3120	6.47
		Kia	K7, K9	4811	3205	6.58
		Renault Samsung	SM7	3625	2997	6.51
		Chevrolet	Alpheon	3415	2691	6.71
		SsangYong	Chairman	6779	3399	5.35
SUV	Gasoline	Hyundai	Tucson, Maxcruz	3131	2467	6.55
		Kia	Sportage, Sorento	2636	1999	6.28
		Renault Samsung	QM5	2508	1998	7.04
		Chevrolet	Trax	2118	1362	8.11
		SsangYong	Tivoli	2026	1597	7.48
	Diesel	Hyundai	Tucson, SantaFe, Maxcruz	3312	2045	9.80
		Kia	Sportage, Sorento, Mohave	3458	2299	9.14
		Renault Samsung	QM3, QM6	2879	1728	11.09
		Chevrolet	Trax, Captiva	2720	1777	10.10
		SsangYong	Tivoli, Rexton	3160	1877	9.24

Note 1: The unit of sales price, engine displacement, and fuel efficiency are 10,000 KRW, cc, and km/1,000 KRW, respectively (for electric vehicles, the engine displacement of similar gasoline or diesel model is employed).

Note 2: If a manufacturer sells several models of the same vehicle type, the average of the specifications, such as sales price, engine displacement, and fuel efficiency is employed.

The scenarios where sales prices decrease and fuel efficiencies improve for eco-friendly vehicles are as follows. First, it is assumed that sales prices of compact and mid-size eco-friendly vehicles decrease from 5% to 50% in the sales price reduction scenarios (P1 ~ P10) as shown in Table 12. Here, the specifications of other vehicle models as well as engine displacements and fuel efficiency of eco-friendly vehicles are assumed to remain the same as the 2017 conditions. Next, it is assumed that the distance which can be traveled with 1,000 KRW increases from 3% to 30% in the fuel efficiency improvement scenarios (E1 ~ E10) as shown in Table 13. In this case, the specifications of other vehicle models as well as engine displacements and sales prices of eco-friendly vehicles are also assumed to remain the same as the 2017 conditions.

Table 12. Scenarios for sales price reduction of hybrid/electric vehicles

			P0	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
Gasoline /Diesel	Specifications of gasoline and diesel cars stay the same with the current condition												
Hybrid /Electric	Sales price [10,000 KRW]	Compact (average)	3,860	3,667	3,474	3,281	3,088	2,895	2,702	2,509	2,316	2,123	1,930
		Mid-size (average)	3,147	2,989	2,832	2,675	2,517	2,360	2,203	2,045	1,888	1,731	1,573
	Engine displacement		Eginie displacement stays the same with the current condition										
	Fuel efficiency		Fuel cost per km stays the same with the current condition										

Note: P0 is the current condition (May 2017) and P1 to P10 are the hypothesized scenarios where the sales price of hybrid and electric cars decrease.

Table 13. Scenarios for fuel efficiency improvement of hybrid/electric vehicles

			E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Gasoline /Diesel	Specifications of gasoline and diesel cars stay the same with the current condition												
Hybrid /Electric	Sales price		Sales price stays the same with the current condition										
	Engine displacement		Eginie displacement stays the same with the current condition										
	Fuel efficiency [km/1,000KRW]	Compact (average)	15.55	16.02	16.48	16.95	17.42	17.88	18.35	18.82	19.28	19.75	20.22
		Mid-size (average)	11.59	11.93	12.28	12.63	12.98	13.32	13.67	14.02	14.37	14.71	15.06

Note: E0 is the current condition (May 2017) and E1 to E10 are the hypothesized scenarios where the fuel efficiency of hybrid and electric cars improve.

Tables 14 and 15 show the results of a scenario analysis using the estimation results of the choice sets 1 and 4, respectively. In addition, Figures 11 and 12 represent how the choice probabilities change based on the scenarios. The results of the scenario analysis show that estimation results of choice set 1 describes consumers more sensitive to sales price reduction and fuel efficiency improvement when compared to the results of choice set 4. In other words, the result of scenario analysis using choice set 1 shows that the choice probability of eco-friendly vehicles increases steeply, but this can be regarded as the overestimated result.

Therefore, the market simulation indicates that the estimated preference can be distorted when a research does not accurately consider the consumers' actual choice set, such as assuming that the consumers' choice set consists of all the possible alternatives in the market. Thus, the result of the empirical study emphasizes that it is important to accurately observe and consider the consumers' actual choice set while analyzing consumer preference.

Table 14. Simulation result of sales price scenarios

Choice set 1		**P0**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
Gasoline		66.9%	66.5%	66.1%	65.6%	65.1%	64.6%	64.0%	63.3%	62.6%	61.9%	61.0%
Diesel		28.0%	27.8%	27.6%	27.4%	27.2%	27.0%	26.7%	26.5%	26.2%	25.9%	25.5%
Hybrid /Electric	Compact	2.2%	2.4%	2.7%	3.0%	3.3%	3.7%	4.1%	4.6%	5.1%	5.6%	6.2%
	Mid-size	3.0%	3.3%	3.6%	4.0%	4.3%	4.7%	5.2%	5.6%	6.1%	6.7%	7.3%
Choice set 4		**P0**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
Gasoline		81.1%	81.1%	80.8%	80.6%	80.5%	80.3%	80.1%	79.9%	79.7%	79.5%	79.3%
Diesel		15.2%	15.2%	15.1%	15.1%	15.1%	15.0%	15.0%	15.0%	14.9%	14.9%	14.8%
Hybrid /Electric	Compact	1.7%	1.8%	1.9%	2.0%	2.1%	2.3%	2.4%	2.5%	2.6%	2.8%	2.9%
	Mid-size	2.0%	2.0%	2.1%	2.2%	2.3%	2.4%	2.5%	2.6%	2.7%	2.9%	3.0%

Table 15. Simulation result of fuel efficiency scenarios

Choice set 1		E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
Gasoline		66.9%	66.1%	65.2%	64.1%	62.7%	61.2%	59.3%	57.2%	54.7%	52.0%	48.9%
Diesel		28.0%	27.6%	27.2%	26.8%	26.2%	25.6%	24.8%	23.9%	22.9%	21.7%	20.4%
Hybrid /Electric	Compact	2.2%	2.7%	3.4%	4.3%	5.4%	6.7%	8.3%	10.2%	12.4%	15.0%	18.1%
	Mid-size	3.0%	3.5%	4.2%	4.9%	5.7%	6.6%	7.6%	8.7%	10.0%	11.3%	12.6%
Choice set 4		**E0**	**E1**	**E2**	**E3**	**E4**	**E5**	**E6**	**E7**	**E8**	**E9**	**E10**
Gasoline		81.1%	81.0%	79.8%	79.0%	78.0%	76.8%	75.4%	73.8%	71.9%	69.8%	67.3%
Diesel		15.2%	15.2%	14.9%	14.8%	14.6%	14.4%	14.1%	13.8%	13.5%	13.1%	12.6%
Hybrid /Electric	Compact	1.7%	2.2%	2.6%	3.2%	4.0%	4.8%	5.9%	7.1%	8.6%	10.4%	12.4%
	Mid-size	2.0%	2.3%	2.6%	3.0%	3.5%	4.0%	4.6%	5.3%	6.0%	6.8%	7.6%

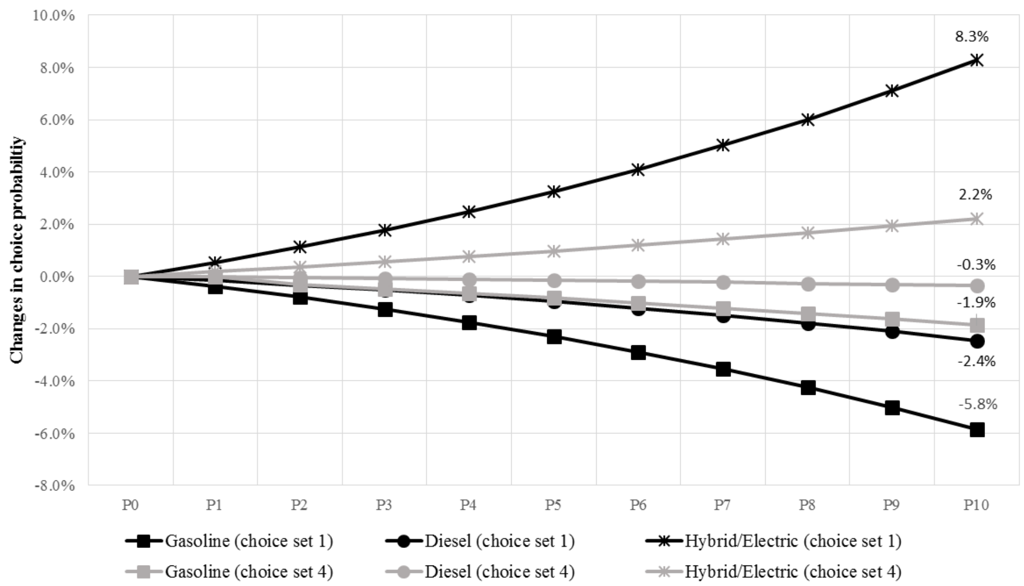


Figure 11. Changes in choice probability for sales price scenarios

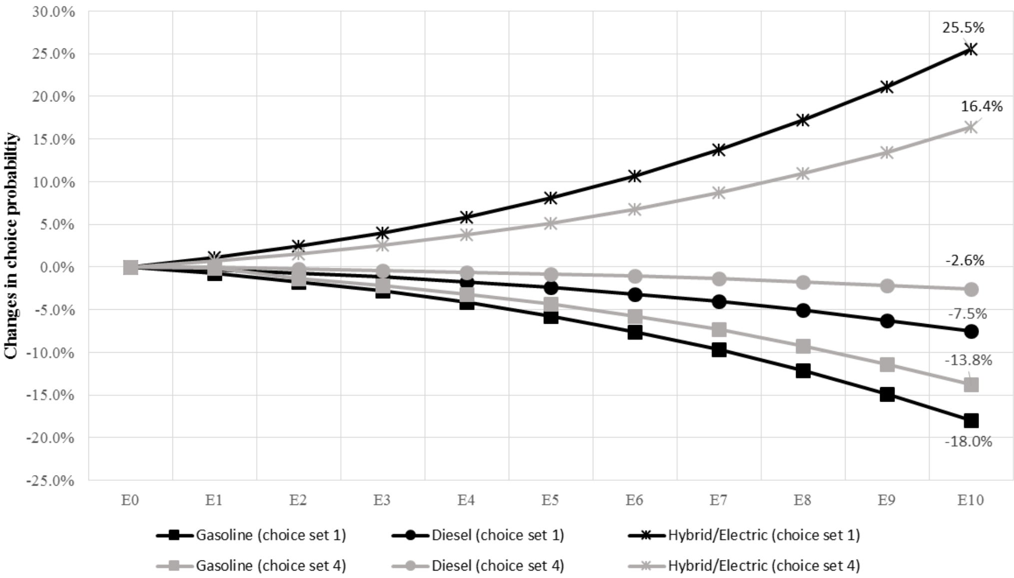


Figure 12. Changes in choice probability for fuel efficiency scenarios

4.2.4 Conclusion and Implications

Several previous studies were interested in identifying the method to reconstruct consumers' choice set and its impact on the estimation results when analyzing consumer preference based on RP data (Swait and Ben-Akiva, 1986; Brownstone, Bunch, and Train, 2000; Swait 2001; Woo, Moon, Lee, Jo, and Lee, 2017). In this empirical study, the theoretical findings of previous studies related to assuming consumers' choice set in consumer preference analysis are empirically confirmed. The results of the empirical study are summarized as follows. First, it is shown that the estimation results of choice sets, which are a choice set consisting of all the possible alternatives and a choice set consisting of the alternatives randomly drawn from the universal choice set, are consistent. Second, if consumer preference is analyzed by using a choice set consisting of all the possible alternatives in the market without taking consumers' actual choice set into account, the alternative specific constants for the alternatives which do not tend to be included in the choice set, are downward biased. Third, the study confirmed that the estimated preference for product attributes is biased when it is assumed that consumers consider all the possible alternatives in their choice set. More specifically, consumers' marginal utility for product attributes tend to be overestimated. When purchasing a new automobile, consumers generally set certain criteria according to their purchasing ability or driving purpose and construct their choice set with alternatives having similar characteristics (Kaplan, Bekhor, and Shiftan, 2011; Kaplan, Shiftn, and Bekhor, 2012; Zolfaghari, Sivakumar, and Polak, 2013). Thereafter, consumers evaluate the attributes of the alternatives in the choice set, thus they may not be that sensitive to the product attributes. If a researcher does not consider the consumers' choice set behavior and assumes that the consumers make a choice within the universal choice set, the analysis could misestimate consumer preference by estimating that consumers sensitively react to the product attributes.

In the next section, I analyze the consumers' choice set formation behavior, and its impact on product choice and usage determination by adopting the proposed model which considers choice set formation in the automobile market.

4.3 Identifying Connection between Choice Set Formation, Product Choice and Usage Determination Behavior

4.3.1 Introduction

In this section, the proposed model explained in chapter 3 is applied to the automobile market. In the choice set formation stage, which is the first stage of the proposed model, consumers' choice set formation behavior depending on individual demographics, perception of the environment, and knowledge of energy policies is observed by using information on the alternatives that consumers actually compared while purchasing a new automobile. In the second stage of the proposed model, which represents the consumers' product choice behavior, the model is constructed under the assumption that consumer preference for a new automobile is affected by the choice set formation behavior, and the relationship between the choice set formation and automobile choice behaviors is identified. Finally, a linear model is employed to analyze how the vehicle type and individual characteristics can affect annual mileage in the third stage, which is the usage determination stage. The model adopts an instrumental variable to correct the endogeneity problem that occurs due to the correlation between product choice and usage determination.

4.3.2 Estimation Results and Discussion

4.3.2.1 Choice Set Formation Stage

In the first stage, the consumers' choice set formation behavior while purchasing a new automobile is analyzed. More specifically, the tendency of choice set formation behavior according to individual characteristics is identified using information on the types of vehicles that the consumer considered in the choice set. Considering this, a multivariate probit model is employed, as introduced in chapter 3, to analyze the consumers' choice set formation behavior in the automobile market.

In this empirical study, variables such as age, household income, knowledge of the environmental externalities, energy policies in the transport sector, perception of the environmental positive effect of eco-friendly vehicles, number of family members, and fuel type of the vehicle previously owned, are used in the model for the choice set formation stage. Table 16 shows how the variables affect the tendency of being included in the choice set for each vehicle type.

Table 16. Estimation result of stage 1 (coefficients)

Type	ASC	Age	Income	Know. Extern	Know. Policy	Environ. H/EV	No. of Family	Gasoline Owned	Diesel Owned
1	**-0.519**	**-0.022**	**-0.094**	-0.242	0.146	**0.334**	-0.163	0.133	-0.370
2	-0.123	**-0.016**	**-0.099**	0.020	0.140	**0.177**	-0.007	0.004	**-0.810**
3	-0.227	**-0.024**	**-0.112**	**0.422**	**-0.241**	-0.050	0.135	**-0.893**	0.407
4	**-1.234**	0.003	**0.088**	**0.318**	-0.076	0.055	**-0.160**	**-0.783**	**-0.772**
5	-0.198	0.002	**0.098**	0.163	-0.102	**-0.199**	-0.047	**0.663**	-0.327
6	**-1.068**	0.005	0.009	0.301	**-0.323**	0.150	-0.136	**-0.423**	**0.579**
7	**-1.001**	**0.016**	-0.021	0.079	**-0.289**	-0.028	-0.056	0.067	-0.241
8	**-1.035**	0.008	**0.108**	0.088	-0.007	**-0.230**	0.030	**0.332**	0.295
9	**-0.565**	0.006	**-0.094**	**-0.146**	0.101	0.065	-0.054	-0.188	-0.003
10	0.178	**0.010**	**-0.120**	0.062	**-0.197**	**-0.224**	**0.112**	**-0.450**	0.117

Note 1: Estimates in bold indicate significance at the 10% level.

Note 2: "Know. Extern" indicates the knowledge of environmental externalities of the transport sector and "Know. Policy" indicates the knowledge of energy policies in the transport sector. In addition, "Environ. H/EV" indicates the perception of environmental positive effect of eco-friendly vehicles (hybrid/electric vehicles) and "No. of Family" indicates the number of family members. Finally, "Gasoline Owned" and "Diesel Owned" represent whether a respondent has owned a gasoline car or diesel car, respectively.

Based on the results, young consumers tend to consider relatively small size vehicles (economy/subcompact and compact sedans) within their choice set and high income consumers tend to include relatively expensive vehicles, such as subcompact eco-friendly sedans, and mid-size and full-size gasoline sedans in their choice set. In addition, consumers who are familiar with the environmental externalities of the transport sector tend to consider compact diesel and compact eco-friendly vehicles in their choice set. This implies that consumers who are aware of the environmental externalities of the transport sector understand that fuel-efficient vehicles have an environmentally positive effect, thus they consider fuel-efficient vehicles, such as compact diesel and eco-friendly sedans in their choice set. Consumers who are aware of the energy policies in the transport sector generally do not tend to consider diesel vehicles in their choice set, and consumers who have a large family are more likely to consider diesel SUVs in their choice set. Moreover, consumers who have previously owned a gasoline vehicle are more likely to consider mid-size and full-size gasoline sedans within their choice set, but do not include mid-size and full-size diesel sedans, and diesel SUVs in their choice set. On the other hand, consumers who have previously owned a diesel vehicle tend to include mid-size diesel sedans in their choice set, but do not consider compact gasoline sedans within their choice set. Lastly, consumers who have previously owned a gasoline or diesel vehicle are less likely to consider eco-friendly vehicles within their choice set, which implies that the consumers who purchase an automobile for the first time tend to include eco-friendly vehicles in their choice set.

Certain alternatives either tend to be considered in the choice set together or not to be included together in the choice at the consumers' choice set formation stage. This tendency can be confirmed by the estimated variance-covariance matrix of a multivariate probit model. The estimation result of the variance-covariance matrix is shown in Table 17.

Table 17. Estimation result of stage 1 (variance-covariance matrix)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1.000	**0.201**	0.130	-0.032	**-0.501**	**-0.871**	-0.133	**-0.124**	**-0.226**	**-0.372**
2		1.000	**0.412**	-0.156	**-0.221**	0.079	**-0.182**	**-0.459**	-0.120	**-0.469**
3			1.000	**-0.688**	-0.111	0.070	-0.120	**-0.699**	0.149	-0.037
4				1.000	0.036	-0.120	**0.391**	**0.338**	**-0.535**	0.076
5					1.000	**0.406**	**0.466**	0.118	0.085	**-0.295**
6						1.000	0.014	-0.008	**0.305**	**0.174**
7							1.000	-0.270	**-0.572**	0.091
8								1.000	**0.234**	**-0.162**
9									1.000	-0.124
10										1.000

Note: Estimates in bold indicate significance at the 10% level.

Economy/subcompact sedans tend to be included in a consumer's choice set with compact gasoline sedans and compact gasoline sedans are less likely to be included together with mid-size and full-size gasoline sedans in the choice set. Compact eco-friendly sedans are likely to be considered together with mid-size eco-friendly sedans, and mid-size gasoline sedans tend to be included in the choice set with mid-size diesel and mid-size eco-friendly sedans. Thus, mid-size sedans are likely to be considered together within the choice set, regardless of the fuel type.

4.3.2.2 Product Choice Stage

The second stage of the proposed model represents the stage in which consumers purchase a new automobile. In the product choice stage, it is assumed that the preference for an automobile is affected by the choice set formation behavior. The model for the product choice stage is constructed by applying a multinomial logit model as shown in equation (3.9) and (3.11). Table 18 shows the estimation result of the model.

Based on the result, it is confirmed that the choice set formation behavior significantly affects consumer preferences of the alternative as the coefficients of the predicted choice set formation probabilities (Prob 1 ~ Prob 10) are significantly estimated. As shown in section 4.2, the consumers' choice set formation behavior can vary according to their individual characteristics. Figure 13 shows how the alternative specific constants can be changed depending on the consumers' choice set formation behavior. Comparing Figures 14 and 15, which show the estimated alternative specific constants by applying choice sets 1 and 4 respectively in section 4.2, the result of the proposed model for the product choice stage implies that the order of the alternative specific constants can change according to the consumers' choice set formation behavior.

Table 18. Estimation result of stage 2

Attribute	Estimates	S.E	P>|z|	90% Confident Interval
ASC1	-0.446	0.465	0.338	[-1.211, 0.320]
ASC2	**-1.150**	0.636	0.070	[-2.195, -0.104]
ASC3	**-3.646**	0.623	0.000	[-4.671, -2.621]
ASC4	**-6.545**	1.642	0.000	[-9.245, -3.844]
ASC5	0.302	0.586	0.606	[-0.662, 1.267]
ASC6	**-3.062**	0.731	0.000	[-4.265, -1.859]
ASC7	**-1.394**	0.841	0.098	[-2.778, -0.010]
ASC8	0.859	0.733	0.241	[-0.347, 2.064]
ASC9	**-3.471**	1.317	0.008	[-5.638, -1.304]
Prob1	**3.551**	1.426	0.013	[1.205, 5.896]
Prob2	**4.074**	1.036	0.000	[2.369, 5.778]
Prob3	**4.271**	2.060	0.038	[0.882, 7.660]
Prob4	**9.796**	3.207	0.002	[4.521, 15.071]
Prob5	**3.048**	0.879	0.001	[1.601, 4.495]
Prob6	**6.469**	2.520	0.010	[2.324, 10.615]
Prob7	1.087	8.033	0.892	[-12.127, 14.300]
Prob8	**5.031**	1.709	0.003	[2.220, 7.842]
Prob9	**14.161**	4.813	0.003	[6.245, 22.077]
Prob10	**4.467**	1.268	0.000	[2.382, 6.552]
Sales price	**-0.628**	0.169	0.000	[-0.906, -0.350]
Fuel efficiency	**0.503**	0.070	0.000	[0.388, 0.617]

Note: Estimates in bold indicate significance at the 10% level.

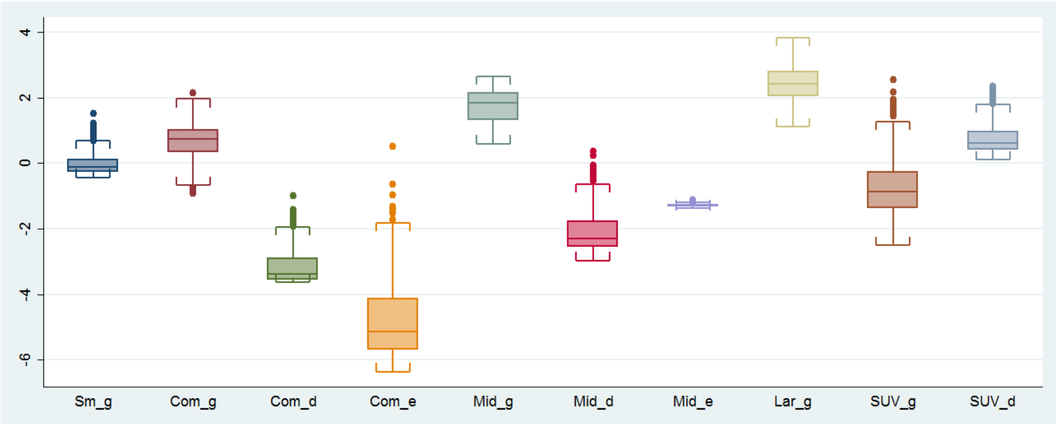


Figure 13. Alternative specific constants affected by choice set formation

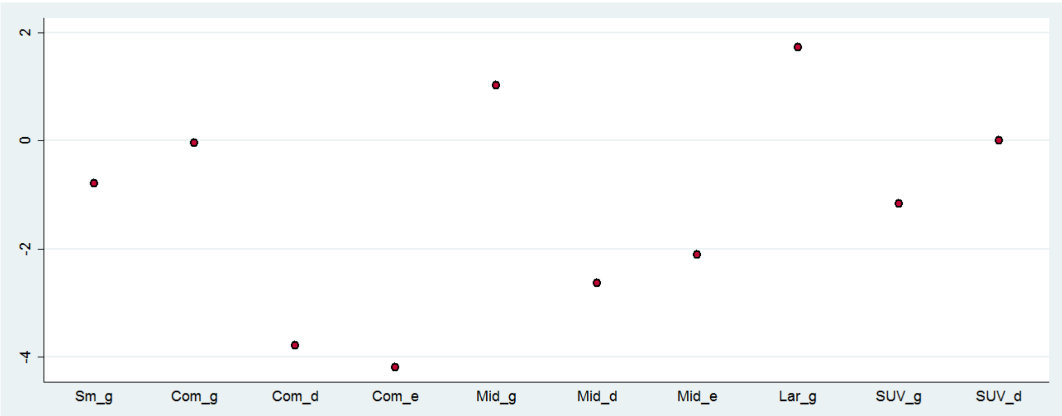


Figure 14. Alternative specific constants (result of choice set 1)

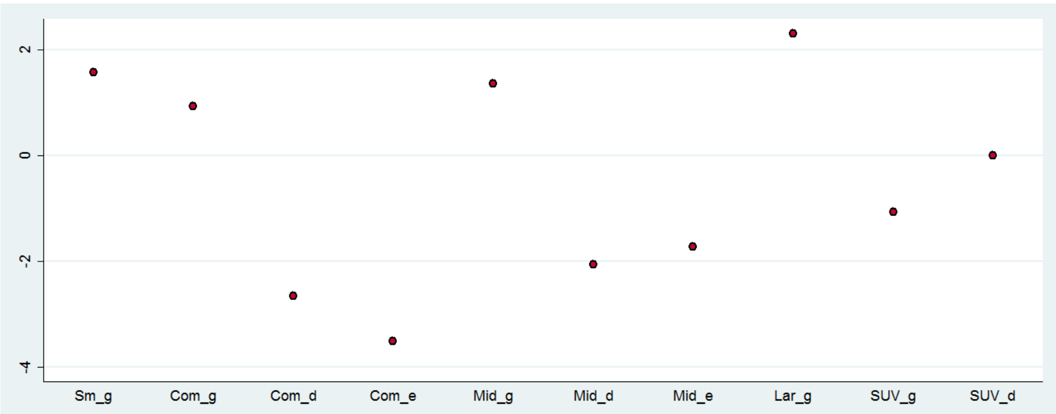


Figure 15. Alternative specific constants (result of choice set 4)

For example, consumers who consider economy/subcompact and compact gasoline sedans within their choice set have relatively high alternative specific constants for that vehicle type (Figure 16). If a consumer includes compact eco-friendly and mid-size gasoline sedans in their choice set, the alternative specific constant for the compact eco-friendly sedans increases considerably (Figure 17).

In addition, the degree of variation of alternative specific constant is different for each alternative as shown in Figure 13. The variation in alternative specific constants for compact eco-friendly sedans and gasoline SUVs is relatively large, while the variation in the alternative specific constants for economy/subcompact gasoline vehicles are relatively small.

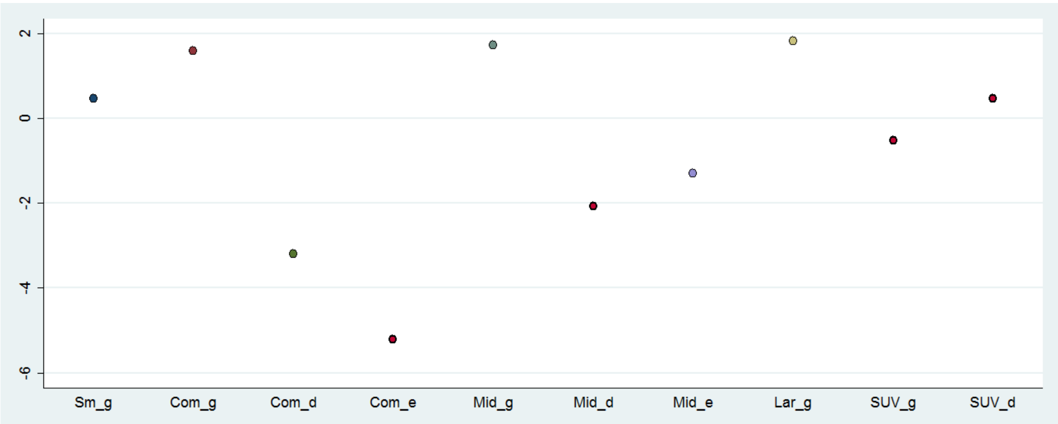


Figure 16. Alternative specific constants ($\{\text{Sm\_g, Com\_g}\} \subset C_{ID=33}$)

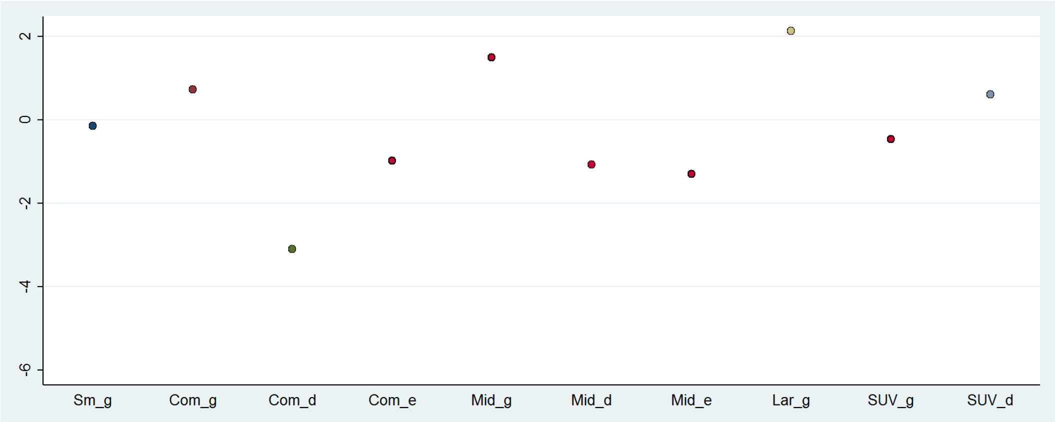


Figure 17. Alternative specific constants ($\{\text{Com\_e, Lar\_g}\} \subset C_{ID=246}$)

4.3.2.3 Usage Determination Stage

In the last stage of usage determination, the consumers make a decision on the mileage of the automobile chosen in the product choice stage based on the vehicle specifications and individual characteristics. In order to analyze this behavior, the model specified in equation (3.17) is employed, which is based on the methodology of Dubin and McFadden (1984). In the analysis model, variables such as engine displacement, SUV dummy, fuel efficiency, and gender of the driver are used[15].

The estimation result of the usage determination stage is shown in Table 19. When the engine displacement increases by 100cc, people drive more at approximately 411km per year. Consumers tend to drive more at approximately 443 km a year if the distance that can be traveled with 1,000 KRW is increased by

[15] The variable selection is based on Woo, Moon, Lee, Jo, and Lee (2017).

one kilometer. In addition, the annual mileage of male drivers is 3,273 km longer than that of female drivers. Moreover, there is no significant difference between the annual mileages of sedans and SUVs, when the other specifications are the same.

Table 19. Estimation result of stage 3

Attribute	Estimates	S.E	P>|t|	90% Confident Interval
Displacement	**4.114**	1.020	0.000	[2.430, 5.797]
SUV	2582.381	3622.592	0.477	[-3396.767, 8561.529]
Fuel efficiency	**443.324**	256.329	0.085	[20.249, 866.399]
Gender	**3273.293**	843.854	0.000	[1880.498, 4666.088]

Note 1: Estimates in bold indicate significance at the 10% level.

Note 2: Cases where the annual mileage is less than 6,000 km or more than 35,000 km (137 samples) are excluded from the data for estimation.

4.3.3 Conclusion and Implications

In this section, the proposed model was applied to the Korean automobile market. Further, the consumers' choice set formation behavior, its impact on preference for automobile alternatives, consumers' mileage affected by automobile specifications, and individual characteristics are analyzed.

The results of this empirical study are summarized as follows. First, it is confirmed that the consumers' choice set formation behavior is significantly affected by the fuel type of the vehicle previously owned, knowledge of environmental externalities, energy policies in the transport sector, perception of eco-friendly vehicles as well as demographics. Thus, the study observed that the behavioral tendency of including vehicle types in the choice set can vary depending on individual characteristics. Second, it is found that consumers' choice set formation behavior can affect the utility that consumers obtain from an automobile. In this study, it is assumed that the alternative specific constant of each vehicle type can vary based on the consumers' choice set formation behavior, particularly the probability that a consumer considers a particular vehicle type within their choice set. Considering the analysis results, as assumed, it was confirmed that the choice set formation behavior significantly affects consumer preference while choosing a product. Third, this empirical study found that the variation of an alternative specific constant caused by the changes of consumers' choice set formation behavior varies among the alternatives. For a certain vehicle type, such as compact eco-friendly sedans, its alternative specific constant can significantly vary based on the consumers' choice set formation behavior.

In the following section, the scenarios of perception changes and technological improvements, which affect consumers' choice set formation behavior and automobile specifications, are constructed and the ripple effects are analyzed using the estimation results in this section. Ultimately, I examine the strengths of the proposed model as a methodology for market forecasting.

4.4 *Ex-ante* Impact Evaluation of Consumers' Perception Change and Technological Improvement on Automobile Market and Environment

4.4.1 Introduction

This section attempts to analyze the changes in the automobile market due to the changes in consumer perceptions and technological improvement, and the following ripple effects. I construct scenarios on consumers' perception changes and technological improvement, and perform the analysis by using the steps as shown in Figure 18.

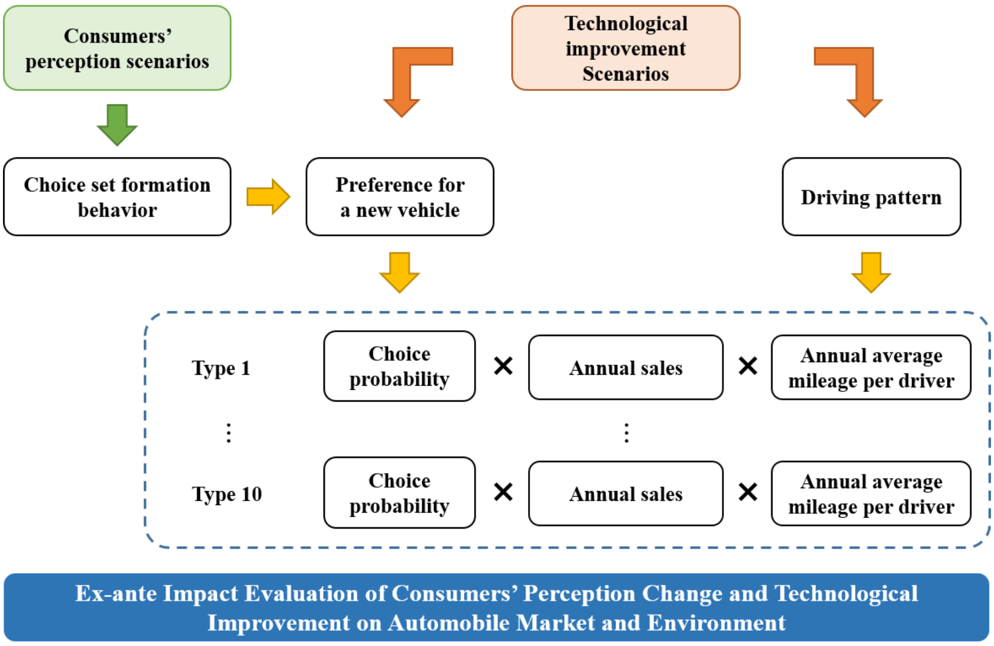


Figure 18. Schematic diagram of the scenario analysis

I set four scenarios as presented in Table 20. In scenarios 1 and 2, the fuel efficiency of gasoline vehicles, which have relatively low fuel efficiency compared to the other types of vehicles, is improved. In scenarios 3 and 4, the sales prices of eco-friendly vehicles (e.g., hybrid and electric vehicles) are lowered[16]. More specifically, scenarios 1 and 2 assume that fuel efficiency of gasoline vehicles is improved from 5% to 20% based on the currently sold domestic vehicles in 2017. Scenarios 3 and 4 assume that the sales price of domestic eco-friendly vehicles are reduced from 5% to 20%. In these cases, scenarios 1 and 3 only assume technology improvement, while scenarios 2 and 4 consider both, technology improvement and changes in consumers' perception. Scenarios 2 and 4 can investigate the effect of changes in choice set formation behavior on the preference of vehicles, and enable to observe the ripple effects of consumers' perception changes. In this section, I analyze the changes in the automobile market under each scenario and the ripple effects, such as changes in energy consumption and greenhouse gas emissions.

[16] For 2017, the average fuel efficiency of vehicles by fuel type is 7.84km/1,000KRW for gasoline vehicles, 11.59km/1,000KRW for diesel vehicles, and 13.57km/1,000KRW for eco-friendly vehicles. In addition, the average sales price is 27.61 million KRW for gasoline vehicles, 27 million KRW for diesel vehicles, and 35.03 million KRW for eco-friendly vehicles.

Table 20. Scenarios of technological improvement and perception change

			C0	C1	C2	C3	C4
Fuel efficiency improvement of gasoline vehicles	**Scenario 1**	Fuel efficiency of gasoline [km/1,000KRW]	7.84 (-)	8.23 (5%↑)	8.62 (10%↑)	9.01 (15%↑)	9.41 (20%↑)
		Knowledge of energy policy [Five-point likert scale]	Perception stays the same with the current condition				
	Scenario 2	Fuel efficiency of gasoline [km/1,000KRW]	7.84 (-)	8.23 (5%↑)	8.62 (10%↑)	9.01 (15%↑)	9.41 (20%↑)
		Knowledge of energy policy [Five-point likert scale]	3.06 (-)	3.56 (0.5p↑)	4.03 (1.0p↑)	4.42 (1.5p↑)	4.70 (2.0p↑)
Sales price reductoin of hybrid/electric vehicles	**Scenario 3**	Sales price of hybrid/electric [10,000 KRW]	3,503 (-)	3,328 (5%↓)	3,153 (10%↓)	2,978 (15%↓)	2,803 (20%↓)
		Knowledge of externality [Five-point likert scale]	Perception stays the same with the current condition				
	Scenario 4	Sales price of hybrid/electric [10,000 KRW]	3,503 (-)	3,328 (5%↓)	3,153 (10%↓)	2,978 (15%↓)	2,803 (20%↓)
		Knowledge of externality [Five-point likert scale]	2.73 (-)	3.22 (0.5p↑)	3.68 (1.0p↑)	4.11 (1.5p↑)	4.46 (2.0p↑)

Note 1: C0 indicates the current condition.

Note 2: In the scenarios of consumers' perception change, the perception variables, which are measured on a five-point likert scale are assumed to increase only for a consumer who does not have the maximum value (five point).

Note 3: Other variables which are not represented above remain in the current condition.

4.4.2 Estimation Results and Discussion

Based on each scenario assumed previously, this section analyzes the change in the automobile market and the following ripple effect (change in energy consumption and CO_2 emissions) of each scenario. The following are the results of the simulation analysis based on the model proposed in this dissertation.

4.4.2.1 Fuel efficiency improvement of gasoline vehicles

Improvement of fuel efficiency for gasoline vehicles, which has low fuel efficiency when compared to others, was assumed in scenarios 1 and 2. Here, I assumed a case where the perception of the consumer is identical to the present and the choice set formation behavior does not change (scenario 1), and another case where the perception of the consumer and the choice set formation behavior change (scenario 2).

In scenario 2, I assumed an improvement in the awareness level of the energy management policy in the transport sector as the change in consumers' perception. As shown in the results in section 4.3, consumers with a high awareness level in the energy policy in transport sector (corporate average fuel efficiency, energy labeling, tax on transportation fuels, etc.) generally do not tend to consider diesel vehicles in their choice set. Therefore, an improvement in the awareness level in the energy management policy in transport sector is expected to boost the effect of fuel efficiency improvement of gasoline vehicles. Tables 21 and 22 illustrate the change of market share for C0 ~ C4 in scenarios 1 and 2. Figure 19 compares these two results.

Table 21. Changes in market share for scenario 1

Type	C0	C1	C2	C3	C4
1	12.68%	13.85%	14.97%	16.04%	17.05%
2	21.59%	23.46%	25.24%	26.93%	28.50%
3	3.68%	3.17%	2.70%	2.27%	1.90%
4	0.63%	0.54%	0.46%	0.39%	0.33%
5	23.18%	24.47%	25.58%	26.50%	27.24%
6	3.53%	3.03%	2.58%	2.18%	1.82%
7	3.38%	2.90%	2.47%	2.08%	1.74%
8	8.77%	8.89%	8.92%	8.88%	8.77%
9	1.74%	1.81%	1.85%	1.89%	1.90%
10	20.81%	17.89%	15.23%	12.84%	10.75%

Table 22. Changes in market share for scenario 2

Type	C0	C1	C2	C3	C4
1	12.68%	14.20%	15.55%	16.73%	17.75%
2	21.59%	24.05%	26.22%	28.08%	29.68%
3	3.68%	3.01%	2.44%	1.97%	1.60%
4	0.63%	0.56%	0.48%	0.41%	0.34%
5	23.18%	25.10%	26.57%	27.64%	28.36%
6	3.53%	2.54%	1.86%	1.41%	1.10%
7	3.38%	2.91%	2.47%	2.07%	1.71%
8	8.77%	9.12%	9.27%	9.26%	9.13%
9	1.74%	1.85%	1.93%	1.97%	1.98%
10	20.81%	16.67%	13.21%	10.47%	8.35%

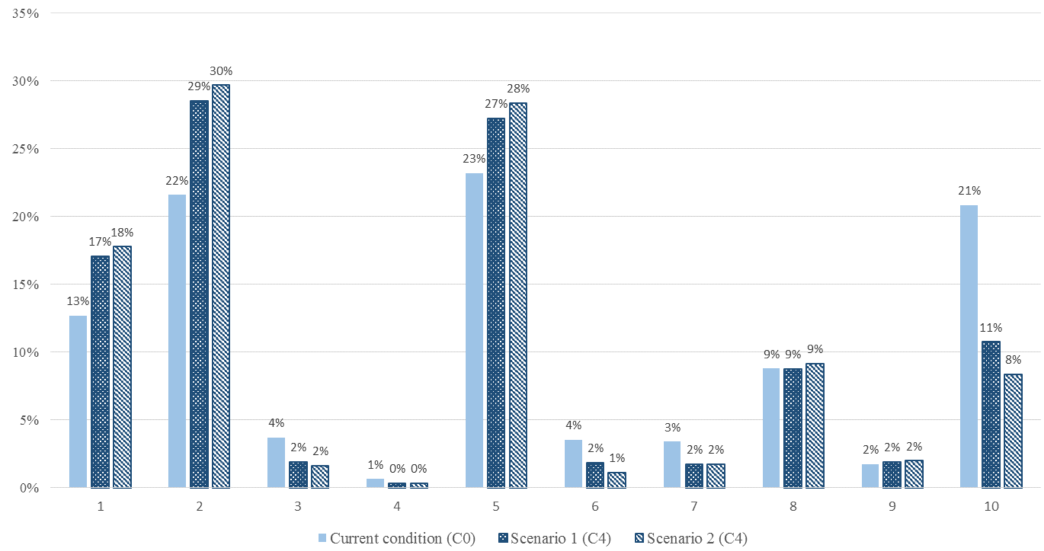


Figure 19. Market shares for scenarios 1 and 2 (C4)

The results show that the market share of gasoline vehicles (vehicle types 1, 2, 5, 8, and 9) increase in both scenarios, 1 and 2. The increase is more steep in scenario 2. This may be attributed to those who are more aware of the energy management policy and tend to exclude diesel vehicles in their choice set. Considering that those who are more aware of the energy management policies may drive their vehicles less than others, they do not tend to include diesel vehicles, which have excellent fuel efficiency, in their choice set.

Next, I analyze the ripple effect of the change in automobile market shares. More specifically, I analyze the change in energy consumption and CO_2 emissions when the market share of each vehicle type changes. Considering this, I assumed the annual sales of new vehicles as 1.3 million[17], and the mileage of each vehicle type is estimated using the results of section 4.3[18]. When estimating energy consumption and CO_2 emissions, I considered that the mileage changes as the fuel efficiency of the gasoline vehicle changes. First, energy consumption is calculated by dividing the annual mileage with the fuel efficiency of each vehicle type. Then, CO_2 emissions are derivied by multiplying the energy consumption with the emission factor for each fuel type (Korea Energy Agency, 2017)[19]. Tables 23 and 24, and Figures 20 and 21 illustrate the results of the estimated energy consumption and CO_2 emissions for each fuel type.

As shown in both scenarios 1 and 2, the total energy consumption and CO_2 emissions decrease as the fuel efficiency of the gasoline vehicle improves. The decrease in total energy consumption may stem from two factors. First, the energy consumption may have decreased as a result of the fuel spent for driving the same distance decreased as the fuel economy improved.

---

[17] Newly registered vehicles in 2016 was 1,279,418 in Korea (Ministry of Land, Infrastructure and Transport, 2017)

[18] The behavior that the consumer increases their energy consumption responding to the drop in relative price from the improvement in energy efficiency is called a rebound effect. Excluding the rebound effect may result in overestimation of the effect of energy efficiency improvement. Therefore, this study assumes that the mileage of the consumer is effected by the vehicle's fuel efficiency. According to the results of section 4.3, as the distance that can be driven with 1000 KRW is increased to one kilometer, the consumer tends to drive 440km more annually.

[19] I assumed 0.00208tCO2/L for gasoline, 0.00258tCO2/L for diesel, and 0.00047tCO2/kWh and 1kWh = 0.294L for eco-friendly vehicles.

Table 23. Changes in energy consumption for scenarios 1 and 2

Scenario 1

Fuel type	**Changes of energy consumption**			
	C1	C2	C3	C4
Gasoline	+ 1.10 % (+ 21,150,271 L)	+ 1.90 % (+ 36,332,750 L)	+ 2.40 % (+ 45,868,570 L)	+ 2.63 % (+ 50,284,506 L)
Diesel	- 5.45 % (- 57,324,346 L)	- 10.41 % (- 109,527,039 L)	- 14.84 % (- 156,213,392 L)	- 18.74 % (- 197,290,039 L)
Hybrid /Electric	- 8.79 % (- 6,144,061 L)	- 16.79 % (- 11,739,179 L)	- 23.95 % (- 16,743,053 L)	- 30.24 % (- 21,145,675 L)
Total	**- 1.39 % (- 42,318,136 L)**	**- 2.80 % (- 84,933,468 L)**	**- 4.19 % (- 127,087,874 L)**	**- 5.54 % (- 168,151,207 L)**

Scenario 2

Fuel type	**Changes of energy consumption**			
	C1	C2	C3	C4
Gasoline	+ 2.36 % (+ 45,193,562 L)	+ 3.84 % (+ 73,483,663 L)	+ 4.56 % (+ 87,335,149 L)	+ 4.72 % (+ 90,310,189 L)
Diesel	- 7.97 % (- 83,914,975 L)	- 14.46 % (- 152,222,260 L)	- 19.54 % (- 205,610,960 L)	- 23.43 % (- 246,600,269 L)
Hybrid /Electric	- 8.42 % (- 5,890,168 L)	- 16.56 % (- 11,576,221 L)	- 23.99 % (- 16,773,385 L)	- 30.53% (- 21,346,180 L)
Total	**- 1.47 % (- 44,611,581 L)**	**- 2.97 % (- 90,314,818 L)**	**- 4.45 % (- 135,049,195 L)**	**- 5.85 % (- 177,636,260 L)**

Table 24. Changes in CO_2 emission for scenarios 1 and 2

Scenario 1

Fuel type	Changes of CO_2 emission			
	C1	C2	C3	C4
Gasoline	+ 2.30 % (+ 43,993 ton)	+ 3.95 % (+ 75,572 ton)	+ 4.98 % (+ 95,407 ton)	+ 5.46 % (+ 104,592 ton)
Diesel	- 14.05 % (- 147,897 ton)	- 26.85 % (- 282,580 ton)	- 38.29 % (- 403,031 ton)	- 48.36 % (- 509,008 ton)
Hybrid /Electric	- 14.05 % (- 9,824 ton)	- 26.85 % (- 18,771 ton)	- 38.29 % (- 26,773 ton)	- 48.36 % (- 33,812 ton)
Total	**- 3.75 % (- 113,729 ton)**	**- 7.44 % (- 225,779 ton)**	**- 11.01 % (- 334,396 ton)**	**- 14.43 % (- 438,229 ton)**

Scenario 2

Fuel type	Changes of CO_2 emission			
	C1	C2	C3	C4
Gasoline	+ 4.91 % (+ 94,003 ton)	+ 7.98 % (+ 152,846 ton)	+ 9.49 % (+ 181,657 ton)	+ 9.81 % (+ 187,845 ton)
Diesel	- 20.57 % (- 216,501 ton)	- 37.31 % (- 392,733 ton)	- 50.40 % (- 530,476 ton)	- 60.45 % (- 636,229 ton)
Hybrid /Electric	- 13.47 % (- 9,419 ton)	- 26.48 % (- 18,511 ton)	- 38.36 % (- 26,821 ton)	- 48.82 % (- 34,133 ton)
Total	**- 4.34 % (- 131,917 ton)**	**- 8.51 % (- 258,398 ton)**	**- 12.37 % (- 375,640 ton)**	**- 15.89 % (- 482,517 ton)**

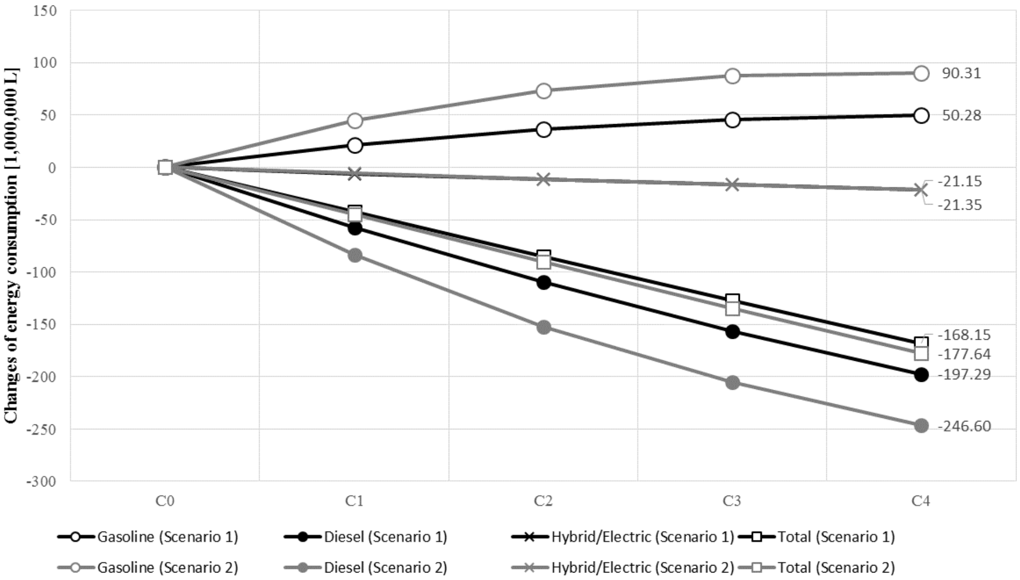


Figure 20. Changes in energy consumption for scenarios 1 and 2

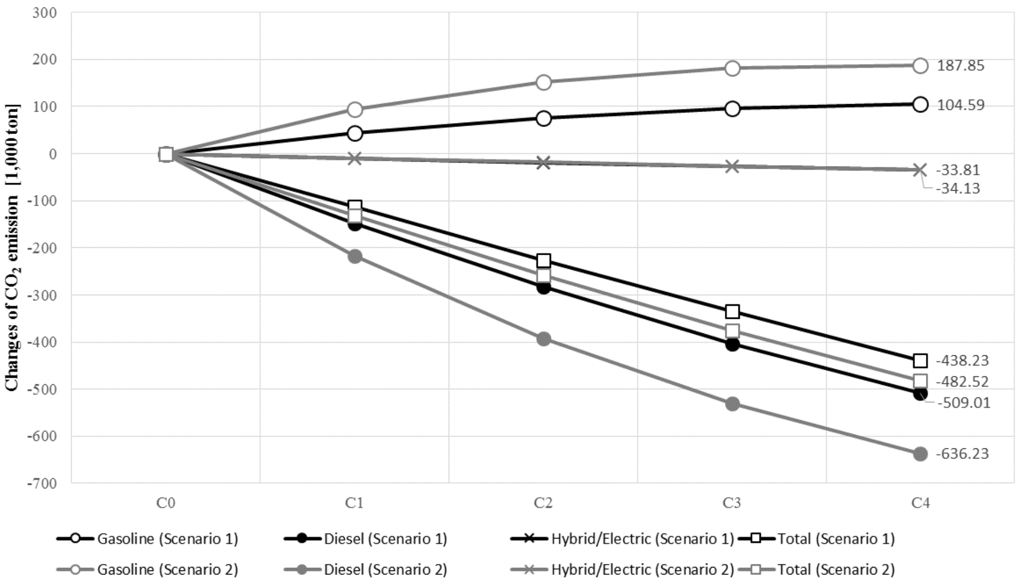


Figure 21. Changes in CO_2 emission for scenarios 1 and 2

Second, as shown in Figure 19, the total energy consumption may have decreased as the market share of diesel vehicles (vehicle types 3, 6, and 10), which have relatively long mileage, decreased. Next, the decrease in total CO_2 emissions may stem from the decrease in total energy consumption, particularly for diesel fuels which have high CO_2 emissions per unit energy.

A decrease in energy consumption and CO_2 emissions was higher in scenario 2 when compared to scenario 1. This is owing to the decrease in diesel vehicle's market share being greater in scenario 2.

4.4.2.2 Sales price reduction of hybrid/electric vehicles

An improvement of fuel efficiency for eco-friendly vehicles (hybrid and electric vehicles), which have high fuel efficiency and low CO_2 emissions when compared to the others, was assumed in scenarios 3 and 4. The scenarios assume that technology improvement for eco-friendly vehicles lower the production cost and thus, lower the sales price. Similar to the previous example, I assumed a case where the perception of the consumer is identical to the present and the choice set formation behavior does not change (scenario 3), and another case where the perception of the consumer and the choice set formation behavior changes (scenario 4).

In scenario 4, I assumed an improvement in the awareness level of the environmental externalities in transport sector as the change in consumers' perception. As shown in the estimation results in section 4.3, consumers with high awareness levels of environmental externalities in the transport sector tend to include compact eco-friendly vehicles in their choice set. Therefore, an improvement in the awareness level of environmental externalities in transportation sector is expected to boost the effect of lowered sales price of eco-friendly vehicles, which improves the market share of these vehicles. Tables 25 and 26 illustrate the change of market share for C0 ~ C4 for scenarios 3 and 4. Figure 22 compares these two results.

Table 25. Changes in market share for scenario 3

Type	C0	C1	C2	C3	C4
1	12.68%	12.63%	12.57%	12.50%	12.43%
2	21.59%	21.50%	21.40%	21.29%	21.16%
3	3.68%	3.67%	3.65%	3.63%	3.61%
4	0.63%	0.70%	0.79%	0.88%	0.98%
5	23.18%	23.09%	22.98%	22.86%	22.73%
6	3.53%	3.51%	3.49%	3.48%	3.46%
7	3.38%	3.71%	4.08%	4.48%	4.91%
8	8.77%	8.73%	8.69%	8.65%	8.60%
9	1.74%	1.74%	1.73%	1.72%	1.71%
10	20.81%	20.72%	20.63%	20.52%	20.40%

Table 26. Changes in market share for scenario 4

Type	C0	C1	C2	C3	C4
1	12.68%	12.56%	12.36%	12.09%	11.74%
2	21.59%	21.38%	21.04%	20.57%	19.99%
3	3.68%	4.27%	4.99%	5.83%	6.67%
4	0.63%	1.03%	1.69%	2.72%	4.15%
5	23.18%	22.96%	22.60%	22.09%	21.47%
6	3.53%	3.49%	3.44%	3.36%	3.26%
7	3.38%	3.69%	4.01%	4.33%	4.64%
8	8.77%	8.68%	8.55%	8.36%	8.12%
9	1.74%	1.33%	1.03%	0.83%	0.69%
10	20.81%	20.61%	20.29%	19.83%	19.27%

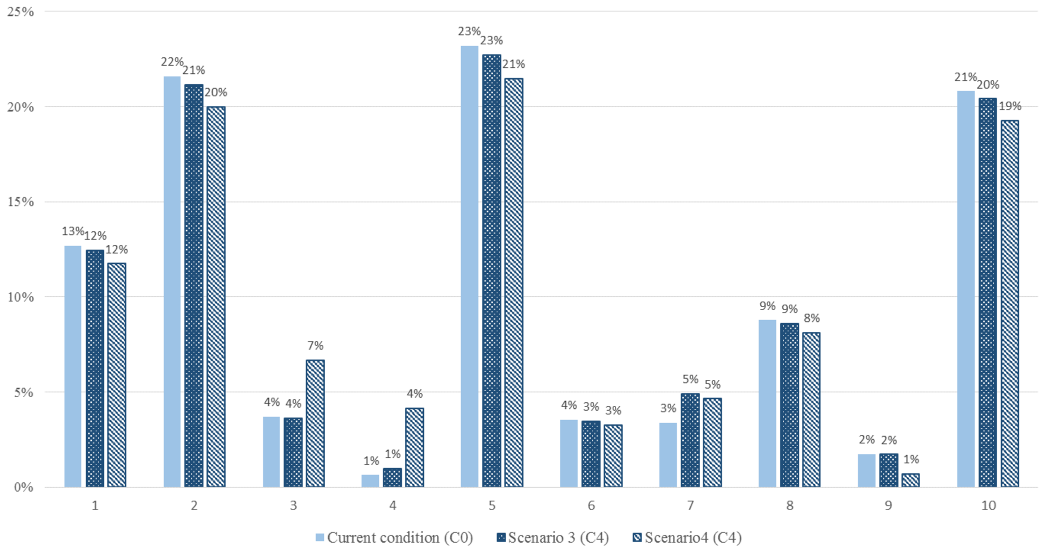


Figure 22. Market shares for scenarios 3 and 4 (C4)

The results show that the market share of eco-friendly vehicles (vehicle types 4 and 7) increase in both, scenarios 3 and 4. However, the change in market share for other vehicle types were different for the two scenarios. Scenario 3 only assumes that the sales price of eco-friendly vehicles decreases due to technology improvement, thus the market share of all other vehicle types excluding eco-friendly vehicles have decreased. On the other hand, scenario 4 assumes that consumers' awareness level of environmental externalities increases when the sales price of eco-friendly vehicles deacreases. With the additional assumption, the market share of compact diesel vehicle (vehicle type 3) is increased. This is due to the estimate in section 4.3, where the probability of including not only compact eco-friendly vehicles but also compact diesel vehicles in their choice set is increased as the awareness level of environmental externalities in transport sector is improved. This could have increased the preference of compact diesel vehicles, which results in higher market share.

Next, I analyze the ripple effect of the change in automobile market shares. More specifically, I analyze the change in energy consumption and CO_2 emissions as the market share of each vehicle type changes. Tables 27 and 28, and Figures 23 and 24 illustrate the results of estimated energy consumption and CO_2 emissions for each fuel type.

Table 27. Changes in energy consumption for scenarios 3 and 4

Scenario 3

Fuel type	Changes of energy consumption			
	C1	C2	C3	C4
Gasoline	- 0.20 % (- 3,910,911 L)	- 0.43 % (- 8,202,572 L)	- 0.67 % (- 12,908,131 L)	- 0.94 % (- 18,062,762 L)
Diesel	- 0.16 % (- 1,733,618 L)	- 0.35 % (- 3,636,013 L)	- 0.54 % (- 5,721,880 L)	- 0.76 % (- 8,006,811 L)
Hybrid /Electric	+ 6.37 % (+ 4,451,600 L)	+ 13.35 % (+ 9,337,030 L)	+ 21.02 % (+ 14,694,140 L)	+ 29.41 % (+ 20,563,108 L)
Total	**- 0.04 % (- 1,192,929 L)**	**- 0.08 % (- 2,501,555 L)**	**- 0.13 % (- 3,935,871 L)**	**- 0.18 % (- 5,506,464 L)**

Scenario 4

Fuel type	Changes of energy consumption			
	C1	C2	C3	C4
Gasoline	- 0.82 % (- 15,704,498 L)	- 1.81 % (- 34,583,392 L)	- 3.01 % (- 57,564,724 L)	- 4.39 % (- 84,006,185 L)
Diesel	+ 0.23 % (+ 2,433,452 L)	+ 0.40 % (+ 4,168,514 L)	+ 0.45 % (+ 4,767,224 L)	+ 0.33 % (+ 3,431,128 L)
Hybrid /Electric	+ 10.84 % (+ 7,579,391 L)	+ 25.71 % (+ 17,973,390 L)	+ 46.07 % (+ 32,209,625 L)	+ 72.50 % (+ 50,686,023 L)
Total	**- 0.19 % (- 5,691,655 L)**	**- 0.41 % (- 12,441,488 L)**	**- 0.68 % (- 20,587,876 L)**	**- 0.98 % (- 29,889,034 L)**

Table 28. Changes in CO_2 emission for scenarios 3 and 4

Scenario 3

Fuel type	Changes of CO_2 emission			
	C1	C2	C3	C4
Gasoline	- 0.42 % (- 8,135 ton)	- 0.89 % (- 17,061 ton)	- 1.40 % (- 26,849 ton)	- 1.96 % (- 37,571 ton)
Diesel	- 0.42 % (- 4,473 ton)	- 0.89 % (- 9,381 ton)	- 1.40 % (- 14,762 ton)	- 1.96 % (- 20,658 ton)
Hybrid /Electric	+ 10.18 % (+ 7,118 ton)	+ 21.35 % (+ 14,930 ton)	+ 33.61 % (+ 23,496 ton)	+ 47.03 % (+ 32,881 ton)
Total	**- 0.18 % (- 5,489 ton)**	**- 0.38 % (- 11,512 ton)**	**- 0.60 % (- 18,115 ton)**	**- 0.83 % (- 25,347 ton)**

Scenario 4

Fuel type	Changes of CO_2 emission			
	C1	C2	C3	C4
Gasoline	- 1.71 % (- 32,665 ton)	- 3.76 % (- 71,933 ton)	- 6.26 % (- 119,735 ton)	- 9.13 % (- 174,733 ton)
Diesel	+ 0.60 % (+ 6,278 ton)	+ 1.02 % (+ 10,755 ton)	+ 1.17 % (+ 12,299 ton)	+ 0.84 % (+ 8,852 ton)
Hybrid /Electric	+ 17.33 % (+ 12,120 ton)	+ 41.11 % (+ 28,740 ton)	+ 73.67 % (+ 51,504 ton)	+ 115.92 % (+ 81,048 ton)
Total	**- 0.47 % (- 14,267 ton)**	**- 1.07 % (- 32,439 ton)**	**- 1.84 % (- 55,931 ton)**	**- 2.79 % (- 84,833 ton)**

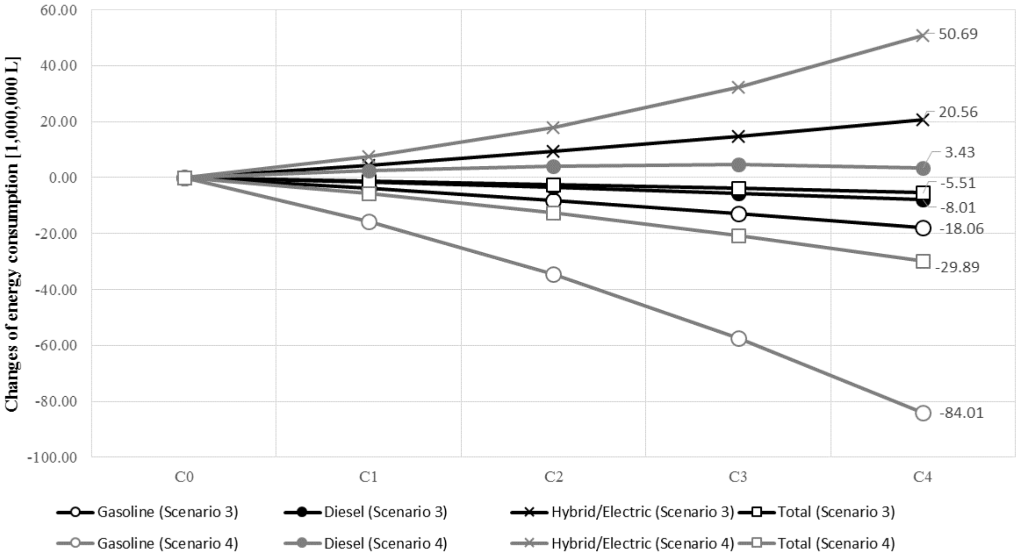


Figure 23. Changes in energy consumption for scenarios 3 and 4

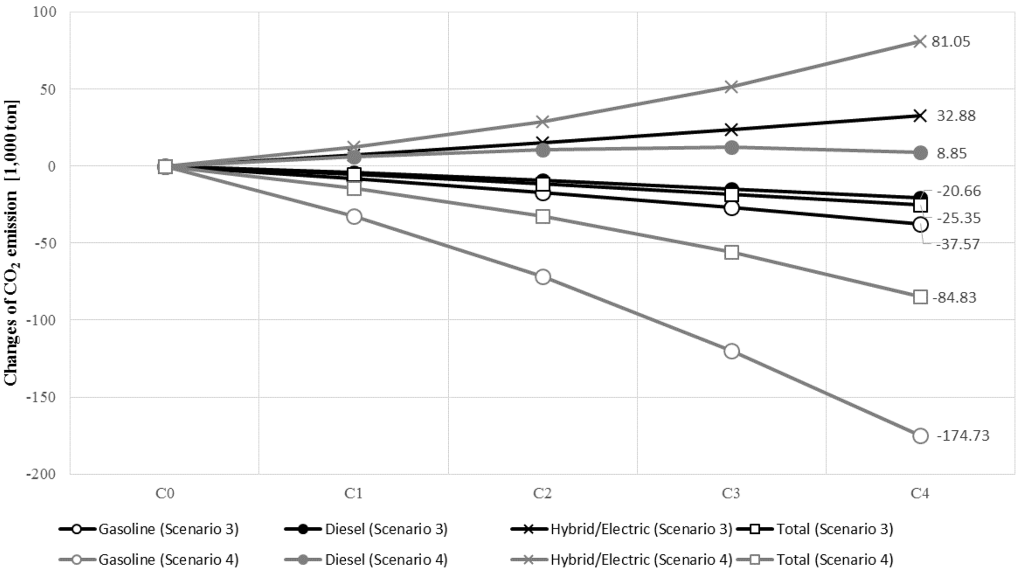


Figure 24. Changes in CO_2 emission for scenarios 3 and 4

The results show that the energy consumption and CO_2 emissions decrease as the sales price of eco-friendly vehicles decrease. This trend is stronger in scenario 4, which considers the improvement in awareness level for environmental externalities. This may be due to the fact that those who are more aware of environmental externalities tend to include compact eco-friendly vehicles and diesel vehicles in their choice set. Therefore, the market share for these vehicles increase when compared to scenario 3 and the total energy consumption decreases. Diesel vehicles have higher CO_2 emissions per mileage when compared to other vehicle types, but the decrease in CO_2 emissions from the decrease in market share for gasoline vehicles is bigger and therefore, the CO_2 emissions in scenario 4, which assumes an improvement in awareness level, is smaller than that in scenario 3.

4.4.3 Conclusion and Implications

The simulation analysis using the estimation results of the proposed model showed the effects of technology improvement and consumers' perception changes on the market and its environmental ripple effect. In this section, scenarios were assumed as cases where the fuel efficiency of gasoline vehicles improves due to the innovation in the automobile market and cases where the sales price of eco-friendly vehicles decreases. Furthermore, I investigated the importance of promoting consumers' perception by analyzing how the market changes due to technology improvements and are affected by changes in the choice set formation behavior. The purpose of the empirical study is to confirm that the market can be changed by choice set formation behavior of the consumer who has limited information processing capacity.

The results confirmed the expected effects actually presented in the two technology improvement scenarios (improvement in fuel efficiency of gasoline vehicles and reduction in sales price of eco-friendly vehicles), which were expected to result in energy savings and environmental improvement. This empirical study also confirmed that the effect of change in choice set formation behavior along with technology improvement can be varied according to the scenarios.

First, it is confirmed that the impact of technology improvement is synergistic when the awareness level of consumers on the energy management policy in the transport sector rises. As the awareness level of consumers on the energy management policy increases, diesel vehicles with relatively longer mileage are less likely to be considered in the choice set when compared to other types of vehicles, and thereby, gasoline vehicles are preferred, resulting in greater energy savings and CO_2 emission reduction.

Next, in the scenario where the sales price of eco-friendly vehicles decreases, it is confirmed that the effect of technology improvement is synergistic when the level of consumer's knowledge on environmental externalities in the transport sector increases. Thus, consumers with a high level of knowledge on environmental externalities in the transport sector tend to be more inclined toward considering compact eco-friendly vehicles in their choice set. One notable point is that as the awareness level of the environmental externalities in the transport sector increases, the overall energy consumption and CO_2 emissions are further reduced, even though the probability that compact diesel vehicles are taken into account in the choice set increases. As the market share of gasoline vehicles dramatically declines and the share of eco-friendly vehicles increases, the overall effect on environment is positive in spite of the increase of the share of diesel vehicels.

The following implications on technology policy are derived from scenario analysis using the proposed model. First, in a market such as the automobile market where consumers cannot evaluate all the alternatives, the consumers' choice set formation behavior has a significant impact on the preference for alternatives. In addition, it is necessary to consider the tendency of consumers' choice set formation in market forecasting as changes in the choice set formation behavior can lead to changes in energy consumption and CO_2 emissions. Second, market can be influenced by the changes in the choice set formation behavior according to consumers' perception change. The ripple effect of technology improvement can be synergistic or diluted, depending on the consumers' perception change. Therefore, in order to reduce energy consumption and CO_2

emissions in the transport sector, the government should improve consumers' perception and allow consumers to consider fuel efficiencient and eco-friendly alternatives into their choice set through subsidies on the purchase or R&D support for firms.

Chapter 5. Summary and Conclusion

5.1 Concluding Remarks and Contributions

As the momentum of technological innovation increases, many products are being introduced into the market in a short period of time. Considering this circumstance, consumers with limited information processing capacity tend to construct the choice set with some alternative among all possible alternatives according to their individual characteristics rather than evaluating all the possible alternatives.

Some of the previous studies which analyzed consumer preference using RP data attempted to incorporate consumers' choice set formation behavior within the choice model, but few of them observed the effect of choice set formation behavior on consumer preference of alternatives. In addition, few studies analyzed the economic and environmental ripple effects of consumers' choice set formation behavior by considering the usage determination behavior as well as product choice with choice set formation behavior. In this dissertation, the new model which consists of the choice set formation stage, product choice stage, and usage determination stage is proposed. Further, the effect of consumers' choice set formation behavior on product choice and usage is analyzed by empirically applying the proposed model.

The academic contributions of this dissertation proposing a new methodology are summarized as follows. First, I proposed a model that can explicitly analyze the effects of the consumers' choice set formation behavior on consumers' preference of the alternatives. By applying the proposed model, not only is the researcher able to observe how the consumers' choice set formation behavior is affected by individual consumer characteristics, but is also able to observe how it affects the consumers' preference of alternatives. Utilizing this model provides valuable implications for policy makers and marketers who aim to increase the adoption rate of innovative products.

Second, the proposed model not only considers consumers' choice set formation stage and product choice stage, but also the usage determination stage. This suggests a framework for forecasting market changes due to consumer behavior change and technological innovation. Through this, it is possible to investigate the synergistic effect on the diffusion of new products due to the improvement of the product from technological innovation and changes in consumers' perception.

The present dissertation identified the potential problems in analyzing consumer preference when the consumer's actual choice set is not considered by utilizing the proposed model and RP data of the automobile market. Considering this, I empirically demonstrated that the estimation of the alternative specific constant may be biased when the consumers' choice set is not appropriately taken into account. I empirically proved the previous studies' arguments on the sensitivity of attributes being overestimated when the consumers' choice set is assumed to be a universal choice set. Also, the empirical analysis showed that alternative specific constants can be estimated differently for each consumer depending on the consumers' choice set formation behavior. Finally, based on the analysis, I proved that environmental effects (energy saving and reduction of greenhouse gas emission) from technological innovation in the automobile industry can be influenced by the consumers' choice set formation behavior.

5.2 Limitations and Future Research Topics

The present study improves the discrete-continuous choice model by incorporating the consumers' choice set formation stage within the model's traditional product choice stage, in addition to the usage determination stage. The limitations of this study and future research topics are as follows.

Firstly, the effect of consumers' choice set formation behavior on the preference of alternatives may be different for each consumer. The present dissertation assumed that the choice set formation behavior has a homogeneous effect on the alternative preference for each consumer. However, further studies may be able to identify the heterogeneous effects of such behavior by applying the hierarchical Bayesian or the latent class model in the product choice stage.

Secondly, the present study constructed the model using only cross-sectional data without taking time dimension into account. Such analysis may be misleading if the product is frequently purchased, as consumers' choice may be affected by past experience or the purchasing environment at different times. Therefore, further studies may be able to analyze consumer preference by using learnable stochastic models. For example, the hidden Markov model can be used to represent the changes in the consumers' choice set formation behavior through time (Goulias, 1999; Netzer, Lattin, and, Srinivasan, 2008; Beyreuther, Carniel, and Wassermann, 2008; Xiong, Chen, He, Guo, and Zhang, 2015).

Lastly, the proposed model is designed to analyze the effect of the consumers' choice set behavior on the consumers' product choice and usage determination only based on RP data. If the data collection method is improved, further studies may be able to expand the model to enable the analysis of new products not yet presented in the market. For example, researchers may be able to acknowledge consumers' choice set formation behavior by constructing a hypothetical choice framework for the process of searching and evaluating future products before the purchase. Based on the hypothetical framework, the researcher can conduct the discrete choice experiment and utilize the proposed model to analyze consumers' preference for future products.

Bibliography

Ahn, J., Jeong, G., & Kim, Y. (2008). A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics*, 30(5), 2091-2104.

Allenby, G. M., & Rossi, P. E. (2006). Hierarchical Bayes models. In *The Handbook of Marketing Research* (pp. 418-440). London, UK: SAGE Publications.

Andrews, R. L., & Srinivasan, T. C. (1995). Studying consideration effects in empirical choice models using scanner panel data. *Journal of Marketing Research*, 32(1), 30-41.

Axsen, J., Mountain, D. C., & Jaccard, M. (2009). Combining stated and revealed choice research to simulate the neighbor effect: The case of hybrid-electric vehicles. *Resource and Energy Economics*, 31(3), 221-238.

Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 12(1), 9-24.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the 'bias heuristic'. *Acta Psychologica*, 50(3), 201-252.

Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of consumer research*, 7(3), 234-248.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of consumer research*, 25(3), 187-217.

Beyreuther, M., Carniel, R., & Wassermann, J. (2008). Continuous hidden Markov models: application to automatic earthquake detection and classification at Las Canãdas caldera, Tenerife. *Journal of Volcanology and Geothermal Research*, 176(4), 513-518.

Bhat, C. R. (2005). A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological*, 39(8), 679-707.

Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, 42(3), 274-303.

Bhat, C. R., & Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B: Methodological*, 36(7), 593-616.

Bhat, C. R., & Sen, S. (2006). Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model. *Transportation Research Part B: Methodological*, 40(1), 35-53.

Bhat, C. R., Sen, S., & Eluru, N. (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B: Methodological*, 43(1), 1-18.

Bhat, C. R., Castro, M., & Khan, M. (2013). A new estimation approach for the multiple discrete–continuous probit (MDCP) choice model. *Transportation Research Part B: Methodological*, 55, 1-22.

Börjesson, M. (2008). Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E: Logistics and Transportation Review*, 44(6), 1025-1038.

Brownstone, D., Bunch, D. S., & Train, K. (2000). Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5), 315-338.

Cantillo, V., & de Dios Ortúzar, J. (2005). A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological*, 39(7), 641-657.

Cascetta, E., & Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, 9(4), 249-263.

Chen-Yi, H., Ke-Ting, C., & Gwo-Hshiung, T. (2007). FMCDM with fuzzy DEMATEL approach for customers' choice behavior model. *International*

Journal of Fuzzy Systems, 9(4), 236.

Chiang, J., Chib, S., & Narasimhan, C. (1998). Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics*, 89(1), 223-248.

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, 347-361.

Dewey, J. (1910). *How We Think*. Boston, MA: D. C. Heath and Company.

Dubin, J. A., & McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345-362.

Engel, J., Blackwell, R., & Kollat, D. (1978). *Consumer Behavior*. Hinsdale, IL: Dryden Press.

Fang, H. A. (2008). A discrete–continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, 42(9), 736-758.

Ferdous, N., Pinjari, A. R., Bhat, C. R., & Pendyala, R. M. (2010). A comprehensive analysis of household transportation expenditures relative to other goods and services: an application to United States consumer expenditure data. *Transportation*, 37(3), 363-390.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62, 451-482.

Goldberg, P. K. (1998). The effects of the corporate average fuel efficiency standards in the US. *The Journal of Industrial Economics*, 46(1), 1-33.

Goulias, K. G. (1999). Longitudinal analysis of activity and travel pattern dynamics using generalized mixed Markov latent class models. *Transportation Research Part B: Methodological*, 33(8), 535-558.

Greene, W. H. (1981). Sample Selection Bias as a Specification Error: Comment. *Econometrica*, 49(3), 795-798.

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), 681-698.

Gupta, A. K., & Souder, W. E. (1998). Key drivers of reduced cycle time. *Research-Technology Management*, 41(4), 38-43.

Hausman, J. A. (1979). Individual discount rates and the purchase and utilization of energy-using durables. *Bell Journal of Economics*, 10(1), 33-54.

Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *The American Economic Review*, 71(4), 662-676.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4), 475-492.

Heckman, J.J. (1978). Dummy Endogenous Variables in a Simultaneous Equation

System. *Econometrica*, 46(4), 931-959.

Hensher, D. A., Rose, J., & Greene, W. H. (2005). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*, 32(3), 203-222.

Hensher, D. A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of applied econometrics*, 21(6), 861-878.

Horowitz, J. L., & Louviere, J. J. (1995). What is the role of consideration sets in choice modeling?. *International Journal of Research in Marketing*, 12(1), 39-54.

Jeong, J., Kim, C. S., & Lee, J. (2011). Household electricity and gas consumption for heating homes. *Energy Policy*, 39(5), 2679-2687.

Jun, S. H., Vogt, C. A., & MacKay, K. J. (2010). Online information search strategies: A focus on flights and accommodations. *Journal of Travel & Tourism Marketing*, 27(6), 579-595.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263-291.

Kaplan, S., Bekhor, S., & Shiftan, Y. (2011). Development and estimation of a semi-compensatory residential choice model based on explicit choice protocols. *The Annals of Regional Science*, 47(1), 51-80.

Kaplan, S., Shiftan, Y., & Bekhor, S. (2012). Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B: Methodological*, 46(2), 291-304.

Keane, M., & Wasi, N. (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics*, 28(6), 1018-1045.

Kim, J., Allenby, G. M., & Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3), 229-250.

Koop, G., Poirier, D. J., & Tobias, J. L. (2007). *Bayesian econometric methods*. New York, NY: Cambridge University Press.

Korea Energy Agency. (2017). *Oil conversion tons calculated*. Retrieved June 12, 2017, from http://co2.kemco.or.kr/toe/toe.aspx.

Lapersonne, E., Laurent, G., & Le Goff, J. J. (1995). Consideration sets of size one: An empirical investigation of automobile purchases. *International Journal of Research in Marketing*, 12(1), 55-66.

Lee, L. F. (1982). Some approaches to the correction of selectivity bias. *The Review of Economic Studies*, 49(3), 355-372.

Lee, L. F. (1983). Generalized Econometric Models with Selectivity. *Econometrica*, 51(2), 507-512.

Liu, L., & Dukes, A. (2016). Consumer Search with Limited Product Evaluation. *Journal of Economics & Management Strategy*, 25(1), 32-55.

Lockwood, M. (1996). Non-Compensatory Preference Structures In Non-Market

Valuation Of Natural Area Policy. *Australian Journal of Agricultural Economics*, 40(2), 85-101.

Lussier, D. A., & Olshavsky, R. W. (1979). Task complexity and contingent processing in brand choice. *Journal of Consumer Research*, 6(2), 154-165.

Mannering, F., & Winston, C. (1985). A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization. *RAND Journal of Economics*, 16(2), 215-236.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229-254.

McFadden, D. (1974a). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (pp. 105-142). New York, NY: Academic Press.

McFadden, D. (1974b). The measurement of urban travel demand. *Journal of public economics*, 3(4), 303-328.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of applied Econometrics*, 447-470.

Ministry of Land, Infrastructure and Transport. (2017). *Total Registered Motor Vehicles*. Retrieved June 13, 2017, from http://stat.molit.go.kr.

Moore, G. (1991). *Crossing the Chasm*. New York, NY: Harper Business.

Mueller, M. G., & de Haan, P. (2009). How much do incentives affect car purchase? Agent-based microsimulation of consumer choice of new cars—Part I: Model structure, simulation of bounded rationality, and model validation. *Energy Policy*, 37(3), 1072-1082.

Netzer, O., Lattin, J. M., & Srinivasan, V. (2008). A hidden Markov model of customer relationship dynamics. *Marketing Science*, *27*(2), 185-204.

Olshavsky, R. W. (1979). Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, 24(3), 300-316.

Olshavsky, R. W., & Granbois, D. H. (1979). Consumer decision making—fact or fiction?. *Journal of Consumer Research*, 6(2), 93-100.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational behavior and human performance*, 16(2), 366-387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York, NY: Cambridge University Press.

Pinjari, A. R., & Bhat, C. (2010). A Multiple Discrete–Continuous Nested Extreme Value (MDCNEV) model: formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological*, 44(4), 562-583.

Puckett, S. M., & Hensher, D. A. (2008). The role of attribute processing strategies in estimating the preferences of road freight stakeholders. *Transportation research part E: logistics and transportation review*, 44(3), 379-395.

Punj, G., & Brookes, R. (2002). The influence of pre-decisional constraints on information search and consideration set formation in new automobile purchases. *International Journal of Research in Marketing*, 19(4), 383-400.

Quattrone, A., & Vitetta, A. (2011). Random and fuzzy utility models for road route choice. *Transportation Research Part E: Logistics and Transportation Review*, 47(6), 1126-1139.

Rekola, M. (2003). Lexicographic preferences in contingent valuation: a theoretical framework with illustrations. *Land Economics*, 79(2), 277-291.

Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 429-440.

Rogers, E.M. (1962). *The Diffusion of Innovations*. New York, NY: The Free Press.

Rossi, P. E., Allenby, G. M., & McCulloch, R. (2012). *Bayesian statistics and marketing*. Chichester, UK: John Wiley & Sons.

Scarpa, R., Gilbride, T. J., Campbell, D., & Hensher, D. A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European review of agricultural economics*, 36(2), 151-174.

Shanteau, J. (1989). Cognitive heuristics and biases in behavioral auditing: Review, comments and observations. *Accounting, Organizations and Society*, 14(1-2), 165-177.

Sherman, J. D., Souder, W. E., & Jenssen, S. A. (2000). Differential effects of the primary forms of cross functional integration on product development cycle time. *Journal of product innovation management*, 17(4), 257-267.

Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing letters*, 2(3), 181-197.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), 99-118.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2), 129.

Spissu, E., Pinjari, A. R., Pendyala, R. M., & Bhat, C. R. (2009). A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel. *Transportation*, 36(4), 403-422.

Swait, J. (2001). Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological*, 35(7), 643-666.

Swait, J., & Ben-Akiva, M. (1986). Analysis of the effects of captivity on travel time and cost elasticities. *Behavioural Research for Transport Policy*, 119-134.

Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B:*

Methodological, 21(2), 91-102.

Train, K. (1986): *Qualitative Choice Analysis: Theory, Econometrics and an Application to Automobile Demand*. Cambridge, UK: MIT Press.

Train, K. (2003). *Discrete choice methods with simulation*. New York, NY: Cambridge University Press.

Train, K., & Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. *Applications of simulation methods in environmental and resource economics*, 117-134.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207-232.

Ursic, M. L., & Helgeson, J. G. (1990). The impact of choice phase and task complexity on consumer decision making. *Journal of Business Research*, 21(1), 69-90.

West, S. E. (2004). Distributional effects of alternative vehicle pollution control policies. *Journal of public Economics*, 88(3), 735-757.

Woo, J., Choi, J. Y., Shin, J., & Lee, J. (2014). The effect of new media on consumer media usage: An empirical Study In South Korea. *Technological Forecasting and Social Change*, 89, 3-11.

Woo, J., Moon, H., Lee, M., Jo, M., & Lee, J. (2017). *Ex-ante* impact evaluation of Corporate Average Fuel Economy standards on energy consumption and the environment in South Korea. *Transportation Research Part D:*

Transport and Environment, 53, 321-333.

Xiong, C., Chen, X., He, X., Guo, W., & Zhang, L. (2015). The analysis of dynamic travel mode choice: a heterogeneous hidden Markov approach. *Transportation*, *42*(6), 985-1002.

Yang, J. (2004). Knowledge management opportunities for cycle time reduction. *International Journal of Innovation and Learning*, 1(2), 192-207

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.

Zimmermann, H. J. (2011). *Fuzzy set theory—and its applications*. New York, NY: Springer Science & Business Media.

Zolfaghari, A., Sivakumar, A., & Polak, J. (2013). Simplified probabilistic choice set formation models in a residential location choice context. *Journal of choice modelling*, 9, 3-13.

Abstract (Korean)

혁신에 의한 기술의 발전으로 다양한 제품이 시장에 출시되면서 소비자의 의사결정과정은 이전보다 훨씬 복잡해지고 있다. 이에 따라 한정된 정보처리능력을 가진 소비자는 구매 가능한 대안들 중에서, 속성을 평가하고 비교할 일부 대안을 선별하는 행동, 즉 선택집합을 구성하는 행동을 보이게 되었다. 소비자의 이와 같은 행동이 빈번해지면서, 혁신기술을 적용한 신제품의 확산에도 영향을 주게 되었다. 구체적으로, 소비자가 선택집합을 구성하는 과정에서 혁신기술을 활용한 신제품이 고려되지 않는 경우 신기술 확산의 양상은 기대한 바를 따르지 못하게 된다. 그러므로 신기술의 확산을 촉진하기 위해서는 소비자의 특성에 따른 선택집합 구성의 양상을 파악하고, 소비자가 선택집합 내에서 신기술을 고려할 수 있도록 하는 정책 및 마케팅 전략의 구성이 필요하다. 뿐만 아니라, 소비자의 실제 선택집합을 고려하여 신기술에 대한 소비자의 선호를 정교히 추정할 필요가 있다.

이에 본 연구에서는 소비자의 선택 집합 구성행동을 소비자의 선택모형에 반영하는 새로운 모형을 제안하고자 한다. 구체적으로, 소비자의 의사결정과정을 선택집합 구성 단계, 제품 선택 단계, 사용량 결정 단계의 3단계로 구성하고, 각 단계의 연계성을 표현할 수 있는 모형을 제안한다. 제안 모형을 통해서는 소비자의 선택집합 구성 행동이 대안의 선호에 미치는 영향을 직접적으로 확인할 수 있다. 뿐만 아니라, 기술 혁신에 따른 제품의 속성 변화 및 개인특성의 변화에 따른 선택집합 구성행동의 변화가 시장에 미치는 효과를

관측할 수 있다는 강점이 있다.

나아가 본 논문은 제안 모형을 자동차 시장에 실증 적용하여 시장 예측 상에서 소비자의 선택 집합 구성 단계를 고려하는 것이 얼마나 중요한지를 보였다. 실증 연구는 다음의 3가지로 구성된다. 1) 현시선호 데이터를 활용하여 소비자의 효용을 분석할 때 연구자가 소비자의 선택집합을 어떻게 가정하는가에 따라 연구결과가 어떻게 달라질 수 있는지를 살펴보았다. 2) 제안모형을 적용하여 자동차 구매 시 소비자의 선택 집합 구성 행동이 어떠한 요소들에 의해 영향을 받으며, 이에 따라 자동차 유형별 선호 어떻게 바뀌는지 확인하였다. 또한 소비자의 차량 주행거리가 소비자의 개인 특성 및 차량 유형에 따라 어떻게 달라지는지도 확인한다. 3) 추정된 결과를 활용하여 기술 혁신에 따른 자동차 속성의 변화, 소비자의 인식 변화에 따른 선택집합 구성 행동의 변화로 구성된 시나리오를 가정하고, 각 시나리오 하에서 자동차 시장의 변화와 그에 따른 파급효과를 분석함으로써 본 연구 모형이 시장 예측의 프레임워크로서 가지는 강점에 대해 확인한다.

주요어: 이산-연속 선택 모형; 소비자 선호; 의사결정과정의 휴리스틱스;
선택집합 구성 행동; 혁신의 확산

학 번: 2014-30287