Choice Data Generation using Usage Scenarios and Discounted Cash Flow Analysis

Ungki Lee

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology,

Daejeon, 34141, South Korea

lwk920518@kaist.ac.kr

Namwoo Kang*

Department of Mechanical Systems Engineering, Sookmyung Women's University,

Cheongpa-ro 47-gil 100, Yongsan-gu, 04310, Seoul, Korea

nwkang@sm.ac.kr

Ikjin Lee*

Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology,

Daejeon, 34141, South Korea

ikjin.lee@kaist.ac.kr

* Co-corresponding authors

Abstract

Discrete choice analysis with hierarchical Bayesian (HB) is one of the most popular methodologies to estimate heterogeneous customers' preferences. Although more choice data results in higher model accuracy, gathering enough choice data of target customers is often expensive and time consuming. This paper proposes a choice data generation method for commercial products whose expected money value is a key factor in consumer choice (e.g., commercial vehicles, financial product, etc.). Using individual usage scenario, a discounted cash flow (DCF) model is generated instead of a utility model, and discount rates of individual consumers are estimated instead of part-worths. Then, numerous synthetic choice data is generated using the DCF model with almost full factorial design, and finally discrete choice analysis with HB approach is conducted using a large amount of synthetic choice data. Preference estimation on hybrid courier truck conversion is used for a case study, and the results show that the usage scenario based HB model outperforms the traditional HB model.

Keywords: Discrete choice analysis, hierarchical Bayesian (HB), discounted cash flow (DCF) analysis, discount rate

1. Introduction

Consumers take into account various aspects of a product (e.g. aesthetic, functional, economic) when purchasing the product, and consumer decision making has attracted much interest in fields such as consumer research [1], operations management [2], design for market systems [3,4], environment [5], food studies [6–9], and health-care applications [10–12]. Discrete choice analysis is a statistical technique that utilizes choice data obtained through surveys to measure consumers' tradeoffs among products or services with various attributes [13]. Discrete choice analysis uses a utility model and assumes that the consumer chooses a product with a larger utility which is the sum of the part-worths that the consumer assigns to the product attributes [14–18]. The choice data needed for estimating part-worths can be in general collected through a survey performed by potential customers [19,20]. In discrete choice analysis, a series of choice questions that consist of a set of products with different attribute combinations are presented and the respondent is asked to choose the most preferred product option among other alternatives [21–23]. A hierarchical Bayesian (HB) approach, which provides a probability of the respondent choosing one product over others as a function of attributes, is used to estimate individual part-worths [24–27]. Then, intermediate attribute values are interpolated using a nature cubic spline to obtain a continuous individual-level utility model [28].

The performance of the discrete choice model depends strongly on the amount of choice data. Particularly, when target customers are specified to a certain group, obtaining choice data can be costly and time consuming, and sometimes it may not be possible to obtain enough data to perform discrete choice analysis. The research question of this study is how to improve the performance of the discrete choice model even in case of lack of choice data. This study confines the scope of the problem to commercial products that are defined as products used for business or investment for the purpose of earning profits. For commercial products, such as financial product, commercial vehicle, etc., consumers will make a reasonable decision considering not only their current investment, but also usage scenario, which is a scenario related to the gains or costs that the product will bring to the customers during the total period of use. In addition, personal experience and learning about risk affect the value of expected future cash flow and therefore have a significant impact on customers' decision [29–31]. In the case of general products, customers consider various functional and non-functional attributes; however, in the case of commercial products, under assumption that the customer considers the expected money value obtained by using the product as the top priority in product selection, we present the possibility that discounted cash flow (DCF) model created by usage scenario and DCF analysis can be applied to discrete choice analysis.

DCF analysis is a valuation technique used to estimate the attractiveness of an investment project and value of an enterprise by considering time-value of money (TVM), which is the concept that money available at the present time is worth more than the identical sum in the future [32–34]. DCF analysis utilizes discounted future cash flow, using a discount rate as the basis for the discounting, to derive present value which is a future sum of stream of cash flows converted to the current value [35,36]. The discount rate includes the cost of risky investments or uncertainty of future values [37]. Then, net present value (NPV), a measure of the difference between the present value of cash inflows and outflows over a period of time, can be calculated based on the usage scenario and discount rate, and the profitability of the investment can be determined [38]. Finally, DCF model can predict consumer's choice on a product by calculating NPV of a given product.

This study starts with the idea of applying the DCF model to discrete choice analysis instead of the utility model. Similar to the utility model where utility of choice options determines choice probability, the DCF model assumes that NPV of the choice options determines the choice probability. In addition, while the objective of the utility model is to estimate part-worths for individual, the objective of the DCF model is to estimate discount rate for individual. We assume that the discount rate varies depending on the individual's experience and propensity to risk. Thus, the discount rate is a criterion for each individual to convert the future cash flow into its present value, and also serves as a factor for determining investment of a given product.

However, the DCF model has a disadvantage in that it does not reflect the preference of other people unlike in the utility model since each DCF model is created by considering each individual only. Therefore, it is necessary to develop a method that can combine the DCF model and HB to improve this drawback by utilizing the advantage of HB shrinkage. Although there are many methods for such a combination, in this study, a method of generating the choice data via DCF model and performing the HB with the generated choice data is proposed. Since the DCF model is created using the usage scenario in addition to the original choice data, the synthetic choice data generated by the DCF model contains the information of the usage scenario that the original choice data does not have, and it is assumed that HB, which learns these synthetic choice data, can indirectly learn the usage scenario. Then, the HB discrete choice model created by the synthetic choice data generated using the DCF model is validated through a case study.

This study proposes a choice data generation method for commerical products. The poposed method consists of three steps: 1) DCF model generation by constructing a cash flow function based on the usage scenario and estimating the discount rate using the DCF analysis. The individual demographic data, such as the duration of use, frequency of use, and the utility of the product, are used as parameters in the usage scenario. The discount rate is obtained by calculating the individual NPV on choice options using the choice data as training data. 2) Design of numerous questionnaires by almost full factorial experiment and simulation of choice results using the DCF model. 3) Generation of synthetic choice data for HB, and part-worths estimation. The proposed method is applied to a case study of hybrid courier truck conversion which is the Korean Government's project [39,40] where three attributes in questionnaires are (1) conversion cost, (2) fuel efficiency improvement percentage, and (3) maintenance cost.

The remainder of this paper is organized as follows. Section 2 introduces the proposed choice data generation method using usage scenarios and DCF analysis. In Section 3, a case study on hybrid courier truck conversion and its results are presented. Finally, Section 4 concludes the study and suggests future research direction.

2. Proposed methods: DCF model and choice data generation

In the DCF model, the usage scenario contains the main information about the product selection of each individual and the choice data is used to find the individual discount rate that is not revealed in the choice data. The first proposed method is to make a DCF model that can calculate NPV and to predict consumer choice, and the second proposed method is to generate the choice data using the DCF model and to perform HB. The information flow of the proposed methods is presented in Fig. 1. The cash flow function that includes the cash flow expected from the product and the duration of the product used can be established by usage scenario consisting of individual demographic data related to product use. Then, each individual's discount rate can be found through optimization techniques using original choice data as training data. After obtaining discount rate of the individual, DCF model that can predict each individual's choice for given product options by calculating NPV is completed and almost full factorial questionnaire sets are simulated to generate synthetic choice data. Finally, the performance of discrete choice model can be improved using the HB approach with large amount of synthetic choice data. Some of the original choice data is used to estimate discount rate, while others are used to calculate hit rate, which is the rate at which the estimated consumer choice matches the actual survey results.

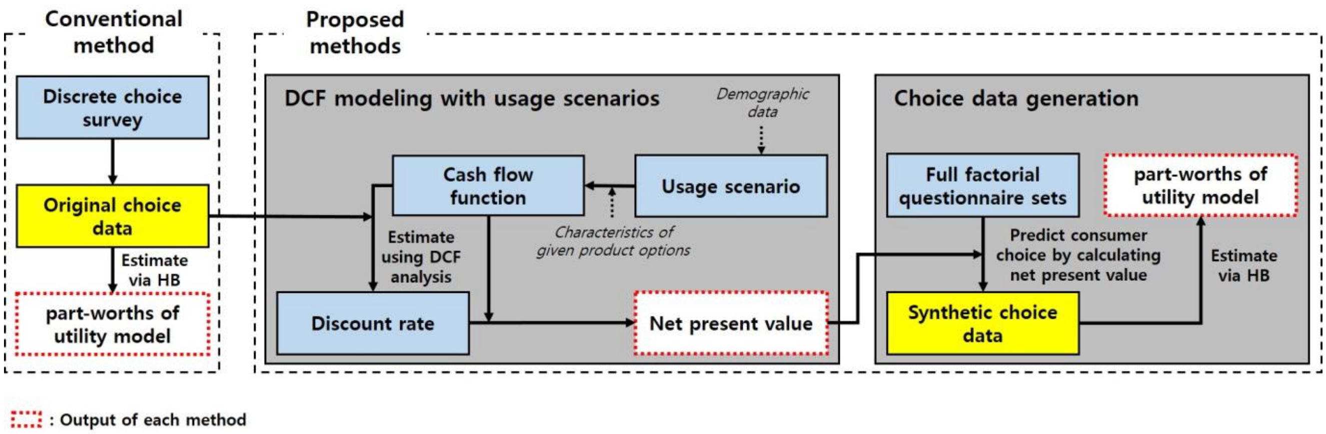


Fig. 1 Information flow of the proposed methods: DCF modeling and choice data generation using DCF model

2.1. DCF modeling with usage scenarios

In the case of commercial products where future cash flows are importantly considered, consumers decide whether they will purchase the product by comparing the benefits that are expected in the future with the amount that they are currently paying. Although the same product has the same initial cost and expected future benefit from the product, the value judgment for initial investment and future benefits differs depending on each customer's personal experience or learning about the risks that may arise in the future. If the expected future benefit is less than the initial investment cost, the consumer will not consider purchasing the product. Furthermore, customers will choose the product that brings the biggest benefit among given product options. Thus, to predict decision making of consumers when various products are given, it must be known how consumers value the current investment and future benefits gained from the product.

This section presents the DCF model that can predict the customer's choice on given product options. As shown in the DCF modeling with usage scenarios in Fig. 1, the individual demographic data, which are importantly considered when each customer chooses a product, are included in the usage scenario in the form of parameters and a cash flow function can be created based on the usage scenario and the characteristics of a given product option. DCF analysis converts a future sum of stream of cash flows into present value considering TVM, and estimates benefit that a consumer would receive from an investment. The discount rate, also known as the opportunity cost of capital and importantly applied for many economic behaviors such as a choice between a lump-sum or an annuity [41,42], reflects the risk of the investment and the uncertainty of future cash flow, and is responsible for discounting future cash flow in DCF analysis. Therefore, NPV, which indicates a net profit subtracting the required investment from the present value of the sum of future cash flows, can be calculated through DCF analysis and the attractiveness of the investment can be measured [43–45]:

$$\pi_{ij} = \sum_{t=1}^{N} \frac{CF_t(\mathbf{z}_j, \mathbf{s}_i)}{(1 + r_i)^t} - p_j \tag{1}$$

where π_{ij} is NPV of product j to individual i; t is the time of cash flow, N is the total period of use; CF_t is the cash flow function; $\mathbf{z}_j$ is characteristics of product j; $\mathbf{s}_i$ is usage scenario of individual i; r_i is discount rate of individual i; and p_j is price (or initial investment cost) of product j. Even if customers use the same product, the cash flow function can be changed depending on the usage scenario of the individual. For example, in the case of hybrid courier truck conversion, the expected benefit of fuel efficiency depends on individual's actual fuel mileage, average daily driving distance, and remaining period of use.

As aforementioned, the discount rate contains the risk of investment and the uncertainty of the future cash flow, therefore, the discount rate varies depending on the individual's experience and perception of the risk of investment, and serves as a measure to convert future cash flow to present value for each individual. For example, if an individual's discount rate is high, the expected future cash flow converted to present value will be small since the risk and uncertainty of investment is highly recognized, and the individual is unlikely to be attracted to the investment. Thus, after establishing the cash flow function and estimating individual discount rate, DCF modeling is completed and consumer's choice on given product options can be estimated by calculating NPV.

2.2. Discount rate estimation

The future cash flow expected from the product is determined by how useful the product is to each individual, therefore, the present value can be calculated using the individual's cash flow function and discount rate. By utilizing each individual's demographic data related to product use, the cash flow function that includes the cash flow expected from the product and the duration of the product used can be established. DCF model assumes that NPV of given product options determines choice probability:

$$P_{ij} = \frac{e^{\pi_{ij}}}{\sum_{j' \in J} e^{\pi_{ij'}}} \tag{2}$$

where P_{ij} is the probability that individual i chooses product j from a set of alternative products J. Then, each individual's discount rate can be found through the optimization technique using the individual's original choice data as training problem:

$$r_i = \underset{r_i}{\operatorname{argmax}} \sum_{j=1}^{n_j} \Phi_{ij} \log P_{ij} \tag{3}$$

where r_i is discount rate of individual i; Φ_{ij} is equal to 1 if individual i chooses product j, and 0 otherwise; and n_j is the number of product alternatives.

After creating the DCF model by estimating the discount rate with a part of the original choice data, the hit rate is calculated using the remaining validation data to present the accuracy of the DCF model, as in the case of HB discrete choice analysis.

2.3. Choice data generation

Since each DCF model is created based on only individual usage scenario and original choice data, the preferences of other people are not reflected, and thus the advantage of HB shrinkage can be used to overcome this drawback by performing HB with synthetic choice data generated by the DCF model. Therefore, this section suggests a method using the DCF model to generate synthetic choice data that can support HB and improve the performance of the discrete choice model. Because the DCF model includes each individual's usage scenario and the discount rate estimated by using usage scenario and original choice data, the DCF model is expected to accurately predict each individual's choice of product and the information of the usage scenario is contained in the synthetic choice data.

After completing the DCF model by establishing the cash flow function and finding the optimum discount rate of each individual, consumer's choice on given product options can be predicted by calculating NPV through the DCF model. Then the individual's choice on synthetic survey questionnaire sets can be estimated through simulation assuming that the respondent's judgment about the product options is contained in the DCF model and the survey not conducted by the respondent can be performed by the DCF model instead of the respondent. Therefore, large volume of choice data that is not actually generated from the respondent but extracted from the DCF model of the respondent can be created. In the synthetic survey questionnaire sets, each questionnaire consists of combinations of various product attribute levels that are drawn from random sampling and the sets are close to almost full factorial design. In the simulation, the DCF model is used to calculate the NPV of the given product options, including the none option, and the synthetic choice data is generated by selecting the option with the largest expected NPV. Then, for the given synthetic survey questionnaire sets, each individual DCF model performs a survey to generate synthetic choice data for each individual.

2.4. HB estimation

This section introduces the conventional HB approach and how to estimate the part-worths of the utility model with the synthetic choice data generated using the DCF model in the proposed method. Discrete choice analysis is a survey and statistical based marketing technique used to estimate the importance of each product attribute and to predict consumer decision making. Consumer choice data is needed to estimate consumer's choice for a product and can be gathered from questionnaire answered by potential consumers using the discrete choice analysis [46–49]. There are several multiple-choice questions in the questionnaire, and a set of designs with different combinations of various attributes is shown to respondents. Then, the respondents select the most preferred design, and none options can be selected when all alternative designs are not satisfactory.

Then, the HB approach, which consists of two levels, is used to quantify respondents' preference [50,51]. An upper level model is pooled across respondents and a lower individual-level model is treated within respondents [52]. The heterogeneity in the individual-level part-worths throughout the respondents' population is explained in the upper level model. Individual-level utility v_{ij}, which is the sum of part-worths of the designed product, can be defined as follows:

$$v_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \boldsymbol{\beta}_{ikl} z_{jkl} \tag{4}$$

where $\boldsymbol{\beta}_{ikl}$ represents the part-worth of the l-th level of the k-th attribute for the i-th individual, and z_{jkl} corresponds to a binary dummy number, which is equal to 1 if level l of the k-th attribute is selected for alternative j and 0 otherwise. An individual's part-worths $\boldsymbol{\beta}_i$ are assumed to be derived from a multivariate normal distribution, $\boldsymbol{\beta}_i \sim N(\boldsymbol{\theta}, \boldsymbol{\Lambda})$, where $\boldsymbol{\theta}$ is a mean vector of individual distributions, and $\boldsymbol{\Lambda}$ is the distribution's covariance matrix. At the lower level, choice probability, which is determined using mixed logit model, can be calculated as follows:

$$P_{ij} = \frac{e^{v_{ij}}}{\sum_{j' \in J} e^{v_{ij'}}} \tag{5}$$

which is similar to the probability of the i-th individual selecting option j from a set of alternatives J. Then, the optimal weight of the upper level and lower level models in estimating each individual's part-worth can be determined using the HB approach. Markov chain Monte Carlo (MCMC) is used to draw individual's part-worths. Though the achieved part-worth coefficients are discrete, interpolation of intermediate attribute values enables individual-level utility models to work with continuous attributes by using a nature cubic spline. In order to evaluate how accurately the utility model is established, the hit rate is calculated: some of the choice data is left as validation data and the utility model created from the remaining data is used to calculate the rate at which the utility model matches the consumer's choice contained in the validation data. For example, if the estimated utility model accurately predicts 3 out of the 5 questions of the respondent's choice, then the hit rate becomes 0.6.

In the case of the existing discrete choice model, the HB approach is performed with the original choice data obtained from the actual respondents, whereas the proposed method estimates the part-worths of the utility model by performing the HB approach with the synthetic choice data. Using the large amount of synthetic choice data generated from the DCF model, individual level part-worths of the utility model can be estimated through the conventional HB approach, and the performance of the discrete choice model can be enhanced, and therefore, the accuracy of consumer choice estimation can be improved. Since the synthetic choice data is generated from the DCF model, an individual's judgment of the value of the product attributes and the perception of the potential risks in the future can be reflected, thus it has the similar effect as a respondent actually responding to a large number of survey questionnaires.

3. Case study results

3.1. Case study on hybrid courier truck conversion

Eco-friendly vehicles are globally popular since they reduce transport-generated $\mathrm{CO_2}$ emissions and harmful impacts on the environment compared with conventional internal combustion engine vehicles [53–55]. The recent rise in oil prices and the need to purify air quality in cities have accelerated the advent of fuel-efficient, low pollutant emission power train technologies, and hybrid electric vehicles (HEVs) become significant [56,57]. Accordingly, converting conventional internal combustion engine vehicles to HEVs becomes an attractive option and research on the topic is actively studied and discussed today [58]. This paper focuses on applying the proposed method to converting small diesel courier trucks to HEVs, and the research subjects are truck operators in South Korea. In order to meet the needs of truck operators and to secure market demand when converting, consumer’s choice depending on conversion cost, fuel efficiency improvement, and maintenance cost that will be reflected in actual product design, should be investigated. The truck operators will decide whether to convert the truck considering conversion cost, which is the initial investment, and the maintenance cost that occurs every four years, and the fuel cost savings due to conversion. Therefore, a survey about hybrid courier truck conversion targeted to truck operators should be preceded to gather choice data and estimate consumer choices. However, it is expensive and difficult to gather the data because the surveyor has to chase and survey truck drivers individually.

In this study, 70 courier truck operators in South Korea are surveyed and the respondents are asked to answer 15 choice questions. The attributes in questionnaires are as follows: (1) conversion cost; (2) fuel efficiency improvement percentage; and (3) maintenance cost. Several multiple-choice questions are presented in the questionnaire, and a set of designs with combinations of various levels of attributes (shown in Table 1) is presented to respondents. The respondents then select the most preferred design. When no satisfactory designs exist, the respondents may pick none of the options. In addition, demographic data of truck operators such as annual income, actual fuel mileage, daily driving distance, current mileage, and target mileage are also included in the survey. The mean and standard deviation of annual income, actual fuel mileage, daily driving distance, current mileage, and target mileage are \$33,036 and \$11,220, 16.0 mpg and 2.6 mpg, 28.6 mi and 15.1 mi, 54,006 mi and 43,084 mi, and 124,948 mi and 41,337 mi, respectively. After gathering choice data from the survey, the HB approach is used to find the individual level part-worths, and the importance of each attribute can be derived.

Table 1 Attribute levels and importance

Attributes	Level				Importance (%)
	1	2	3	4	
Conversion cost (\$)	1,000	2,500	4,500	6,000	61.3
Fuel efficiency improvement percentage (%)	15	20	25	30	22.5
Maintenance cost (\$/4 years)	350	450	550	650	16.2

The equation of DCF model for hybrid courier truck conversion can be given as follows:

$$\pi_{ij} = \sum_{t=1}^{N} \frac{A_{i,t} - F_{ij,t}}{(1+r_i)^t} - \sum_{t=1}^{N/4} \frac{M_{j,t}}{(1+r_i)^t} - p_j \tag{6}$$

where $F_{ij,t} = \frac{D_{i,t}}{FM_{i,t}} \times Fuel\ price \times \left(1 - \frac{I_j}{100}\right) \times 365$

where N is the total period of use; $A_{i,t}$ and $M_{j,t}$ indicate annual income of individual i and maintenance cost of conversion option j at time t, respectively; $F_{ij,t}$ is fuel cost of individual i with conversion option j at time t; $D_{i,t}$ and $FM_{i,t}$ are daily driving distance and actual fuel mileage of individual i at time t, respectively; I_j is fuel efficiency improvement percentage of conversion option j; and p_j is the conversion cost of conversion option j.

3.2. Comparison and discussion of model performance

To present and investigate the novelty of the proposed methods, three different methods that estimate consumer choice are compared: (1) Method 1 is the conventional method shown in Fig. 1 that estimates part-worths using the utility model for HB discrete choice analysis; (2) Method 2 is the first proposed method using DCF model that estimates the discount rate using usage scenario and choice data, and predicts consumer choice by calculating NPV of given product options (DCF modeling with usage scenarios in Fig. 1); and (3) Method 3 is the second proposed method that improves the accuracy of consumer choice estimation by supporting HB discrete choice analysis with the synthetic choice data generated using the DCF model (choice data generation in Fig. 1). To assess the accuracy of the consumer choice estimation of the three methods that use different models, 5 choice data chosen from the respondent's 15 choice data are used to calculate hit rate as described in Section 2.4. Since the hit rate may vary depending on the set of test data, the mean value of the hit rates obtained from 300 sets of survey questions and test data is used in this paper.

Fig. 2 presents the hit rate results of consumer choice estimation depending on the number of respondents and survey questions using only the original choice data as in the conventional consumer choice estimation that utilizes utility model for discrete choice analysis (Method 1). Also the mean value of hit rates obtained from 200 sets of respondents is used in case that the number of respondents is less than 70. As the number of survey questions becomes small, hit rate tends to decrease because individual's choice data used to estimate individual's part-worths is not sufficient. In addition, due to the nature of the HB approach, the individual level part-worths is estimated not only using individual data but also leveraging the population data; therefore, as the number of respondents reduces, the population data becomes smaller and thus the hit rate decreases.

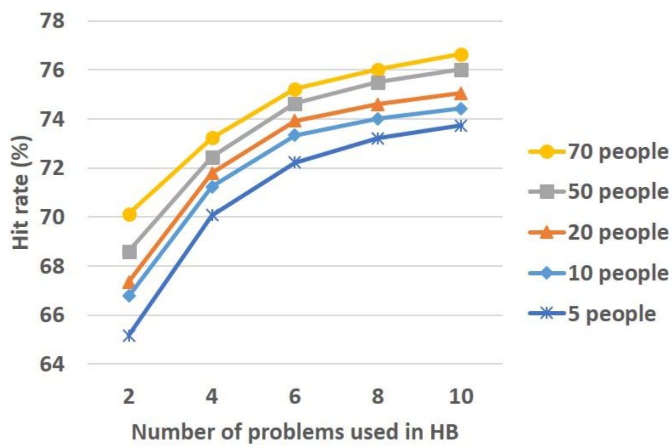
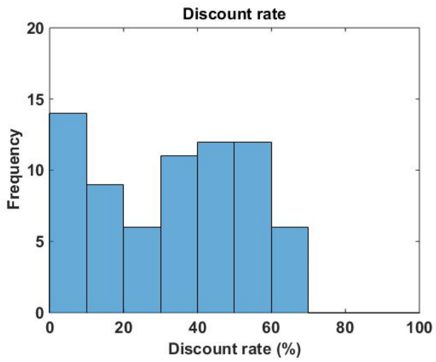


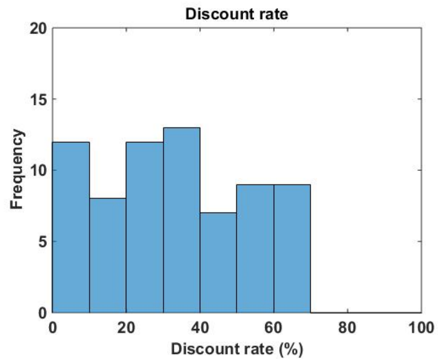
Fig. 2 Hit rate of consumer choice estimation based on actual choice data (Method 1)

In Method 2, DCF model that predicts consumer choice can be completed after finding the optimum discount rate through the optimization technique using the individual's choice data as training problem, as described in Section 2.2. We solve the optimization problem in Eq. (3) and find each respondent's optimum discount rate using genetic algorithm (GA) for global search and sequential quadratic programming (SQP) for local search. The GA population size is 300 and computation requires 1 hour on average using a standard desktop (Intel i7 6900 CPU @ 3.20 GHz and 64.0 GB RAM).

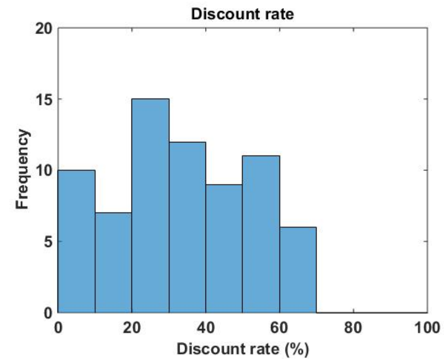
Fig. 3 and Table 2 show 70 respondents' optimum discount rate results and distributions depending on the number of training problems. As shown in Table 2, the mean value of the optimum discount rates tends to increase as the number of training problems increases. A large variance in the optimum discount rate shows that respondents' assessment of future gains or outcomes expected from the product are highly heterogeneous.



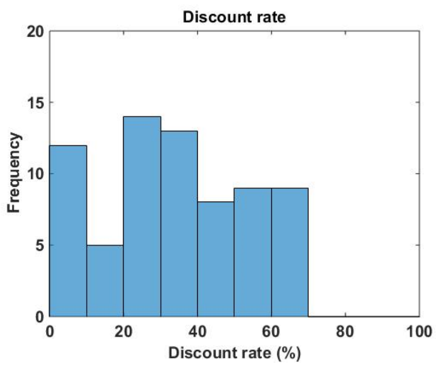
(a) 2 training problems



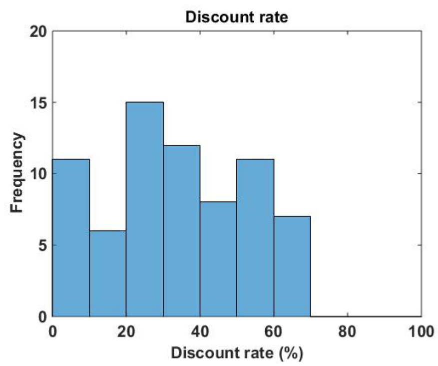
(b) 4 training problems



(c) 6 training problems



(d) 8 training problems



(e) 10 training problems

Fig. 3 Optimum discount rate distributions depending on the number of training problems (70 people)

Table 2 Mean and standard deviation of optimum discount rates

	Number of training problems				
	2	4	6	8	10
Mean	32.7%	32.8%	32.9%	33.6%	33.7%
(Std)	(20.3%)	(20.1%)	(18.7%)	(19.6%)	(19.4%)

After finding the optimum discount rate, DCF model can predict each respondent's choice on given product options based on NPV calculation, and the hit rate of the DCF model, which indicates how accurately the DCF model predicts the choice of respondent, is shown in Fig. 4 (Method 2). The results show that the DCF model predicts the choice of respondents more accurately as the number of training problems used to find the optimal discount rate increases. This indicates that the more the training data, the more accurately the discount rate of each respondent can be estimated, and therefore, the DCF model could more accurately represent the respondent's value judgment of the future profits resulted from personal experience and perception of the risk of investment. This is because, even if optimization is performed using original choice data as training data, there may exist several discount rate candidates when training data is small. However, as the training data increases, the range of discount rates becomes narrower and approaches the discount rate that best represents each individual's judgment. Compared with the hit rate results of Method 1 using the choice data of 70 respondents, the hit rate of Method 2 tends to be higher for all number of training problems (4.4% point higher on average). In the case of Method 1 using the utility model for discrete choice analysis, the performance of the utility model is affected by the number of respondents, while the proposed Method 2 is not affected by the number of respondents since the DCF model is created for each respondent and does not affect the model of other respondents.

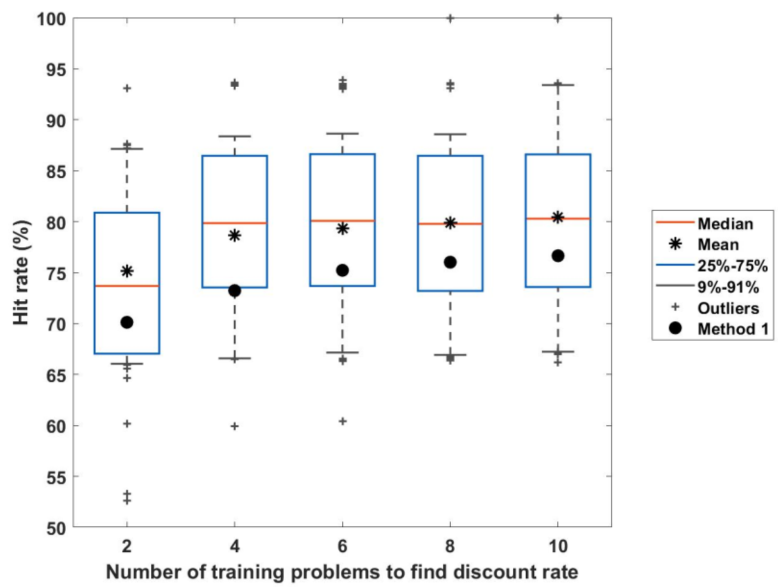
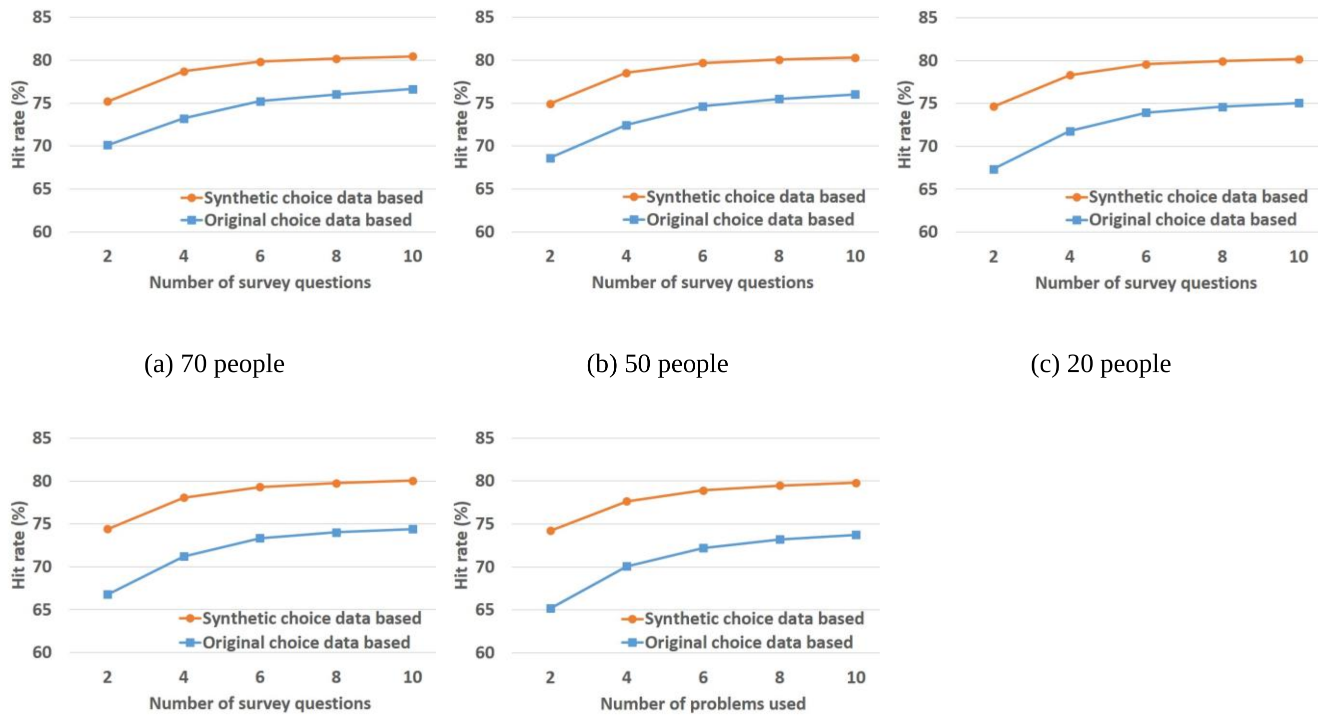


Fig. 4 Hit rate of DCF model depending on the number of training problems to find discount rate (70 people)

As explained in Section 2.3, HB is performed with synthetic choice data created by DCF model to solve the problem of Method 2 that does not reflect the preferences of other customers. Using the DCF model, synthetic choice data can be generated by performing survey questionnaires that consist of combinations of various product attribute levels. In Method 3, a large amount of synthetic choice data generated from the DCF model are used to estimate the part-worths through HB approach. Comparisons of the hit rate results of consumer choice estimation obtained using only original choice data (Method 1) and using the synthetic choice data (Method 3) are shown in Fig. 5.



(a) 70 people (b) 50 people (c) 20 people

(d) 10 people (e) 5 people

Fig. 5 Comparisons of the hit rate of the estimated consumer choice obtained using only original choice data (Method 1) and using the synthetic choice data (Method 3)

In Method 3 using the synthetic choice data, the number of survey questionnaires used is equal to the number of training problems in finding the optimum discount rate. Almost full factorial design sets of survey questionnaires are used to generate synthetic choice data and the hit rate is converged for a given number of survey questions and respondents in Method 3. The result of Method 3 is highly robust to the number of respondents as compared to Method 1 because the amount of synthetic choice data for each respondent is already large so that a large amount of choice data can assist even when the number of respondents is small. The hit rate is significantly affected by the number of training problems to find discount rate, and is slightly influenced by the number of respondents. As the training data used to find the optimum discount rate increases, the DCF model becomes more precise and more accurate synthetic choice data can be generated, therefore, more accurate consumer choice estimation becomes possible. The results show that the overall hit rate is improved when the consumer choice is estimated using the synthetic choice data generated by the DCF model since a large amount of synthetic choice data play a supporting role not only for the data of each respondent and also for population data. The hit rate increases drastically when the number of respondents and survey questionnaires used is small (insufficient choice data), but the hit rate is also improved in the opposite case when the choice data is relatively sufficient. Table 3 shows how the accuracy of consumer choice estimation is improved when estimating consumer choice based on the synthetic choice data compared to estimating based on original choice data only.

Table 3 Accuracy improvement in consumer choice estimation

Number of people	Number of survey questions				
	2	4	6	8	10
70	7.24%	7.06%	6.13%	5.49%	4.96%
50	9.23%	8.41%	6.78%	6.07%	5.64%
20	10.84%	9.10%	7.64%	7.14%	6.82%
10	11.43%	9.60%	8.17%	7.75%	7.57%
5	13.89%	10.82%	9.28%	8.56%	8.23%

In addition, the results in Fig. 5 show that the hit rate obtained using the synthetic choice data is improved as the number of respondents increases since HB approach leverages population data to estimate individual level part-worths and the amount of population data increases as the number of respondents increases [59]. Because the synthetic choice data generated from Method 2 is used as choice data for HB approach in Method 3, the hit rate results of Method 3 tend to be affected by the hit rate results of Method 2. The hit rate of Method 3 tends to be higher than that of Method 2 when the number of people is large: 0.67% point, 0.37% point, and 0.16% point lower when the number of people is 5, 10, and 20, respectively; and 0.02% point and 0.19% point higher when the number of people is 50 and 70, respectively. Both of the above statements indicate that as the number of questionnaires used and respondents increases, hit rate increases because more accurate synthetic choice data can be generated, and the amount of total choice data increases, at the same time.

To further explore the characteristics of the DCF model, comparison of the importance of product attributes when HB is performed with original choice data and synthetic choice data is presented in Table 4. When carrying out HB with synthetic choice data generated by the DCF model, the importance of conversion cost increases (from 61.3% to 75.1%) and the importance of fuel efficiency improvement percentage decreases (22.5% to 16.8%). This indicates that the DCF model is relatively resistant to initial investment costs and is negative for future gains. In the survey, people are involved in errors originating from being emotional and irrational: since the survey question is focused on fuel efficiency improvement, customers tend to concentrate on this issue and consider it more important. However, in the case of the DCF model, usage scenarios, which are accurately quantified by each individual, are used so that more objective judgment without bias can be made.

Table 4 Comparison of the importance of product attributes

	Attributes		
	Conversion cost	Fuel efficiency improvement percentage	Maintenance cost
Original choice data based	61.3%	22.5%	16.2%
Synthetic choice data based	75.1%	16.8%	8.1%

4. Conclusion

The presented work suggests a method that predicts the consumer's choice on a product by using individual usage scenario and discount rate estimated by DCF analysis to calculate NPV for given product options. The DCF model that can predict consumer choice can be established after finding the optimum discount rate, which indicates the respondent's evaluation on the future cash flows, through the optimization technique based on the cash flow function and using the respondent's choice data as the training problems. The cash flow function depends on the characteristics of product and the usage scenario of each individual. In addition, a method that generates synthetic choice data for HB estimation using the DCF model and improves the accuracy of consumer choice prediction is proposed in this paper. The case study results show that predictive power of the DCF model is higher than the conventional discrete choice analysis using utility model and the overall hit rate of estimated consumer choice is improved when using synthetic choice data obtained from the DCF model for HB approach. The number of questionnaires and respondents affect the accuracy of the consumer choice estimation in terms of the number of training problems used to obtain the optimum discount rate and the amount of population data, respectively.

One of the main contributions of this paper is that a model is suggested to predict the consumer's choice of a commercial product by considering individual usage scenario and estimating the discount rate not revealed in the choice data. In addition, higher accuracy of the consumer choice estimation can be achieved even in the case of lack of choice data caused by small number of questionnaires and respondents. As the number of questionnaires increases, respondent becomes tired of repeated choice questions and the response may be distorted. Furthermore, conducting surveys on potential customers is time consuming and costly in some cases. This study therefore has the advantage of enabling accurate consumer choice estimation with each respondent's demographic data and a small amount of questionnaires although the number of respondents is not sufficient.

The existing consumer choice estimation considers only the current response of the respondent, but the suggested method considers an individual's evaluation on future cash flows through extracting the discount rate. That is, by adding the discount rate to the consumer choice estimation, the time axis can be considered and more accurate estimation can be made. Future work can calibrate the results of estimated consumer choice obtained from original choice data and synthetic choice data. An accurate estimate of the population that provides information to HB approach when the number of respondents is small, and more case studies in finance etc. also can be added in future work.

5. Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2017R1C1B2005266) and a grant (17TLRP-C135446-01, Development of Hybrid Electric Vehicle Conversion Kit for Diesel Delivery Trucks and its Commercialization for Parcel Services) from Transportation & Logistics Research Program (TLRP) funded by Ministry of Land, Infrastructure and Transport of Korean government.

6. References

[1] Bettman, J. R., Luce, M. F., and Payne, J. W., 1998, "Constructive consumer choice processes," J. Consumer Res., **25**(3), pp. 187–217.
[2] Karniouchina, E.V., Moore, W. L., van der Rhee, B., and Verma, R., 2009, "Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research," Eur. J. Oper. Res., **197**(1), pp. 340–348.
[3] Frischknecht, B. D., Whitefoot, K., and Papalambros, P. Y., 2010, "On the suitability of econometric demand models in Design for Market Systems," J. Mech. Des., **132**(12), pp. 57–68.
[4] Kang, N., 2014, "Multidomain demand modeling in design for market systems," Ph.D. thesis, University of Michigan, Ann Arbor, MI.
[5] Alriksson, S., and Öberg, T., 2008, "Conjoint analysis for environmental evaluation: a review of methods and applications," Environ. Sci. Pollut. Res., **15**(3), pp. 244–257.
[6] Halbrendt, C. K., Wirth, E. F., and Vaughn, G. F., 1991, "Conjoint analysis of the mid-Atlantic food-fish market for farm-raised hybrid striped bass," S. J. Ag. E., **23**(1), pp. 155–163.
[7] Darby, K., Batte, M. T., Ernst, S., and Roe, B., 2008, "Decomposing local: a conjoint analysis of locally produced foods," Am. J. Ag. E., **90**(2), pp. 476–486.
[8] Annunziata, A., and Vecchio, R., 2013, "Consumer perception of functional foods: a conjoint analysis with probiotics," Food Qual. Prefer., **28**(1), pp. 348–355.
[9] Caputo, V., Scarpa, R., Nayga, R. M., and Ortega, D. L., 2018, "Are preferences for food quality attributes really normally distributed? An analysis using flexible mixing distributions," J. Choice Model., **28**, pp. 10–27.
[10] Bridges, J. F. P., Kinter, E. T., Kidane, L., Heinzen, R. R., and McCormick, C., 2008, "Things are looking up since we started listening to patients: recent trends in the application of conjoint analysis in health 1970–2007," Patient, **1**(4), pp. 273–282.
[11] Marshall, D., Bridges, J. F. P., Hauber, A. B., Cameron R., Donnally, L., and Fyie, K., 2010, "Conjoint analysis

applications in health – how are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008," Patient, **3**(4), pp. 249–256.
[12] Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., and Regier, D. A., 2011, "Conjoint analysis applications in health – a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force," Value Health, **14**(4), pp. 403–411.
[13] Green, P., and Srinivasan, V., 1990, "Conjoint analysis in marketing research: new developments and directions," J. Mark., **54**(4), pp. 3–19.
[14] Michalek, J. J., Feinberg, F. M., and Papalambros, P. Y., 2005, "Linking marketing and engineering product design Decisions via analytical target cascading," Journal of Product Innovation Management, **22**(1), pp. 42–62.
[15] Lewis, K. E., Chen, W., Schmidt, L. C., and Press, A., 2006, "Decision making in engineering design," ASME Press, New York.
[16] Raghavarao, D., Wiley, J. B., and Chitturi, P., 2011, "Choice-based conjoint analysis models and designs," CRC Press, New York.
[17] Kang, N., Feinberg, F. M., and Papalambros, P. Y., 2013, "A framework for enterprise-driven product service systems design," Proceedings of the 19th International Conference on Engineering Design, Seoul, Korea, Aug 4-Aug 7, ISBN: 978-1- 904670-47-6.
[18] Wong, T., Brownstone, D., Bunch, D. S., 2019, "Aggregation biases in discrete choice models," J. Choice Model., **31**, pp. 210–221.
[19] Kang, N., Ren, Y., Feinberg, F. M., and Papalambros, P. Y., 2016, "Public investment and electric vehicle design: a model-based market analysis framework with application to a USA-China comparison study," Des. Sci., **2**, p. e6.
[20] MacKerron, G., 2011, "Implementation, implementation, implementation: old and new options for putting surveys and experiments online," J. Choice Model., **4**, pp. 20–48.
[21] Louviere, J. J., and Woodworth, G., 1983, "Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data," J. Mark., **20**, pp. 350–367.
[22] DeSarbo, W. S., Ramaswamy, V., and Cohen, S. H., 1995, "Market segmentation with choice-based conjoint analysis," Market. Lett., **6**(2), pp. 137–148.
[23] Halme, M., and Kallio, M., 2011, "Estimation methods for choice-based conjoint analysis of consumer preferences," Eur. J. Oper. Res., **214**(1), pp. 160–167.
[24] Train, K., 2001, "A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit," Paper Presented in University of California, Berkeley, pp. 1–13.
[25] Rossi, P., Allenby, G., and McCulloch, R., 2005, "Bayesian statistics and marketing," Wiley, New York.
[26] Orme, B., 2009, "The CBC/HB system for hierarchical Bayes estimation version 5.0 technical paper," Technical Paper Series, Sawtooth Software, Sequim, WA.
[27] Hein, M., Kurz, P., and Steiner, W. J., 2019, "On the effect of HB covariance matrix prior settings: A simulation study," J. Choice Model., **31**, pp. 51–72.
[28] Kang, N., Feinberg, F. M., and Papalambros, P. Y., 2017, "Autonomous electric vehicle sharing system design," J. Mech. Des., **139**(1), p. 011402.
[29] Botsch, M., and Malmendier, U., 2015, "Inflation experiences and contract choice—evidence from residential mortgages," Work Pap., Bowdoin Coll.
[30] Kuchler, T., and Zafar, B., 2015, "Personal experiences and expectations about aggregate outcomes," Fed. Reserve Bank N.Y., New York, Staff Rep. 748.
[31] Malmendier, U., and Nagel, S., 2016, "Learning from inflation experiences," Q. J. Econ., **131**(1), pp. 53–87.
[32] Myers, S. C., 1984, "Finance theory and financial strategy. Interfaces, **14**, pp. 126–137.
[33] Wee, H. M., and Law, S. T., 2001, "Replenishment and pricing policy for deteriorating items taking into account the time value of money," Int. J. Prod. Econ., **71**(1–3), pp. 213–220.
[34] Kruschwitz, L., and Löffler, A., 2005, "Discounted cash flow," Wiley, Chichester.
[35] Ross, S. A., 1995, "Uses, abuses and alternatives to the net present value rule," Finan. Manage., **24**(3), pp. 96–102.
[36] Weiss, E., and Majkuthová, S., 2005, "Discounted cash flow (DCF) assessment method and its use in assessment of a producer company," Metalurgija, **45**(1), pp. 67–71.
[37] Brealey, R. A., Myers, S. C., Marcus, A. J., Maynes, E. M., and Mitra, D., 2006, "Fundamentals of corporate finance," McGraw-Hill. Ryerson, Toronto.
[38] Srivastava, R. K., Shervani, T. A., and Fahey, L., 1998, "Market-based assets and shareholder value: a framework for analysis," J. Mark., **62**(1), pp. 2–18.
[39] Ministry of Land, Infrastructure and Transport (MOLIT), 2016, "Report on domestic and international development

status of eco-friendly hybrid diesel-electric trucks optimized for parcel delivery service," Korea Agency for Infrastructure Technology Advancement, KAIA.
[40] Koh, S. R., Hur, S. H., and Kang, N., 2019, "Feasibility study on the Korean government's hybrid conversion project of small diesel trucks for parcel delivery services," J. Clean. Prod., **232**, pp. 559–574.
[41] Warner, J. T., and Pleeter, S., 2001, "The personal discount rate: evidence from military downsizing programs," Am. Econ. Rev., **91**(1), pp. 33–53.
[42] Stiglitz, J., 1982, "The rate of discount for benefit-cost analysis and the theory of second-best," R. Lind (ed.), Discounting for Time and Risk in Energy Policy, John Hopkins University Press, Baltimore.
[43] Smith-Daniels, D. E., and Smith-Daniels, V. L., 1987, "Maximizing the net present value of a project subject to materials and capital constraints," J. Oper. Manage., **7**, pp. 33–45.
[44] Remer, D. S., and Nieto, A. P., 1995, "A compendium and comparison of 25 project evaluation techniques. Part 1. Net present value and rate of return methods," Int. J. Prod. Econ., **42**(1), pp. 79–96.
[45] Sanchez Fernandez, E., Bergsma, E. J., De Miguel Mercader, F., Goetheer, E. L.V., and Vlugt, T. J. H., 2012, "Optimisation of lean vapour compression (LVC) as an option for post-combustion CO2 capture: net present value maximisation," Int. J. Greenh. Gas Control, **11**, pp. S114–S121.
[46] Chrzan, K., and Orme, B., 2000, "An overview and comparison of design strategies for choice-based conjoint analysis: Sawtooth Software," Technical Paper Series, Sawtooth Software, Sequim, WA.
[47] Netzer, O., Toubia, O., Bradlow, E., Dahan, E., Evgeniou, T., Feinberg, F., Feit, E., Hui, S., Johnson, J., Liechty, J., Orlin, J., and Rao, V., 2008, "Beyond conjoint analysis: advances in preference measurement," Market. Lett., **19**(3), pp. 337–354.
[48] Train K., 2009, "Discrete choice methods with simulation," Cambridge university press.
[49] Lee, U., Kang, N., and Lee, I., 2019, "Selection of optimal target reliability in RBDO through reliability-based design for market systems (RBDMS) and application to electric vehicle design," Struct. Multidisc. Optim., pp. 1–15.
[50] Allenby, G. M., Arora, N., and Ginter, J. L., 1995, "Incorporating prior knowledge into the analysis of conjoint studies," J. Marketing Res., **32**(2), pp. 152–162.
[51] Lenk, P. J., DeSarbo, W. S., Green, P. E., and Young, M. R., 1996, "Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs," Market. Sci., **15**(2), pp. 173–191.
[52] Kruschke, J. K., 2013, "Bayesian estimation supersedes the t test," J. Exp. Psychol. Gen., **142**, pp. 573–603.
[53] Tilagone, R., Venturi, S., and Monnier, G., 2006, "Natural gas - an environmentally friendly fuel for urban vehicles: the SMART demonstrator approach," Oil gas sci. technol., **61**(1), pp. 155–164.
[54] Situ, L., 2009, "Electric vehicle development: The past present and future," Proc. 3rd Int. Conf. PESA, pp. 1–3.
[55] Hamada, K., Nagao, M., Ajioka, M., and Kawai, F., 2015, "SiC-emerging power device technology for next-generation electrically powered environmentally friendly vehicles," IEEE Trans. Electron. Devices, **62**(2), pp. 278–285.
[56] Fontaras, G., Pistikopoulos, P., and Samaras, Z., 2008, "Experimental evaluation of hybrid vehicle fuel economy and pollutant emissions over real-world simulation driving cycles," Atmos. Environ., **42**(18), pp. 4023–4035.
[57] Gallagher, K. S., and Muehlegger, E., 2011, "Giving green to get green? Incentives and consumer adoption of hybrid vehicle technology," J. Environ. Econ. Manag, **61**(1), pp. 1–15.
[58] EERE, 2018, "Hybrid and plug-In electric vehicle conversions," U.S. Department of Energy, https://www.afdc.energy.gov/vehicles/electric_conversions.html
[59] Orme, B., and Howell, J., 2009, "Application of covariates within Sawtooth software's CBC/HB program: theory and practical example," Technical Paper Series, Sawtooth Software, Sequim, WA.