Master's Dissertation in Engineering

Vehicle Fuel-type Preference Related Elements, Comparing their Importance Depending on Each Fuel-type

자동차 연료유형 선호도와 관련된 요인들의
연료유형별 중요도 차이 비교

February 2023

Graduate School of Seoul National University
Technology Management, Economics, and Policy Program
Mina Kim

Vehicle Fuel-type Preference Related Elements, Comparing their Importance Depending on Each Fuel-type

지도교수 구윤모

이 논문을 공학석사학위 논문으로 제출함
2023 년 2 월

서울대학교 대학원
협동과정 기술경영경제정책 전공
김민아

김민아의 공학석사학위 논문을 인준함
2023 년 2 월

위 원 장 이 종 수 (인)

부위원장 구 윤 모 (인)

위 원 최 현 홍 (인)

Abstract

Vehicle Fuel-type Preference Related Elements, Comparing their Importance Depending on Each Fuel-type

Mina Kim
Technology Management, Economics, and Policy Program
The Graduate School
Seoul National University

This study aims to present a new method for analyzing conjoint survey data to understand the differences in preference-related elements (PREs) depending on four major types of vehicle fuel: gasoline, diesel, electric, and hydrogen. PREs include vehicle features, such as price and class, and customer features, such as age and gender. Previous studies have been focused on discovering the list of PREs. However, comparing the importance of each PRE on preference toward certain vehicle fuel-type has yet to be sufficiently studied. Understanding the relative importance of PREs is required to make effective strategies and policies, especially for efficient target segmentation. In addition, this study included vehicle status quo information as part of PREs. Vehicle status quo, the current status of owned vehicles, such as the number of vehicles owned and their fuel-types, classes, and purchase price, is scarcely studied considering its importance. The vehicle status quo is the customer's answer to the previous vehicle purchasing process and includes every unrevealed corner of the process. Therefore, it is inevitably core information to be included, however challenging because of the scarcity of adequate data and correlation with other personal features. In this study, the vehicle status quo data is included in our analysis, which was possible because of the application of the random forest classifier (RFC). The non-rigidity of the RFC enables the highly correlated features to be analyzed for their importance. As the result, this study showed that the PREs' ranking of importance differs depending on the vehicle fuel-type. Furthermore, we analyzed the non-linear relationship between the PRE and preference towards each vehicle fuel-type with partial dependence plots.

Keywords: Vehicle Fuel-type Preference, Zero-emission Vehicle, Target Segmentation, Conjoint Survey, Machine Learning, Shapley Additive Explanation

Student Number: 2021-21881

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background

Throughout the history of the vehicle industry, considerable research has been conducted on new types of vehicle fuel. Since the emergence of gasoline, which marked the beginning of the internal combustion engine, several other fuel-types have been developed. Most were fossil fuels, such as diesel, liquefied petroleum, and compressed natural gas. However, as governments have attempted to achieve the Paris Agreement goals, zero-emission vehicles (ZEVs), including battery electric vehicles (BEVs) and fuel-cell electric vehicles (FCEVs) (Gota et al., 2018; Logan et al., 2020) have been commercialized. Hence, electricity and hydrogen are the latest emergent fuel-types for cars.

Along with the emergence of new vehicle fuel-type options, the difference of preference towards each vehicle fuel-type depending on customers was recognized. For instance, some people preferred diesel vehicles over gasoline because of their relatively cheaper fuel cost. Some preferred electric vehicles but only the BEVs but not the FCEVs, due to their insufficient refueling infrastructure. However, as much as their preferences differed, the reasons of their preference also differed. Understanding of the influence of features became vital for both the vehicle manufacturers and the governments trying to increase the market share of ZEV.

1.2 Research Objectives

Along with the introduction of new vehicle fuel-types, numerous studies on transition policies have been conducted. Subsidies are arguably one of the most widely used inductive policies for new technological products, including vehicles. However, in many cases where economical resources are limited, establishment of the most appropriate ratio of investment in the subsidy and infrastructure is required (Penna & Geels, 2012). Therefore, there have been many studies on the ZEV subsidy policy and infrastructure investment (Köhler et al., 2010). In addition, numerous studies have been conducted on car owners' features, mainly focusing on basic socioeconomic information or their driving patterns (Berigai Ramaiaha et al., 2018).

Despite the number of high-impact studies on investment in infrastructure for new fuel-type vehicles and the preference of car owners for certain types of vehicle fuel, research on the differences in the importance of preference-related elements (PREs) among the fuel-types is insufficient. In other words, previous studies focused on revealing "what the elements are" rather than "which elements must be considered relatively more important." The importance of each PRE could differ depending on each fuel-type. Although it is a simple concept, it is crucial because the ranking of PRE importance determines the policy and its priority. PREs would be the vehicle's features, such as the price, class, infrastructure, as well as the customer's personal information.

Understanding the PREs of each vehicle fuel-type is essential for building efficient and reasonable policies, particularly when the government is willing to persuade car owners to purchase specific options (Hoen & Koetse, 2014). Moreover, implementing a practical approach requires the determination of the type and intensity of the plan and the target customer group. Therefore, the element relatedness of the preference for each vehicle fuel-type is necessary to provide critical guidelines for both the segmentation procedure and plan building of policy implementation.

In this study, we assume that the relationship between preferences for each fuel-type, i.e., gasoline, diesel, electric, and hydrogen, and elements, including the car owners' demographic, vehicle status quo data, and the vehicle option information, are diverse. As Saarela and Jauhiainen (2021) shown, the feature importance differ depending on the technique. The conventional method to analyze the conjoint survey data is logistic regression. However if there are highly correlated features, logistic regression might just arbitrary choose one of those (Saarela & Jauhiainen, 2021). Therefore, a new method to analyze the data is in need.

1.3 Research Outline

In this study, as new research framework for new products with a limited amount of data, the ways to gain implications from various perspectives using survey data will be introduced. Survey is required to make demand predictions on new technologies, since their historical data do not exist. In cases when the products include various features, which is in most cases, conjoint survey is widely used to find the trade-offs between the features. For instance, a vehicle with outstanding performance, wonderful design, prestigious brand, and most importantly in affordable price is obviously the most preferable. Revealing such a simple tendency of each element is not meaningful in most of the cases, but the trade-off is.

The remainder of this paper is organized as follows. The Chapter 2 presents a detailed literature review on the elements related to preferences for vehicle-fuel-type and the methodologies used to discover them. The methods used in this study, such as the random forest classifier (RFC) and the Shapley value, are also introduced in this chapter. In Chapter 3, the model's dataset and hyperparameters are explained. The analysis results are presented in Chapter 4. Finally, the paper is concluded with Chapter 5, including the main contributions and the limitations of this research.

Chapter 2. Literature Review

There have been many studies on the elements affecting the choice of ZEV. In Section 2.1, the PREs studied in previous studies are discussed. Subsequently, in Section 2.2, the methodology used in this research is introduced and compared with the conventional method, the discrete choice model (DCM).

2.1 Vehicle Fuel-type Preference Elements

The PREs of vehicle fuel-type are either of the two; vehicle features and customer's personal features. Refueling or recharging infrastructure is considered as the vehicle features since it is determined by the fuel-type of the vehicle. On the other hand, vehicle status quo is considered as the customer's personal features, since it is part of each individual's information. The studies covering external factors, such as Contestabile et al. (2017) are separated from the vehicle properties.

Table 1. Studies in vehicle fuel-type preference-related elements and tendency

Authors	Year	Objectives	Tendency Studied
Ahmed	2013	Vehicle Properties	Cost of Hybrid, plug-in hybrid and electric vehicles
Junquera	2016	Vehicle Properties	Charging time and driving range of EV being critical

Byun	2018	Vehicle Properties	Charging time, price, and maintenance cost demand
Ahmed	2022	Vehicle Properties	Vehicle type and electrification preference
Kurani	2022	Vehicle Properties	Safety, reliability, driving range, and duration of EVs
Contestabile et al.	2017	External Factors	Incentives for reducing burden of price for BEV
Byun	2018	External Factors	ZEV charging facility demand
Choi & Koo	2019	External Factor	Policy persistence and appearance affecting ZEV sales
Greene et al.	2020	External Factors	Hydrogen refueling infrastructure as a challenge for FCEV
Ewing & Sarigöllü	1998	Customer's Socioeconomic Characteristics	Driving patterns influencing EV preferences
Mills	2008	Customer's Socioeconomic Characteristics	Environmentally-active consumers and ZEV diffusion
Knez et al.	2021	Customer's Socioeconomic Characteristics	Influence of gender and interest in green-products
Jaiswal et al.	2022	Customer's Socioeconomic Characteristics	Indian EV potential customer segmentations

Choi	2019	Vehicle Status Quo	Relation between current and preferred vehicle fuel-type

Table 1 summarizes studies that revealed PREs of vehicle fuel-types. The tendency revealed are described in the following subsections: Section 2.1.1 vehicle related features and 2.1.2 customer related features.

2.1.1 Vehicle Related Features

The expectations and concerns of consumers are raised when new products, including vehicles with unfamiliar fuel-types, are introduced to the market (Nastjuk et al., 2020). In addition, the fanciness of advanced technology interests people, whereas the lack of verification demanding a quantity of time leads to hesitation as well.

Essential PREs are the properties of a vehicle. Apart from hesitation, concerns about discomfort and irrationally high costs (Al-Alawi & Bradley, 2013) impede people from purchasing vehicles with new fuel-types. In the case of electric vehicles (EVs), discomfort is mainly related to the charging time, driving range, and maintenance cost (Ahmed & Roorda, 2022; Byun et al., 2018; Junquera et al., 2016). Moreover, because vehicles are products directly related to the life of drivers, safety and reliability are also strongly considered (Kurani, 2022).

The discontinuity of adequate policies (Choi & Koo, 2019) also defers people from purchasing ZEVs. The most worrisome aspect of a new fuel-type vehicle is the lack of sufficient refueling or recharging in the case of plug-in infrastructure. Because vehicles are essential to most owners' daily lives, the discomfort of preparation is not easily forgiven (Junquera et al., 2016).

Because of their importance on ZEVs, there has been considerable research on vehicle battery recharging and hydrogen refueling infrastructure (Al-Alawi & Bradley, 2013; Greene et al., 2020). Most of these studies have focused on the importance of the rapid introduction of recharging and refueling infrastructure in adopting new fuel-type vehicles. Others have focused on realistic technological or social obstacles that policymakers or players in the industry were facing. These papers provide practical applicable steps to overcome obstacles and fulfill the necessity of meeting the Paris Agreement (Gota et al., 2018).

2.1.2 Customer Related Features

Similar to other products, every consumer's preference for a vehicle differs. More specifically, preference for vehicle features, including the fuel-type, class (hatchback, sedan, or SUV), fuel cost, price, and refueling or recharging comfort level, are highly related to the background of potential buyers (Ewing & Sarigöllü, 1998; Jaiswal et al., 2022; Knez et al., 2021; Mills, 2008). The most commonly considered factors are age, sex, income, education level, occupation, average driving mileage per year, residential area, and family member status (Ahmed & Roorda, 2022; Byun et al., 2018; Choi & Koo, 2019).

Although these personal factors include most of the significant segmentation criteria of general issues, vehicle status quo data are excluded, even if they are highly important (Choi & Koo, 2019). The status quo data are crucial because they provide previous vehicle purchase decisions. They include the influence of all personal factors. One reason that hinders researchers from considering the vehicle status quo is the difficulty in adequately gathering vehicle status quo data.

Despite these studies, those focusing on the influence of fundamental elements on each other (game theory) are relatively scarce. The elements include the comfort level of the BEV recharging system and the hydrogen refueling system for FCEVs compared to internal combustion engine vehicles (ICEVs), the price and vehicle class of the ones purchasable in the market, and the policies. With Shapley additive explanations (SHAP), which will be further explained in Section 2.2, we could analyze the effect of elements on each other (Hagenauer & Helbich, 2017; Lundberg & Lee, 2017).

2.2 Feature Importance Analysis for Choice Models

This study focuses on presenting a new method to analyze the feature importance from conjoint survey data. In this section, explanation about why conjoint survey is in need (Section 2.2.1) and the conventional way to analyze it (Section 2.2.2) will be given. In the following sections, the new method will be described step by step from the Random Forest classifier (Section 2.2.3), calculation of Shapley Value (Section 2.2.4), and Partial Dependence Plot (Section 2.2.5). In each section, the additional advantages these methods have compared to the conventional way is described.

2.2.1 Conjoint Survey

Discrete choice model is one of the most widely used models to predict diverse behaviors (McFadden, 1974; Train, 2009), such as a household's decision to purchase a car (Bhat & Guo, 2007; Paredes et al., 2017), choosing a freight mode (Ahmed & Roorda, 2022), and selecting the travel mode (Hagenauer & Helbich, 2017). Discrete choice models depict how human behavior or preference lead to particular decisions and are also used to project the demand of an upcoming market to solidify policy explorations (Paredes et al., 2017). Researchers select discrete choice models because they guarantee the interpretability of how a variable influences the decision-making process (Paredes et al., 2017).

Conjoint survey is one of the widely used methods to build discrete choice model (Green & Rao, 1971). As a representative method of analyzing consumer utility, by evaluating the product itself, it is possible to predict the product to be selected by estimating the consumer's utility for each attribute of the product (Green & Srinivasan, 1978). In cases where previous historical data is not available, such as new technology acceptance, the survey data is required to predict the market. If the target item is constituted with diverse features, revealing the trade-off between the features in terms of utility is the key. Therefore, conjoint survey is required for this study, which is focusing on the new vehicle fuel-type preference.

This study applied the part-worth function model, assuming that each attribute level has a separate partial value with full-profile method, presenting respondents using all of the properties they want to use in conjoint analysis. For the estimation of the coefficients, dummy variable regression, converting data into either zero or one is used (Train, 2009).

2.2.2 Multinomial Logit Regression

Discrete choice analysis includes two interrelated steps: specification of the behavioral model, which subsequently moves on to the estimation of the parameters of that model (Train, 2009). Multinomial logistic regression models (MNLs) are most frequently used in discrete choice modeling, including classification problems (Hagenauer & Helbich, 2017). Logistic regression is a statistical technique for developing a model that predicts which group of individual observations can be classified. It is used when the objects to be analyzed are divided into two or more groups and applied when part of the dependent and independent variables are categorical qualitative variables (Train, 2009). The coefficients of the MNL portray the amount of influence a specific variable has on the output, which can be directly understood as coefficients of the importance of variables (Ahmed & Roorda, 2022).

Although MNLs are statistically firm and their coefficients are intuitively explainable, they have a few limitations. First, MNLs assume that each pair of alternatives is independent of the presence or features of the other (McFadden, 1974). Therefore, when this assumption is invalid, which is the case for most real-world problems, the coefficients cannot provide precise results. In addition, the prediction performance of the model is outperformed by machine learning models, such as random forest models, in many cases, including transportation choice problems (Ahmed & Roorda, 2022).

2.2.3 Random Forest Classifier

Supervised machine learning (Cunningham et al., 2008) methods are robust for discovering patterns, especially nonlinear ones (Choudhury et al., 2021). Owing to the well-known tradeoff between the explainability and performance of data-driven models, nonlinear models such as ensemble models loosen model rigidity and increase prediction performance (Kleinberg et al., 2015). This study used one of the most widely used random forest models (Breiman, 2001) because it is advantageous for datasets on a relatively small scale (Hagenauer & Helbich, 2017; Strobl et al., 2008). Random forest builds its model by producing different decision trees on samples and obtaining their majority vote for a classification case or average in the case of regression (Breiman, 2001; Polikar, 2012).

For random forests or most other tree-based machine learning models, variable importance is calculated as the decrease in performance (Choudhury et al., 2021). It is measured as the decrease in the node impurity weighted by the probability of reaching the node. However, the feature importance obtained in this way cannot show how each feature affects the probability of being classified as a specific class of multiclass classification. Therefore, the feature importance from the RFC package focuses only on the performance drop of the model as a whole.

2.2.4 Shapley Value

SHAP by Lundberg and Lee (2017) is a widely used method to explain the individual predictions of regression or classification problems, both binary and multiclass. SHAP is based on computing the Shapley values from coalitional game theory (Molnar et al., 2018). The Shapley value is the average marginal contribution of a feature to the output (Lundberg et al., 2020). This is calculated as follows: Eq. (1)

$$\Phi_i\ M = \Sigma(v(S \cup \{i\}) - v(S))(|S|!(M-|S|-1)!)/(M-1)!) \cdots\cdots \text{Eq. (1)}$$

Φ_i is the Shapley value of the ith feature, M is the number of all features, |S| is the number of elements in the set S, and v(S) is the value expected with the features in S contributing to the prediction. In the case of multiclass classification problems, the value is the score of each class (Lundberg & Lee, 2017).

Unlike linear regression model coefficients, Shapley values can be computed by including a group of correlated features (Lundberg et al., 2018) because the concept of SHAP does not require the features to be independent. Hence, the effect of each feature is not in a form of taking portions of the output. The effect of every possible interaction sets are considered. In the case of features with high correlation, their Shapley values will be similar instead of the total summation being constant. In addition, the rankings of feature importance for each class of multiclass classification can be compared because they are calculated separately.

2.2.5 Partial Dependence Plot

A partial dependence plot of the Shapley value of each feature for each class is also available, providing a better understanding of how features affect the output (Choudhury et al., 2021). The SHAP partial dependence plots depict each data point on a plot with the x-axis showing the feature value and the y-axis showing the Shapley value of that feature of the data point (Lundberg et al., 2020). In multiclass classification problems, the value is the score of each class (Lundberg & Lee, 2017).

The application of machine learning and explainable artificial intelligence (XAI) tools along with DCM is increasing (Choudhury et al., 2021). Ahmed and Roorda (2022) used SHAP-based variable importance to compare the PREs of each vehicle type: van, single-unit truck, trailer, and passenger car. Li et al. (2022) analyzed patients' online physician choices with Shapley values. Ji et al. (2022) applied SHAP to understand the nonlinear relationships and interaction effects between cycling distance and the surrounding environment. The demand for interpreting the models without model constraints and comparing the feature importance between the options can be solved with SHAP assisting conventional DCMs (Choudhury et al., 2021).

Chapter 3. Data and Methods

Section 3.1 discusses the data to be analyzed, the way of their collection, the questions asked, why those questions were selected, and attributes of the respondents. In the following section, Section 3.2, the model used to analyze the data is described.

3.1 Data Description

This study analyzed data from a conjoint survey conducted in December 2020 for the follow-up research of Choi and Koo (2019). Survey data were selected to analyze the PREs because the target is the adoption of a new technology, and hence, there is a need for historical market data. More specifically, a conjoint survey was conducted to reveal the tradeoff between these elements. This preference itself is apparent. However, the relative importance varies among consumers. For instance, people prefer lower costs, but customers will only choose a vehicle with confirmed safety, even if it is not remarkably cheap.

Conducted by a professional survey company, 516 South Korean car owners answered eight conjoint questions, totaling 4,128 conjoint question-and-answer pairs. The respondents were gathered based on their age, gender, and residential area. The ages of the respondents ranged from 20 to 59 years because South Koreans are allowed to receive their driver’s licenses as they turn 20. To gain further meaningful implications from the survey, the percentage of EV owners was controlled at 12% of the whole respondents, while the actual stock share of EVs in 2020 in South Korea was approximately 3%.

Vehicle attribute	Vehicle A	Vehicle B	Vehicle C	Vehicle D
1. Fuel type	Gasoline	Diesel	Electric	Hydrogen
2. Refuel/Recharge convenience	50%	100%	10%	10%
3. Vehicle class	Sedan	Sedan	SUV	Sedan
4. Fuel cost (KRW/km)	100	50	50	100
5. Purchase price (million KRW)	30	40	30	20
Choose the most preferred option				O

Figure 1. An example of a conjoint question set

The conjoint questions are shown in Figure 1. Five vehicle attributes of the four vehicle options were given. The respondents were informed that the elements not shown on the cards remained identical. The first attribute was the type of fuel used by the vehicle, either gasoline, diesel, or electric, as in the BEV, or hydrogen as in the FCEV. The second attribute was the level of refueling or recharging convenience. It was on a relative scale, the convenience level of gasoline vehicles in 2020 being 100%. The following attributes were the vehicle class, either sedan or SUV, the fuel cost measured in Korean won per kilometer, and the price of each vehicle. The attribute levels for the conjoint question sets are listed in Table 2. They are determined based on actual market values.

Table 2. Conjoint question attributes and levels of attributes

Attribute	Attribute levels
1. Fuel-type	1. Gasoline
	2. Diesel
	3. Electric
	4. Hydrogen
2. Refuel/Recharge convenience	1. 10%
	2. 50%
	3. 100%
3. Vehicle class	1. Sedan
	2. SUV
4. Fuel cost	1. 50 KRW/km (0.05 USD/km)
	2. 100 KRW/km (0.09 USD/km)
	3. 150 KRW/km (0.14 USD/km)
5. Purchase price	1. 20 million KRW (18,000 USD)
	2. 30 million KRW (28,000 USD)
	3. 40 million KRW (37,000 USD)
	4. 50 million KRW (46,000 USD)

Table 3. Personal questions and responding forms

Personal question	Responding form
1. Age	Natural number
2. Gender	Binary choice (Male / Female)
3. Level of final education	Multi-class choice (High school graduate or under / Undergraduate / Graduate or more)
4. Political orientation	Multi-class choice (Extreme progressive / Mild progressive / Neutral / Mild conservative / Extreme conservative)
5. Understanding of zero-emission vehicle policy	Multi-class choice (Know absolutely nothing / Not knowing well / Neutral / Knowing relatively well / Knowing perfectly well)
6. Understanding of zero-emission vehicle technology	Multi-class choice (Know absolutely nothing / Not knowing well / Neutral / Knowing relatively well / Knowing perfectly well)
7. Monthly household income	Natural number in KRW
8. Number of students in the household	Natural number
9. Number of vehicles owned	Natural number
10. Average driving	Natural number in km

mileage per year	
11. Average purchase price of currently owning vehicles	Natural number in KRW
12. Most frequently used vehicle's class	Multi-class choice (Light car, 1000cc under / Compact sedan, 1000~1600cc / Midsize sedan, 1600~2000cc / Large sedan, 2000cc or more / Small SUV / Midsize SUV / RV)
13. Most frequently used vehicle's fuel-type	Multi-class choice (Gasoline / Diesel / LPG / Electric / Hydrogen)
14. Second frequently used vehicle's class	Multi-class choice (Light car, 1000cc under / Compact sedan, 1000~1600cc / Midsize sedan, 1600~2000cc / Large sedan, 2000cc or more / Small SUV / Midsize SUV / RV)
15. Second frequently used vehicle's fuel-type	Multi-class choice (Gasoline / Diesel / LPG / Electric / Hydrogen)
16. Third frequently used vehicle's class	Multi-class choice (Light car, 1000cc under / Compact sedan, 1000~1600cc / Midsize sedan, 1600~2000cc / Large sedan, 2000cc or more / Small SUV / Midsize SUV / RV)
17. Third frequently used vehicle's fuel-type	Multi-class choice (Gasoline / Diesel / LPG / Electric / Hydrogen)

Along with the eight conjoint questions, each respondent answered 17 personal questions, including demographic and vehicle status quo questions, as shown in Table 3. For questions from 14 to 17, only those with a second (19%) or third (1%) car answered. Therefore, the number of cars owned was close to the reality, 24% and 3%, according to Korean Statistical Information Service (KOSIS). The gender ratio represented the actual market well, with 74% male and 26% female. However, the age distribution showed a gap, especially in the 50s (15% instead of 41%, the Ministry of Land, Infrastructure, and Transport of South Korea). For better implications, the ZEV owner ratio was controlled such that it was higher than the actual value.

3.2 Model Description

We trained a RFC model and then analyzed it using Shapley values. The classification used 33 features; 17 were from personal questions and 16 (four attributes of four vehicle options) were the conjoint question attributes, excluding the fuel-type. Fuel-type was the subject of classification, forming four classes. In other words, the model was built to predict the fuel-type of the most preferred vehicle among the four options based on the features, including the respondents' features and the attributes of the questions. With 516 respondents answering eight conjoint questions, the model was trained with 4,128 sets. Since the purpose of the model was not to provide classification accuracy but to be analyzed afterward, all sets were used as training sets. The RFC was used with default hyperparameters, as shown in Table 4.

Table 4. Hyperparameters of random forest classifier and the default values used in the model

Hyperparameter	Value or status used in the model
Number of estimators	100
Function to measure the quality of splits	Gini impurity
Maximum depth of the tree	Unlimited
Minimum number of samples required to split an internal node	2
Number of features to consider when looking for the best split	Square root value of number of features
Whether bootstrap samples are used	True
Randomness of the bootstrapping of the samples	True

Chapter 4. Results

In this chapter, the results of the analysis are discussed. Section 4.1 depicts the comparison of the model's prediction performance between MNL and RFC. The following Section 4.2 presents the overall importance of the elements combining every fuel-type. Finally, Section 4.3, the elements that ranked top 3 importance to at least one of the fuel-type and the relevant element will be discussed of its tendency with partial dependence plots.

4.1 Comparison with Multinomial Logit Regression

A RFC was trained to the dataset and then had been analyzed using Shapley values it using Shapley values. In this section, the prediction performance of MNL and RFC will be compared. The MNL was achieved with the software STATA. Part of the dataset had to be simplified in order to proceed the model, since a number of features were highly correlated. As both the features of alternatives (conjoint cards) and the respondents (personal information) were considered, mixed logit model was used. In discrete choice, an individual chooses the alternative that yields the highest value of utility. The utility of choosing fuel-type j for a driver i was as follows: Eq. (2)

$$U_{i,j} = \mathbf{X_{i,j}\beta_i} + \mathbf{W_{i,j}\alpha} + \mathbf{Z_i\delta_a} + \varepsilon_{i,j} \quad \text{Eq. (2)}$$

$\boldsymbol{\beta_i}$ is random coefficients that vary over respondents, $\mathbf{X_{i,j}}$ is a vector of alternative-specific variables, $\boldsymbol{\alpha}$ are fixed coefficients on $\mathbf{W_{i,j}}$, which is a vector of alternative-specific variable, $\boldsymbol{\delta_a}$ are fixed alternative-specific coefficients on $\mathbf{Z_i}$ a vector of case-specific variables. $\varepsilon_{i,j}$ is a random error term. The BEV was set as base alternative.

RFC was trained with the sklearn library of python. The hyperparameters were shown in Table 4. The task was to predict which of the four conjoint survey options each respondent chose as the most preferable one. The test-train set split ratio was 0.3 to 0.7.

Table 5. Prediction performance comparison between Multinomial Logit regression model and random forest classifier

Model	Prediction accuracy
Multinomial Logit Regression	0.45
Random Forest Classifier	0.57

Table 6. The heatmap of actual and predicted choices of Multinomial Logit Regression model

Actual	gasoline	79	41	50	51
	diesel	28	37	29	15
	electric	151	99	379	131
	hydrogen	27	20	42	58
		gasoline	diesel	electric	hydrogen
		Predicted			

Table 7. The heatmap of actual and predicted choices of each Random Forest Classifier

Actual	gasoline	179	40	44	23
	diesel	41	98	45	14
	electric	80	59	310	51
	hydrogen	44	33	67	112
		gasoline	diesel	electric	hydrogen
		Predicted			

From Table 5, we can tell that the prediction accuracy of RFC is higher than that of MNL by 27% which is 12 percent point. The heatmap of actual and predicted choices of each models are shown in Table 6 and Table 7.

4.2 Element importance depending on each vehicle fuel-type

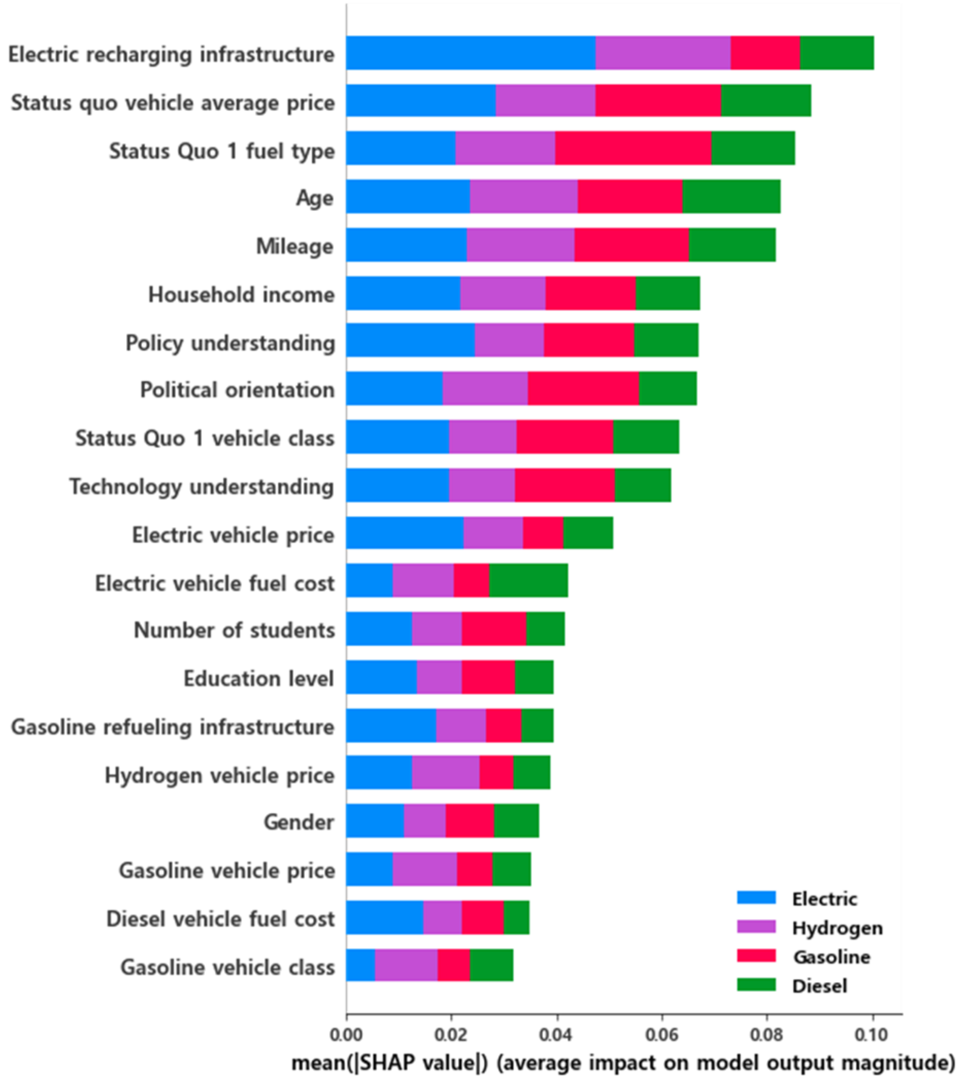

Figure 2. Average absolute Shapley value plot

From Figure 2, we can compare the importance of each element to that of the others when it comes to choosing the vehicle fuel-type out of the four options: gasoline, diesel, electric, and hydrogen. The proportion of each fuel-type in terms of element importance can be determined based on the assigned colors. Although there was a mild similarity in the overall element importance between the fuel-types, several elements showed notable differences.

Among all the elements considered, the comfort level of the EV recharging infrastructure showed the highest importance. While there was a noticeable gap between ranks one and two, ranking number two to five showed a relatively minor difference between each other, forming a "second tier group." It included the average price of currently owned vehicles, fuel-type of the most frequently used vehicle, respondent's age, and total driving mileage per year.

With another relatively large gap from the second-tier group, the third-tier group included household income, understanding of the ZEV subsidy policy, political orientation, class of the most frequently used vehicle, and level of understanding of ZEV technology. Following the EV price, the other elements are clustered into a relatively less influential group.

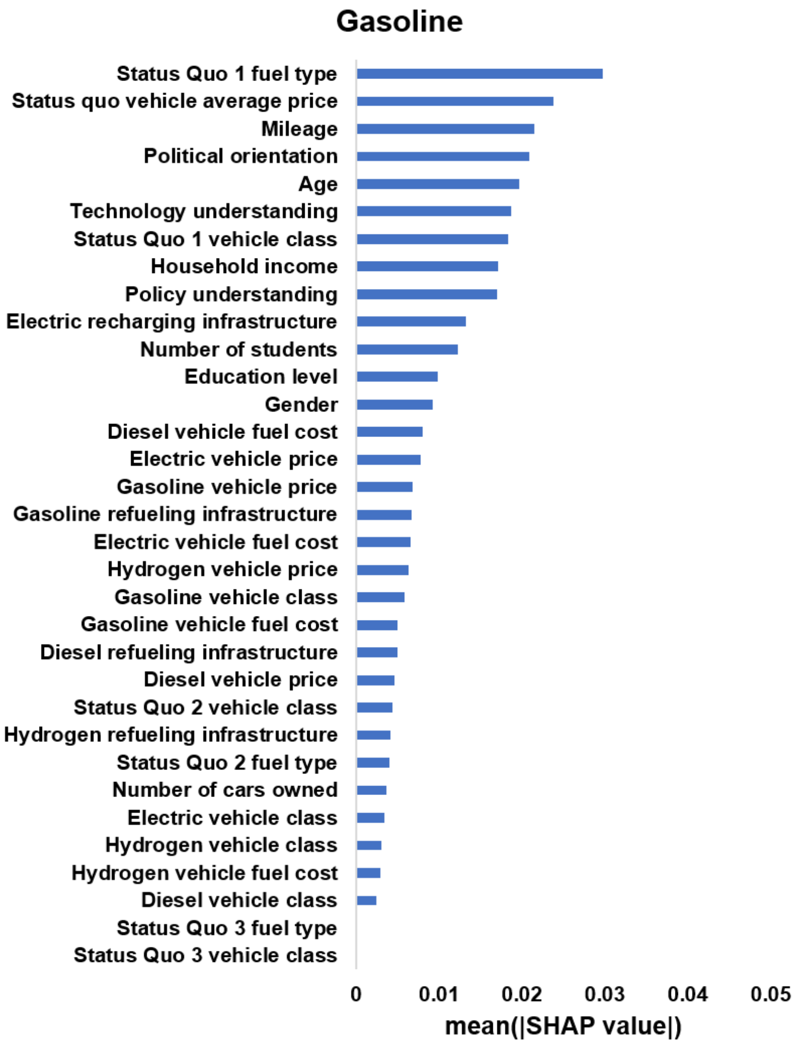

Figure 3. Element importance for preference for gasoline

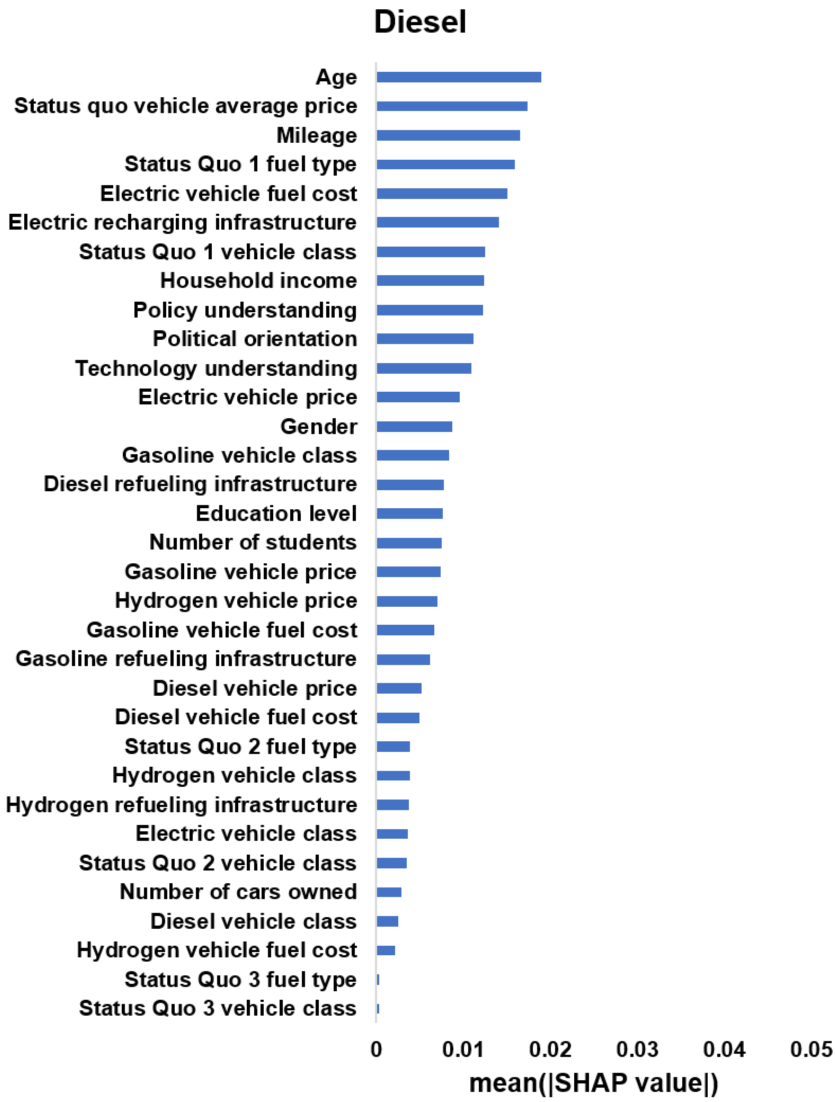

Figure 4. Element importance for preference for diesel

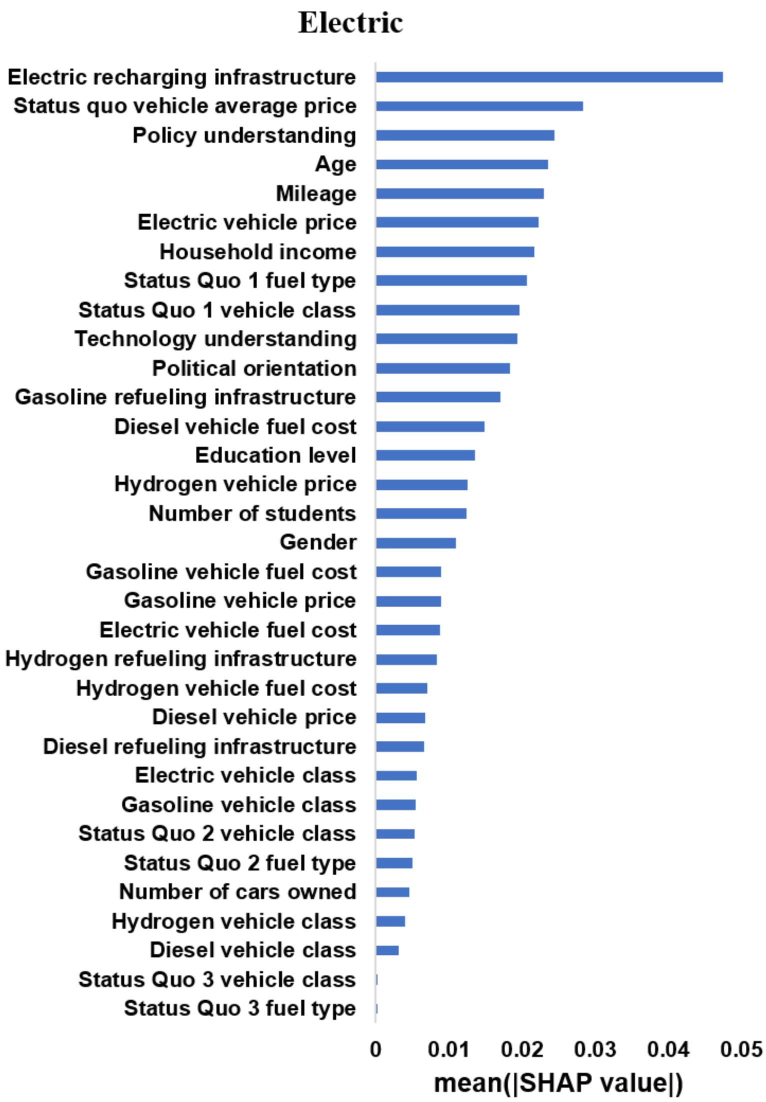

Figure 5. Element importance for preference for electric

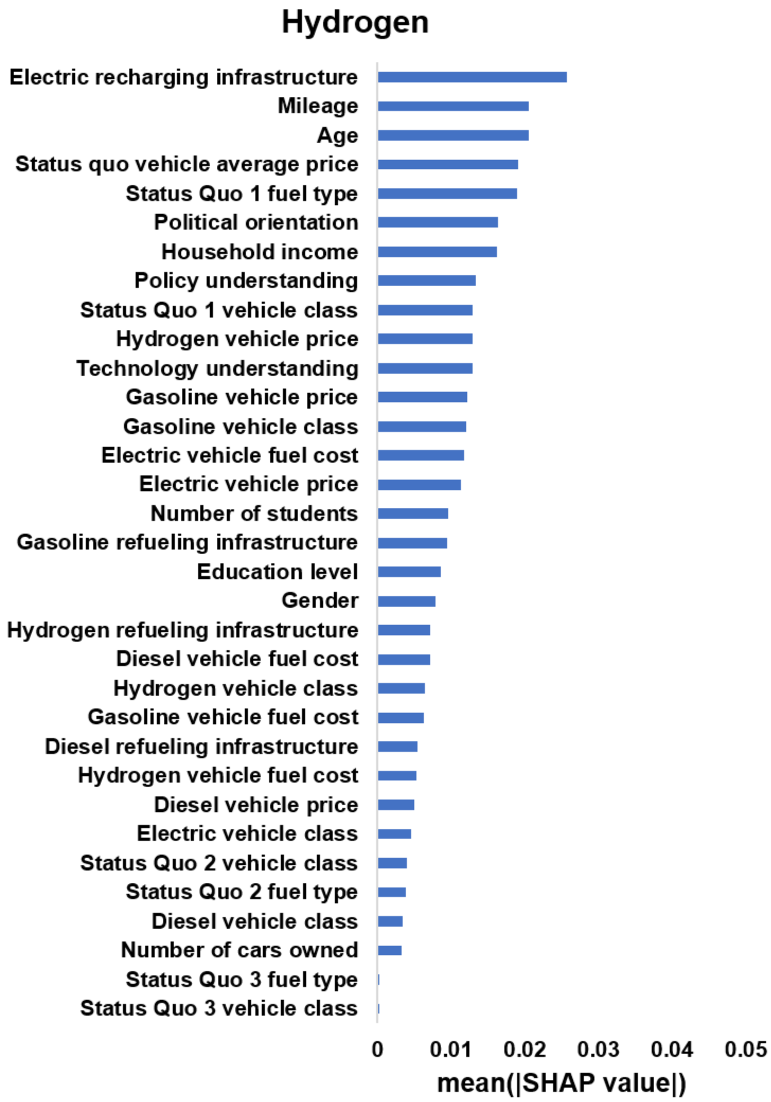

Figure 6. Element importance for preference for hydrogen

The most noticeable point in Figure 3, Figure 4, Figure 5, and Figure 6 is that the rank of element importance varies depending on the fuel-type. The preference for gasoline showed the highest relevance to the fuel-type of the most frequently used vehicle, with no other fuel-type. In addition, the comfort level of the recharging infrastructure was far less considered for gasoline than other fuel-types. This leads to the unique insight that even when the electric vehicle recharging infrastructure is well supplied, the preference for gasoline would still be present. However, the relative preference compared to electric vehicles may change.

In the case of diesel, the mean Shapley value of the elements was smaller than those of the other fuel-types. This indicates that the elements considered in the model are less related to the preference for diesel than the other elements. The gaps between the mean absolute Shapley values are also less distinguishable. Age showed the highest preference relative to diesel. Based on the partial dependence plot of age on diesel preference, the elderly showed lower preference for diesel. Notably, for both gasoline and diesel internal combustion engine vehicles, information about the vehicle itself are less relatable than the personal status.

On the other hand, preference towards electric vehicles showed a significant relationship with the level of recharging infrastructure given. The mean absolute Shapley value of recharging infrastructure level was approximately 70% higher than that of the second-ranked element, the average price of currently owned vehicles. In addition, household income ranked higher with the two ZEVs preferences than ICEVs.

One of the most intriguing points in Figure 6 is that the recharging infrastructure level of BEVs showed a higher relatedness than hydrogen refueling infrastructure level of FCEVs. This could imply that people consider FCEVs to be substitutions for BEVs. Another difference between Figure 5 and Figure 6 is that understanding the technology of

ZEVs showed higher relatedness towards the preference of FCEVs than that of the policy. However, BEVs showed an opposite trend.

4.3 Effect of Major Elements

In this section, some of the elements are selected and their tendency with the preference are analyzed through partial dependence plots. The selected PREs are ranked higher than the third for at least one of the fuel-type, or those that are highly related to them. Section 4.3.1 is about the recharging and refueling infrastructure of ZEV, Section 4.3.2 is about household income and the status quo vehicle average price, Section 4.3.3 discusses the status quo vehicle fuel-type, Section 4.3.4 describes about age and gender of the car owner, Section 4.3.5 explains the yearly mileage of the driver and the fuel cost of each fuel-type, and finally Section 4.3.6 covers the understanding level of ZEV subsidy policy and ZEV technology.

4.3.1 Electric Recharging and Hydrogen Refueling Infrastructure

Comparing the partial dependence plots of the BEV recharging and FCEV refueling infrastructure levels on the preference for each fuel-type also has notable implications for policy formulations, especially those related to infrastructure investment. Understanding the differences between fuel-types would lead to more precise predictions of policy implementation results.

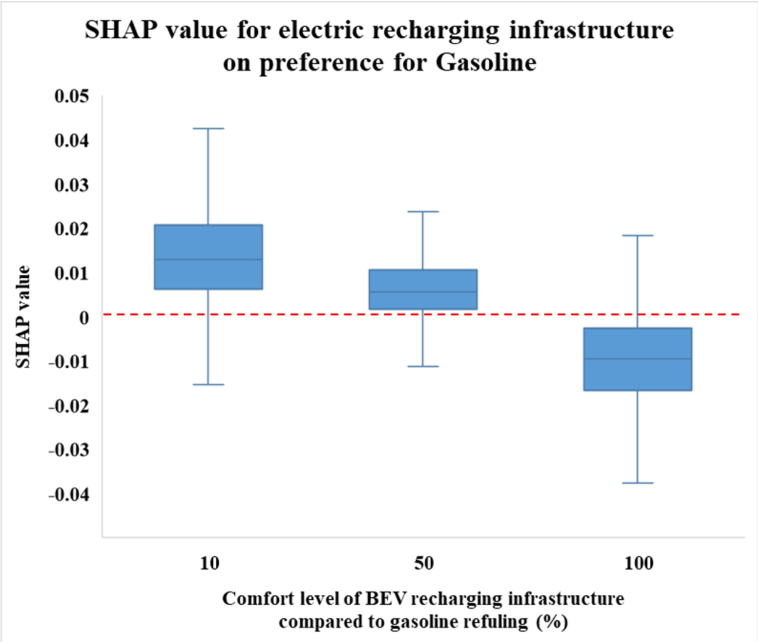

Figure 7. Partial dependence plot of recharging infrastructure level for gasoline

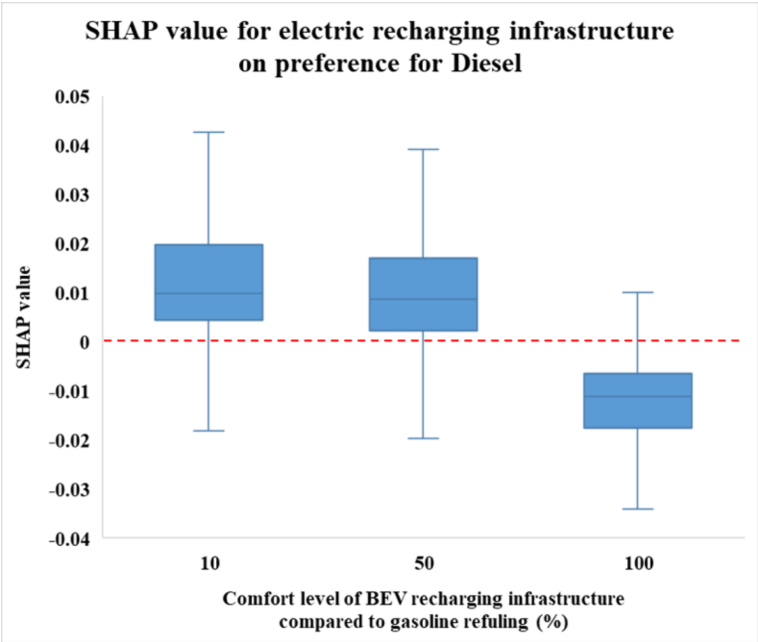

Figure 8. Partial dependence plot of recharging infrastructure level for diesel

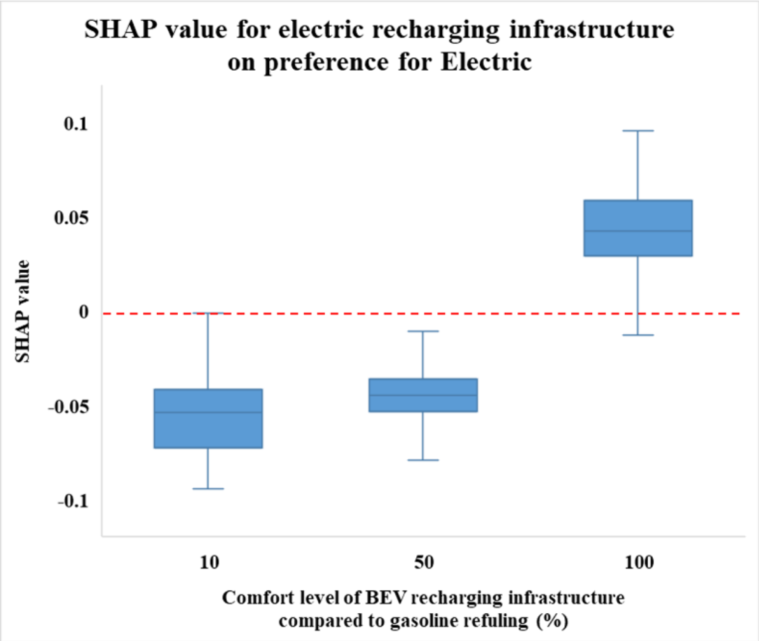

Figure 9. Partial dependence plot of recharging infrastructure level for electric

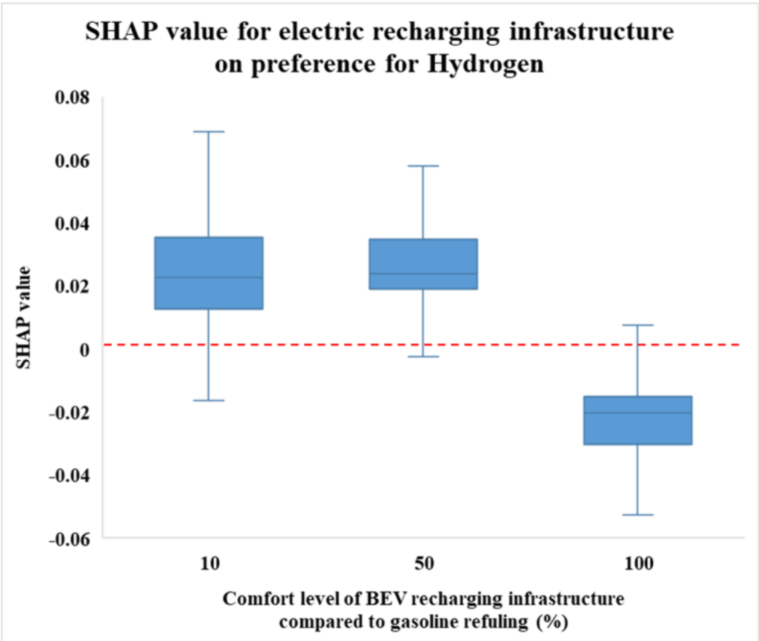

Figure 10. Partial dependence plot of recharging infrastructure level for hydrogen

Figure 7, Figure 8, Figure 9, and Figure 10 portray the impact of the recharging infrastructure level of BEVs on the preference for each fuel-type: gasoline, diesel, electric, and hydrogen. Preference for BEVs increased as the recharging infrastructure improved, while all the other three fuel-types showed the opposite tendency. The noticeable point was that this tendency was in a nonlinear form. The preference level remained similar when the recharging comfort level was 10% and 50%. As the recharging infrastructure reached a level similar to that of current gasoline vehicle refueling, the preference towards BEVs soared, and the others dropped. It is also to be noted that the preference towards FCEV was more vividly affected by the BEV recharging infrastructure than with the two ICEVs.

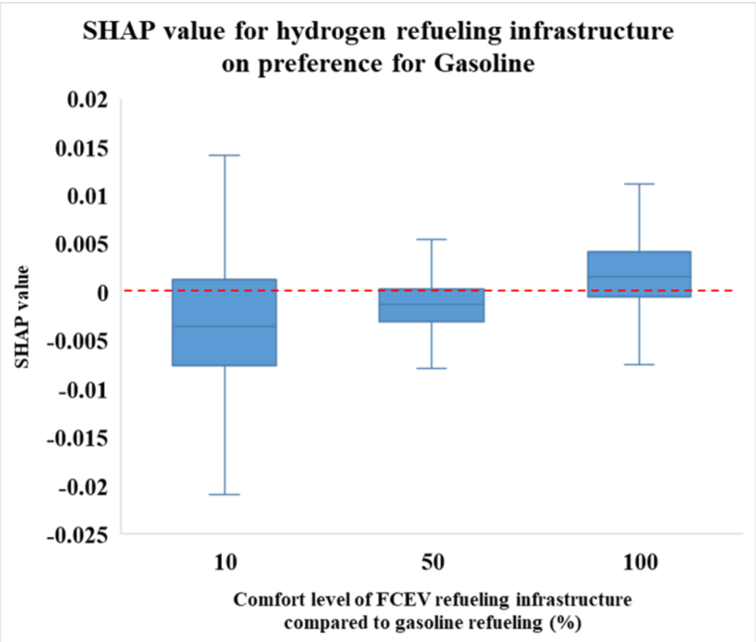

Figure 11. Partial dependence plot of hydrogen refueling infrastructure level for gasoline

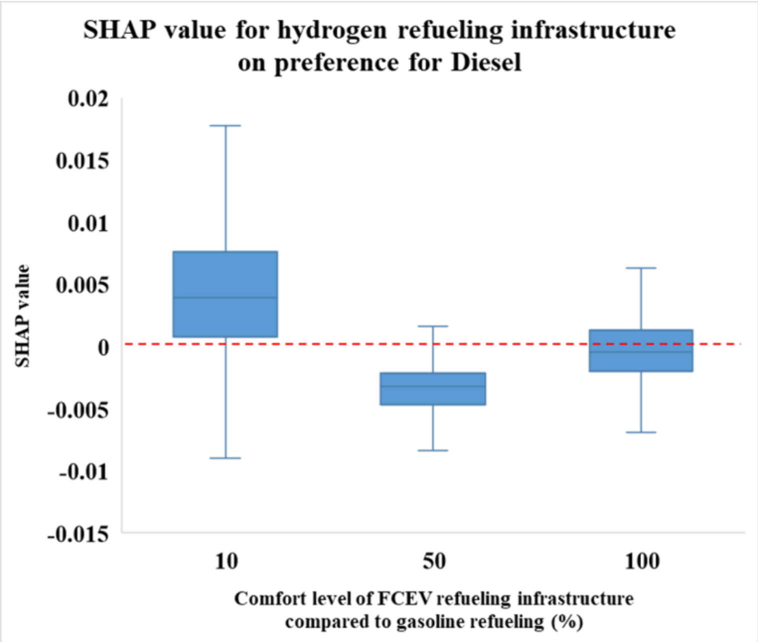

Figure 12. Partial dependence plot of hydrogen refueling infrastructure level for diesel

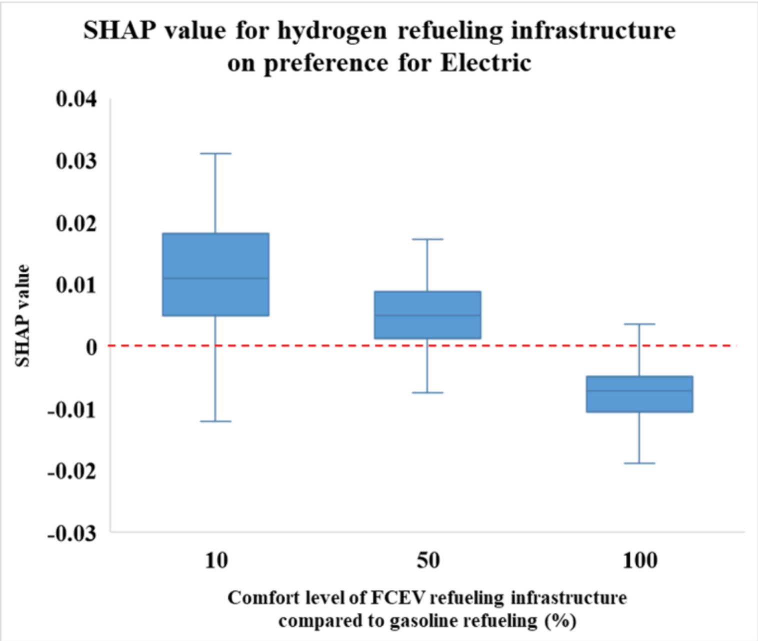

Figure 13. Partial dependence plot of hydrogen refueling infrastructure level for electric

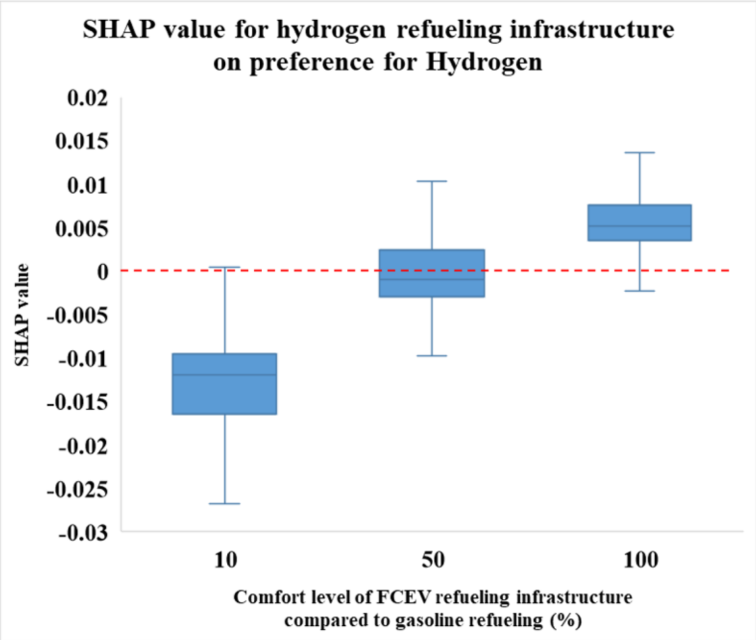

Figure 14. Partial dependence plot of hydrogen refueling infrastructure for hydrogen

Figure 11, Figure 12, Figure 13, and Figure 14 display the influence of hydrogen refueling infrastructure level on the preference for each fuel-type: gasoline, diesel, electric, and hydrogen. The preference for ICEVs remained neutral and not strongly affected by the change in the hydrogen refueling infrastructure level. However, the BEV preference decreased as hydrogen refueling achieved a better comfort level. This tendency was far more linear than vice versa. The preference for FCEVs increased in a nonlinear pattern, showing a discrete gap from 10% to 50% of the hydrogen refueling infrastructure level. On the other hand, the enhancement from 50% to 100% had a less distinguishable increase in the preference for the FCEV.

From Figure 9, Figure 10, Figure 13, and Figure 14, it can be observed that the relationship between the BEV and FCEV is substitutional. In addition, Figure 7, Figure 8,

Figure 11, and Figure 12 show that investment in recharging and hydrogen refueling infrastructure has a favorable impact on the preference for ICEVs, but the relative preference decreases. To strengthen the preference for BEVs, the recharging infrastructure must rapidly reach a level comparable to that of current gasoline refueling infrastructure. On the other hand, the preference for FCEVs will soar only if the hydrogen refueling infrastructure level reaches at least 50% of the comfort level of the current gasoline refueling case. As these findings are applied to infrastructure investment policies, they are more efficient in terms of both time and cost.

4.3.2 Status Quo Vehicle Average Price and Household Income

The average price of currently owning vehicles was considered as highly important elements, ranked as the second place for gasoline, diesel, and electric, and as the fourth place for hydrogen (Figure 3, Figure 4, Figure 5, and Figure 6).

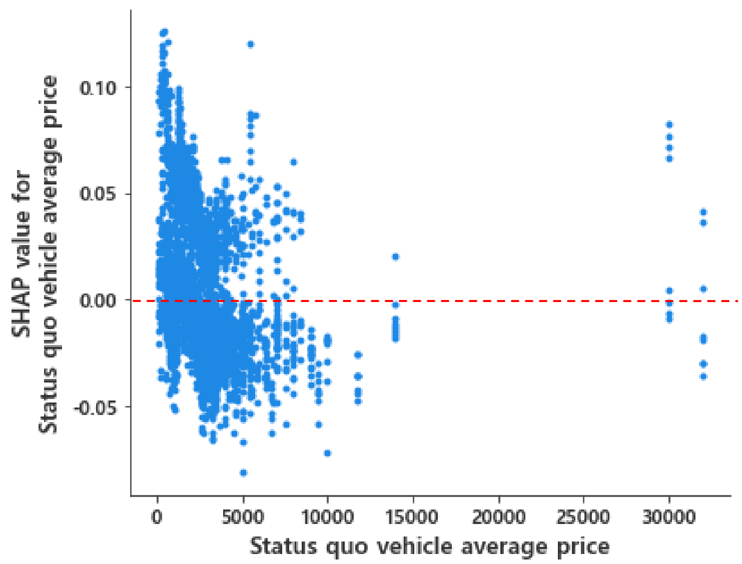

Figure 15. Partial dependence plot of shapley value of status quo vehicle average price on preference towards gasoline

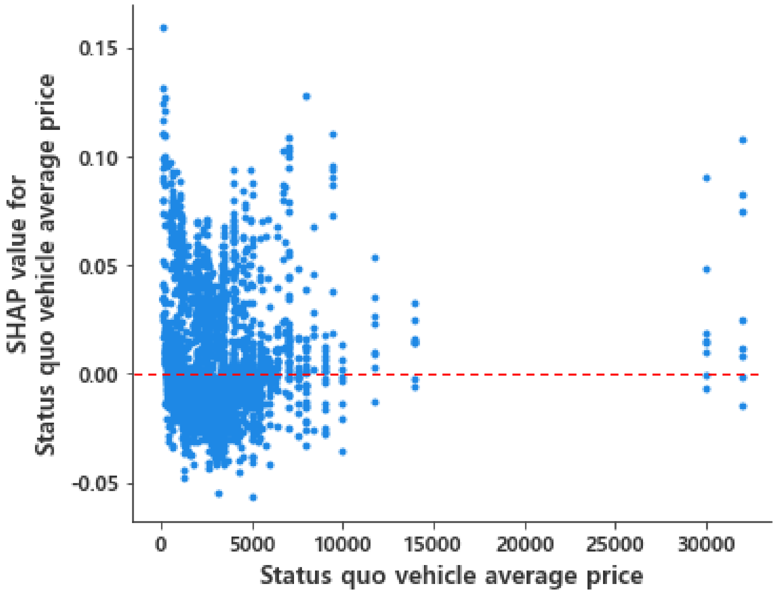

Figure 16. Partial dependence plot of shapley value of status quo vehicle average price on preference towards diesel

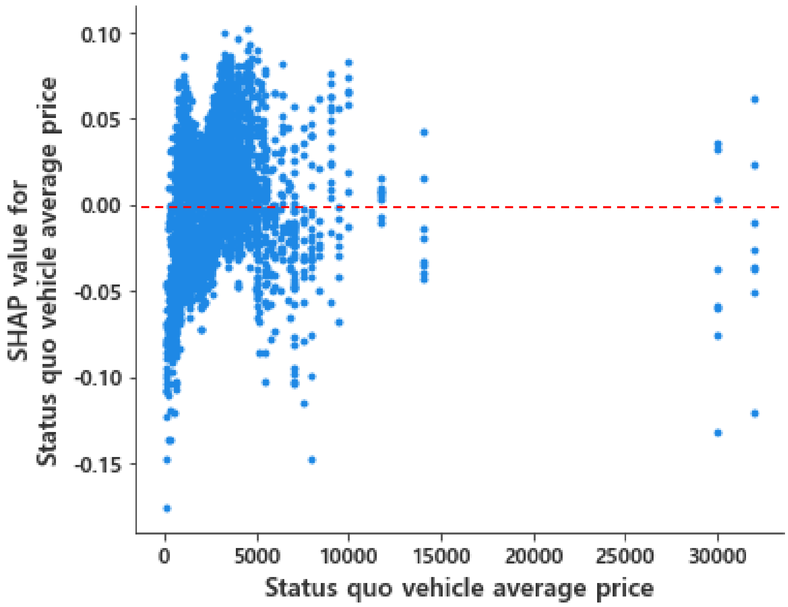

Figure 17. Partial dependence plot of shapley value of status quo vehicle average price on preference towards electric

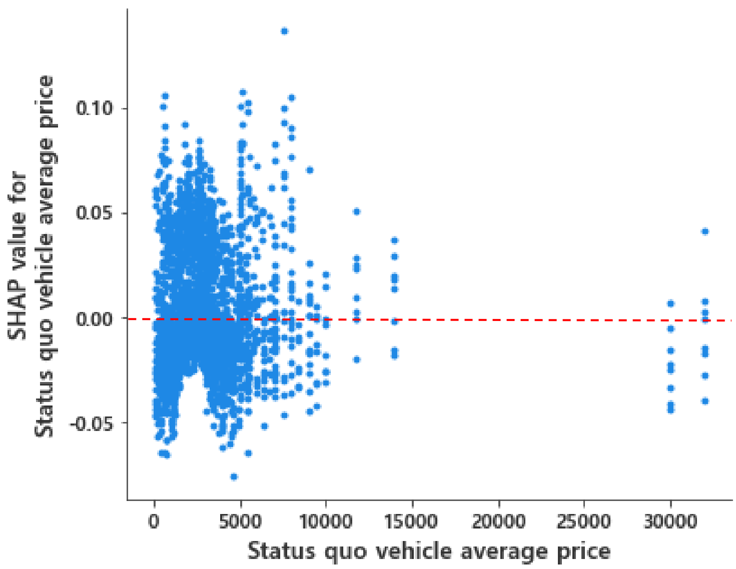

Figure 18. Partial dependence plot of shapley value of status quo vehicle average price on preference towards hydrogen

The x-axis of Figure 15, Figure 16, Figure 17, and Figure 18 are the average price of currently owning vehicles. Their units are 10,000 KRW. Figure 15 shows that the preference towards gasoline drops as the average price of currently owned vehicles increases until it reaches approximately 40 million KRW (37,000 USD). From that point, the linearity weakens. Similar tendency was shown in Figure 16, with the preference towards diesel.

However, for the ZEVs, the preference remained less affected by the average price of currently owning vehicles in the range, below 40 million KRW. Instead, from 30 million KRW to 50 million KRW, the preference towards BEV increased. It is noticeable that the opposite tendency was shown for the FCEV in the same range. For the status quo average price higher than 50 million KRW, the number of respondents were insufficient to discover a solid tendency.

Monthly household income was considered to be highly correlated with the status quo vehicle average price. Those with larger income would purchase more expensive vehicles than the others. Therefore, along with the status quo average price, the partial dependence plots of household income were also analyzed in this section.

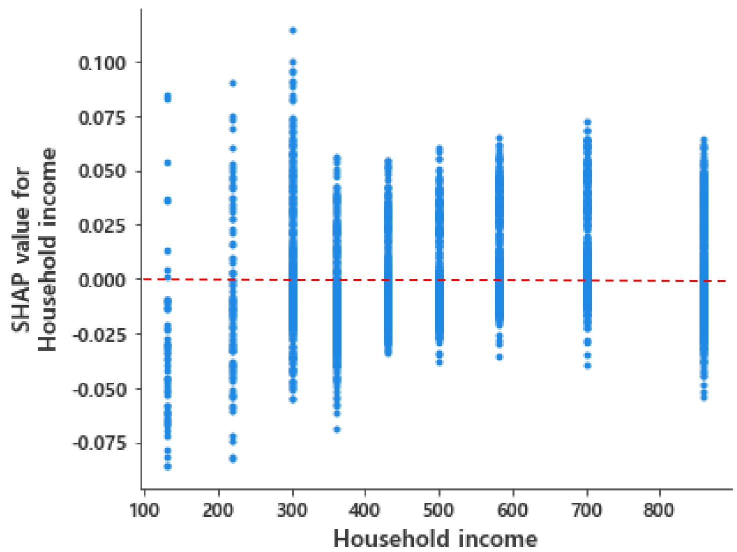

Figure 19. Partial dependence plot of shapley value of household income on preference towards gasoline

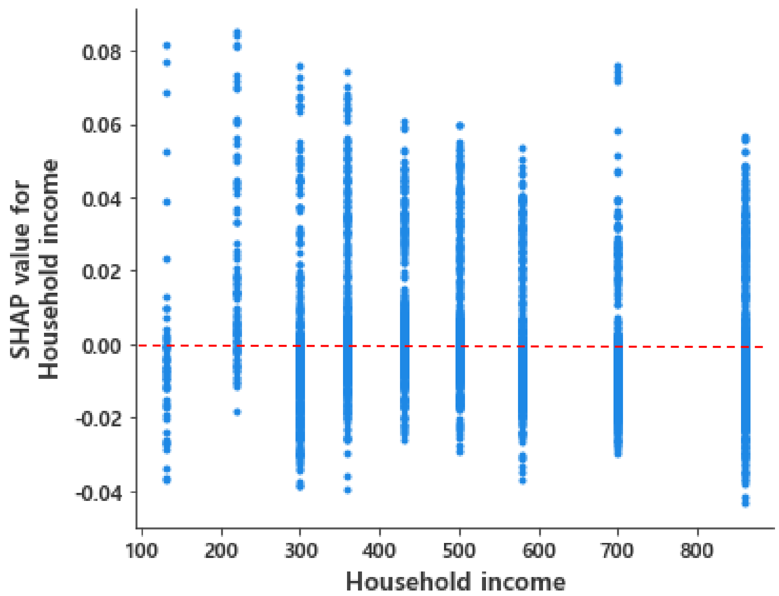

Figure 20. Partial dependence plot of shapley value of household income on preference towards diesel

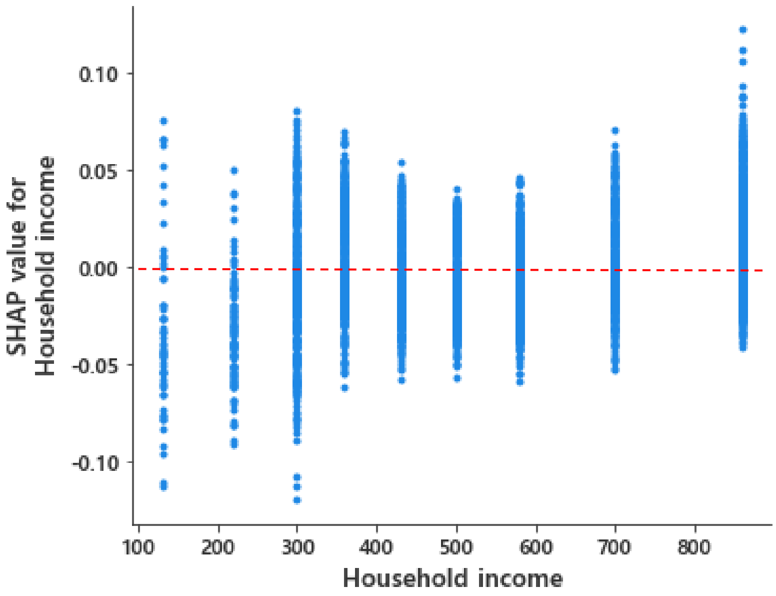

Figure 21. Partial dependence plot of shapley value of household income on preference towards electric

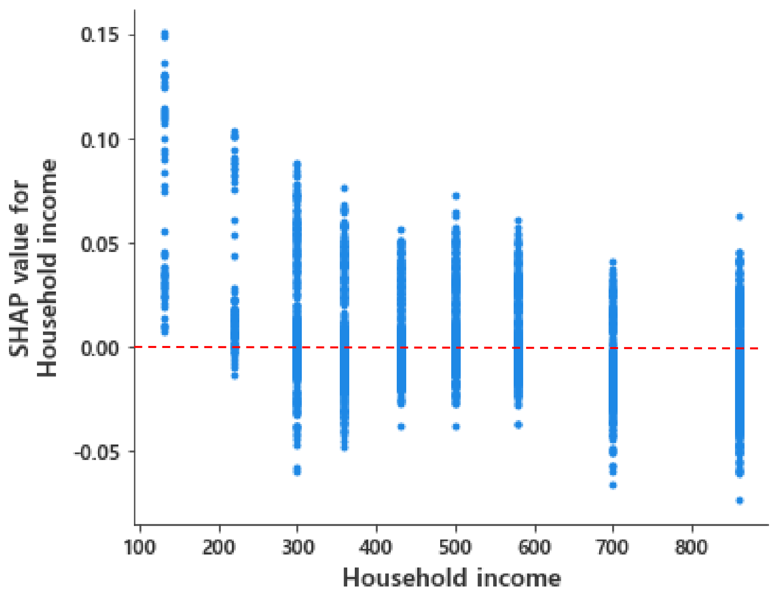

Figure 22. Partial dependence plot of shapley value of household income on preference towards hydrogen

Figure 19, Figure 20, Figure 21, and Figure 22 are the partial dependence plots of shapley value of household income on the preference towards each vehicle fuel-type, gasoline, diesel, electric, and hydrogen. Since the respondents were asked to choose the range of monthly household income from the given options, rather than responding in a continuous numerical form, the x-axis shows the discrete household income values. Their units are 10,000 KRW.

From Figure 19, Figure 20, and Figure 21, it was shown that the household income is not highly related with the preference towards each vehicle fuel-type, gasoline, diesel, and electric. However, in case of the hydrogen, as shown in Figure 22, in the range where household income is less than 3 million KRW a month, the preference towards FCEV decreased as the income level increased.

4.3.3 Status Quo Vehicle Fuel-type

Status quo, especially the one that is the most frequently used, fuel-type was considered as highly related one to the preference towards each fuel-type, ranked number one for gasoline, four for diesel, eight for electric, and five for hydrogen.

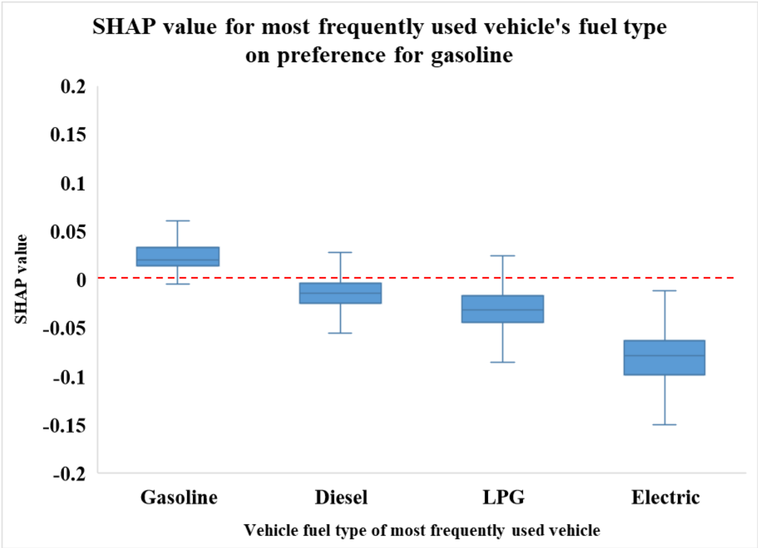

Figure 23. Partial dependence plot of shapley value of vehicle fuel-type of most frequently used vehicle towards gasoline

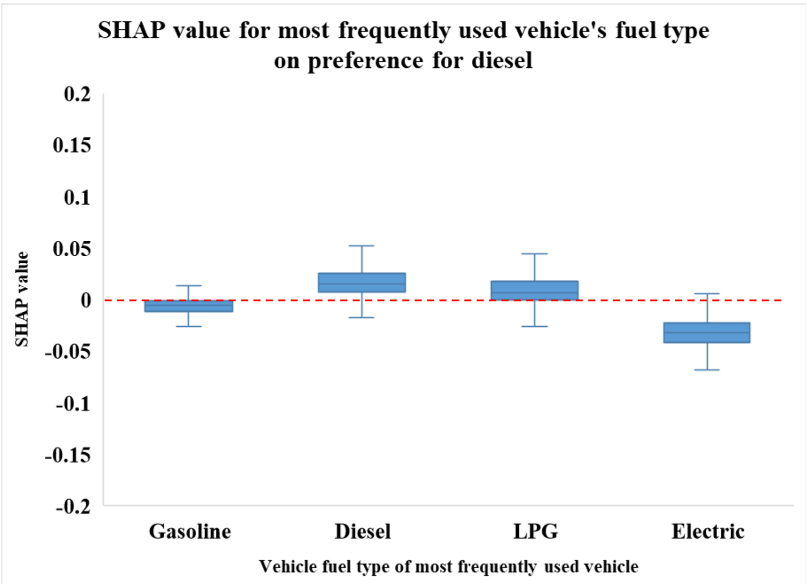

Figure 24. Partial dependence plot of shapley value of vehicle fuel-type of most frequently used vehicle towards diesel

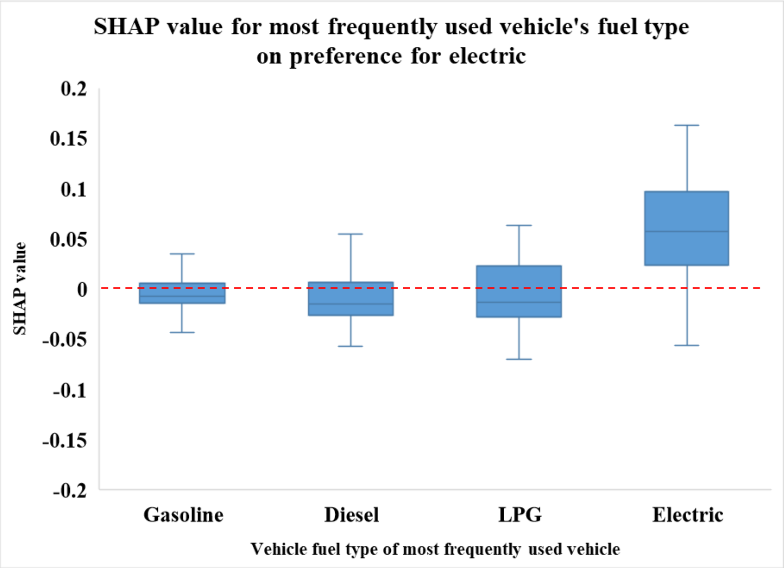

Figure 25. Partial dependence plot of shapley value of vehicle fuel-type of most frequently used vehicle towards electric

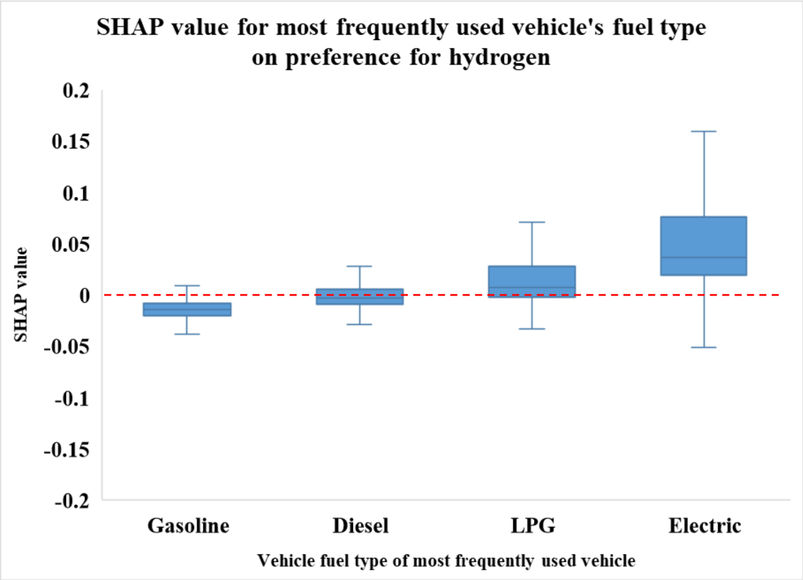

Figure 26. Partial dependence plot of shapley value of vehicle fuel-type of most frequently used vehicle towards hydrogen

Figure 23 shows the partial dependence plot of the most relevant element on the gasoline preference, the fuel-type of the most frequently used vehicle. It informs that those whose most frequently used vehicle is a gasoline vehicle show distinct preference towards gasoline. An interesting point to be noted is that BEV drivers show less preference for gasoline than diesel or LPG drivers.

4.3.4 Age and Gender

Age and gender are the most frequently used criteria to segment people. It is because not only they are the most approachable data but also implies a lot of information within it. Age was considered as one of the highly related elements for preference towards each vehicle fuel-type, ranked the fifth for gasoline, the first for diesel, the fourth for electric, and the third for hydrogen. On the other hand, gender ranked as one of the least related elements for every fuel-type's preference.

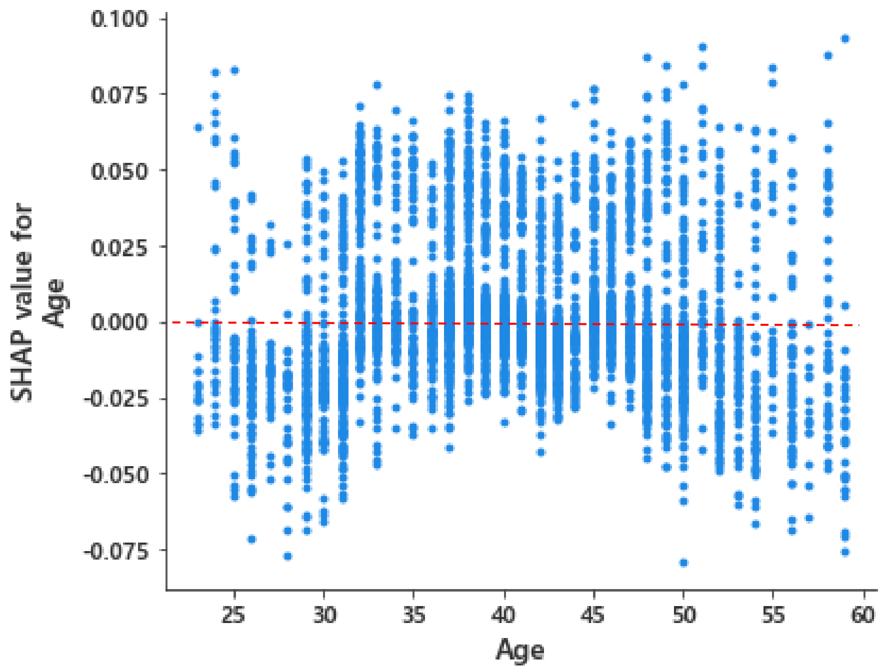

Figure 27. Partial dependence plot of shapley value of driver's age towards gasoline

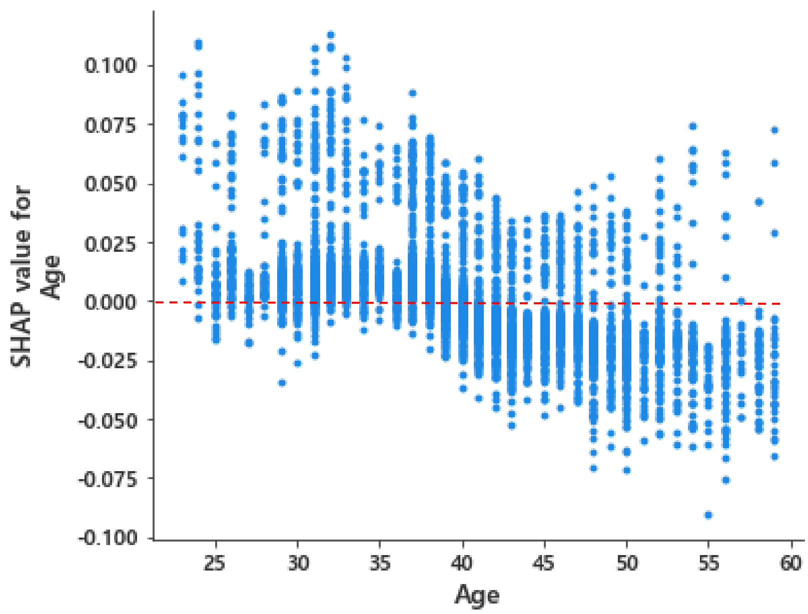

Figure 28. Partial dependence plot of shapley value of driver's age towards diesel

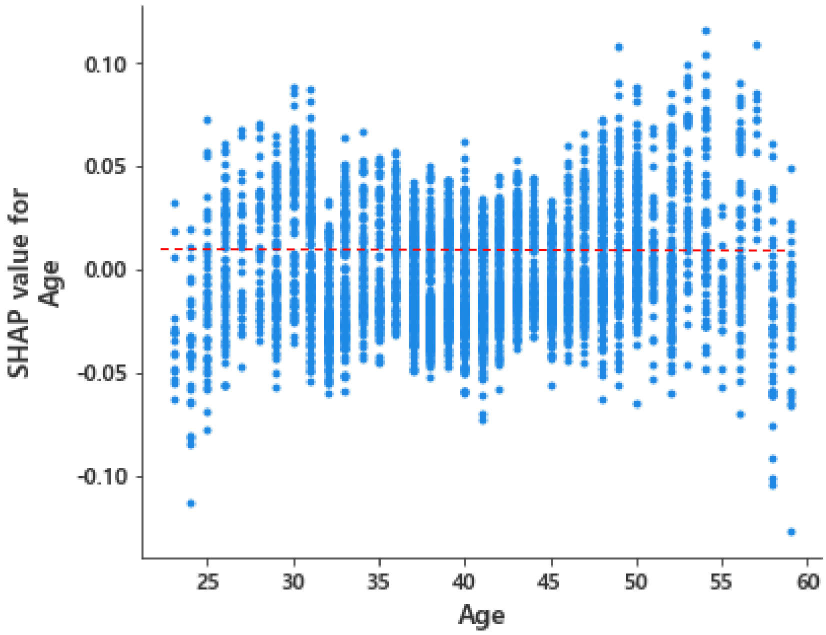

Figure 29. Partial dependence plot of shapley value of driver's age towards electric

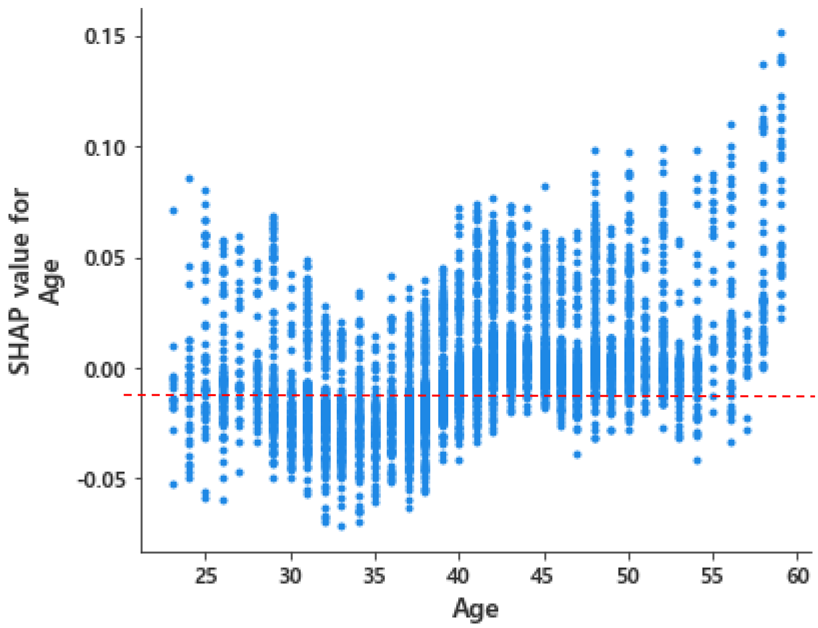

Figure 30. Partial dependence plot of shapley value of driver's age towards hydrogen

Figure 27 and Figure 29 shows that age is relatively less related with the preference towards gasoline and BEV. On the other hand, Figure 28 showed that preference towards diesel vehicle decreased as the age increased. Especially for those who were younger than 35 vividly preferred diesel, while those who are older than late 40s showed the opposite tendency. Figure 30 depicted that the preference towards hydrogen FCEV increased as the driver's age increased. Those who are younger than 35 showed relatively neutral preference towards FCEV, while those who are older than early 40s preferred it.

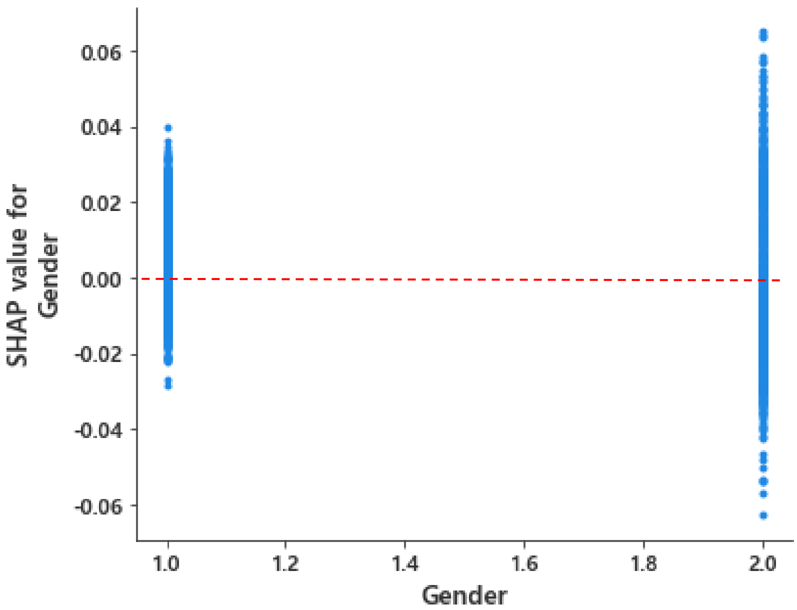

Figure 31. Partial dependence plot of shapley value of driver's gender towards gasoline

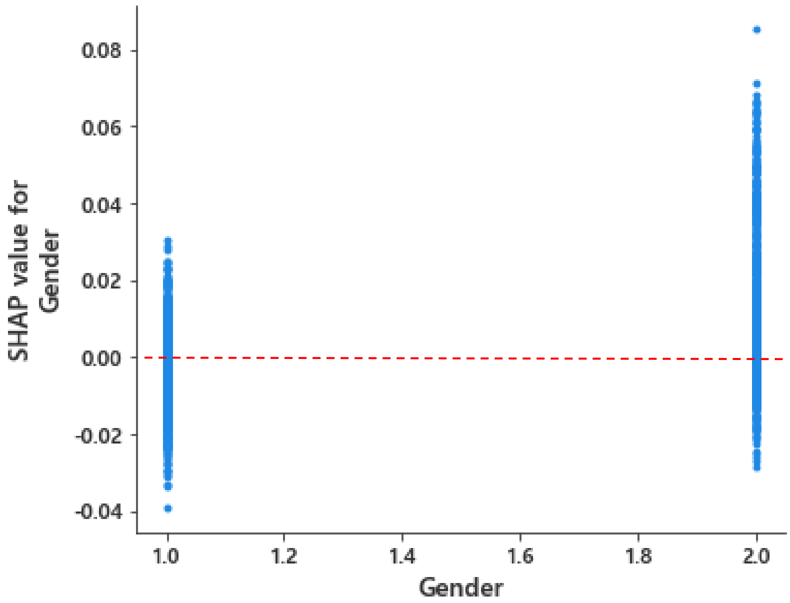

Figure 32. Partial dependence plot of shapley value of driver's gender towards diesel

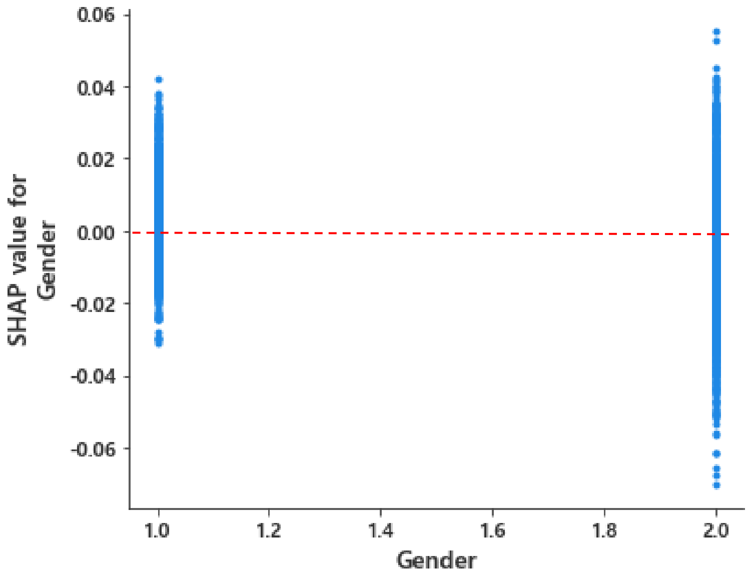

Figure 33. Partial dependence plot of shapley value of driver's gender towards battery electric vehicle

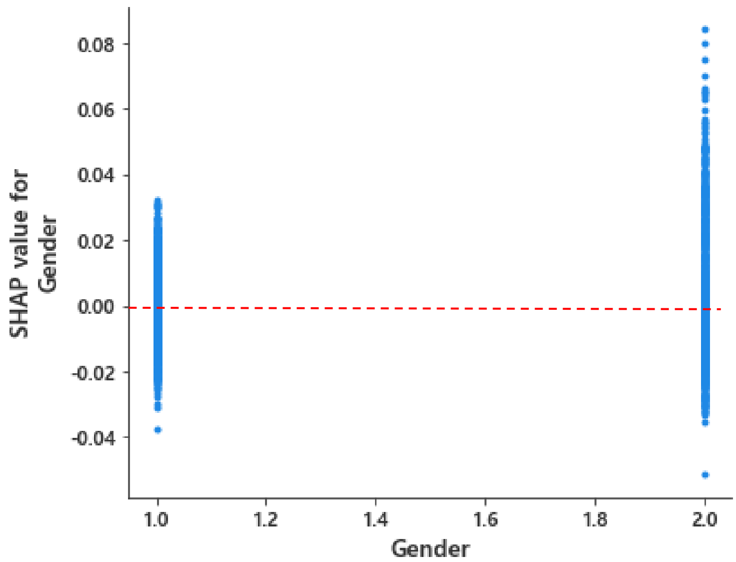

Figure 34. Partial dependence plot of shapley value of driver's gender towards hydrogen fuel-cell electric vehicle

Figure 31, Figure 32, Figure 33, and Figure 34 showed the partial dependence plots of shapley value of driver's gender on preference towards each vehicle fuel-type, gasoline, diesel, electric, and hydrogen. The value "1" stands for male and "2" stands for female. For every fuel-types, gender did not show solid relatedness with preference towards them.

4.3.5 Mileage and Fuel Cost

Yearly driving mileage of the driver showed high importance for preference towards each vehicle fuel-type, ranked number three for gasoline and diesel, five for electric, and two for hydrogen.

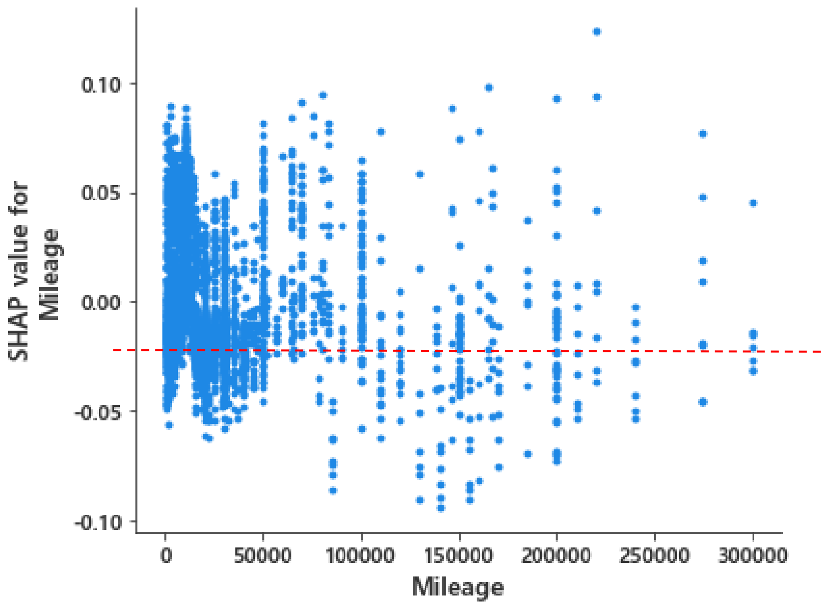

Figure 35. Partial dependence plot of shapley value of driver's yearly driving mileage on preference towards gasoline

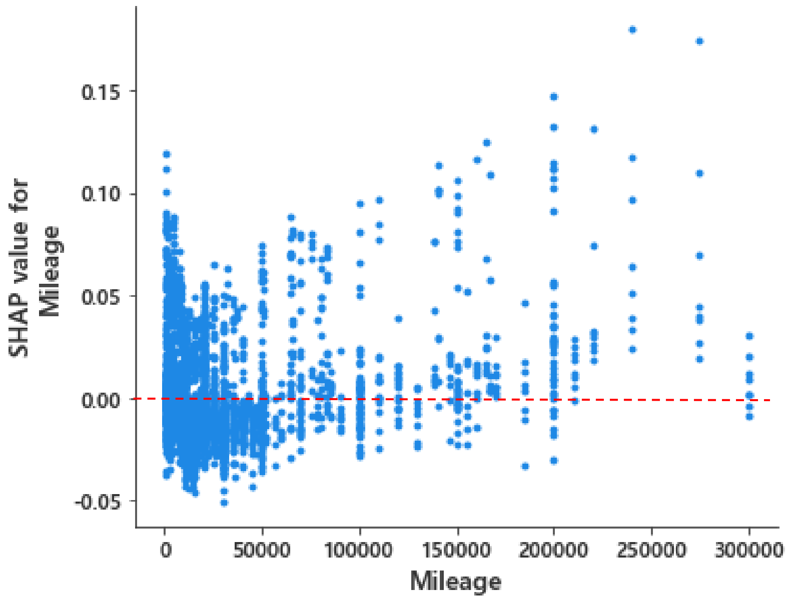

Figure 36. Partial dependence plot of shapley value of driver's yearly driving mileage on preference towards gasoline

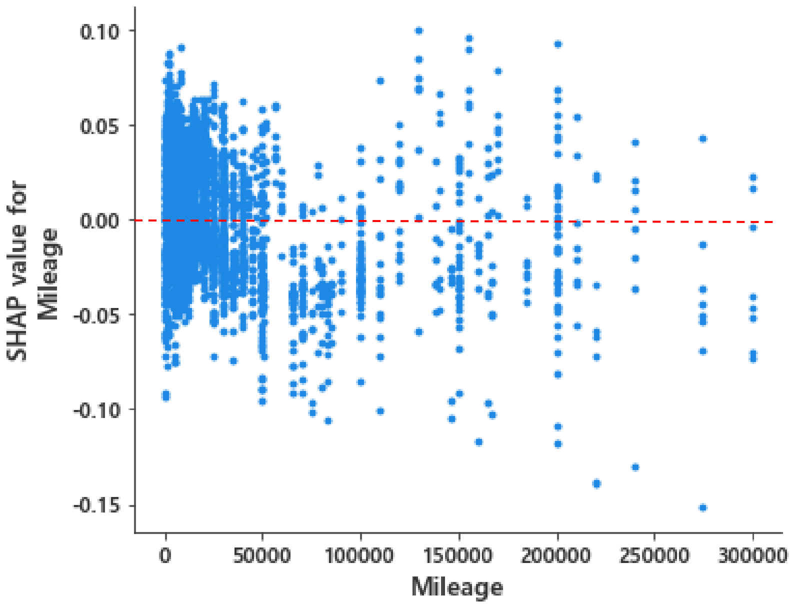

Figure 37. Partial dependence plot of shapley value of driver’s yearly driving mileage on preference towards electric

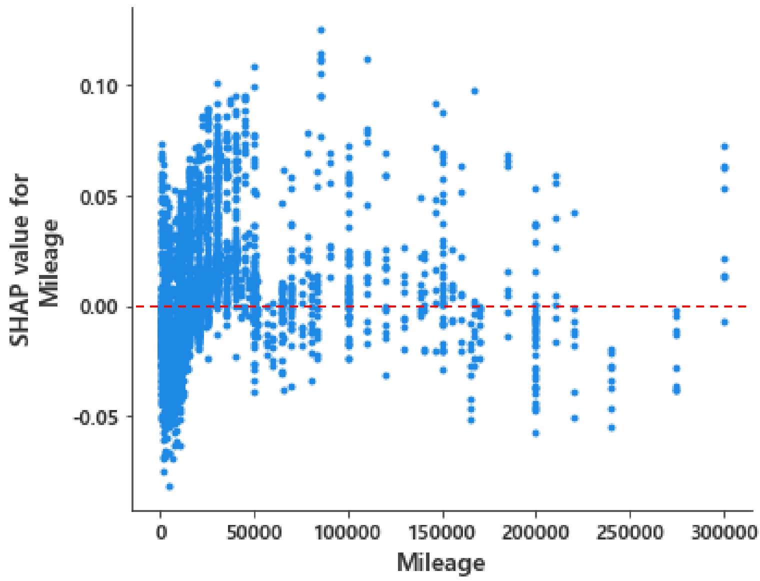

Figure 38. Partial dependence plot of shapley value of driver’s yearly driving mileage on preference towards hydrogen

Figure 35, Figure 36, Figure 37, and Figure 38 show the partial dependence plots of shapley value of driver's yearly driving mileage on preference towards each vehicle fuel-type. The mileage in x-axis is depicted in the unit of kilometers. Figure 35, Figure 36, and Figure 37 showed that mileage does not have vivid relatedness with the preference towards gasoline, diesel, and BEV. However, in case of FCEV, as the mileage increased, the preference towards FCEV increased as well.

Since mileage is highly correlated with the fuel cost, the fuel cost of each vehicle fuel type's shapley values are also analyzed in partial dependence plots.

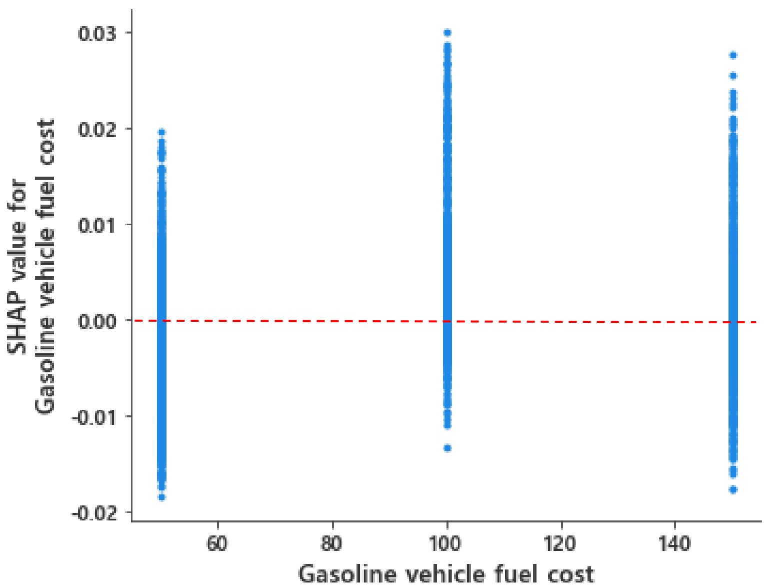

Figure 39. Partial dependence plot of shapley value of fuel cost of gasoline vehicle fuel cost on preference towards gasoline

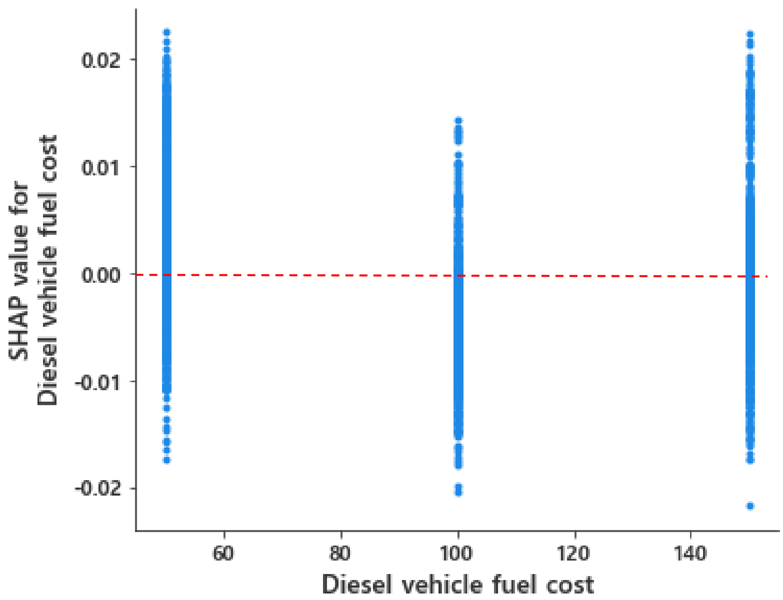

Figure 40. Partial dependence plot of shapley value of fuel cost of diesel vehicle fuel cost on preference towards diesel

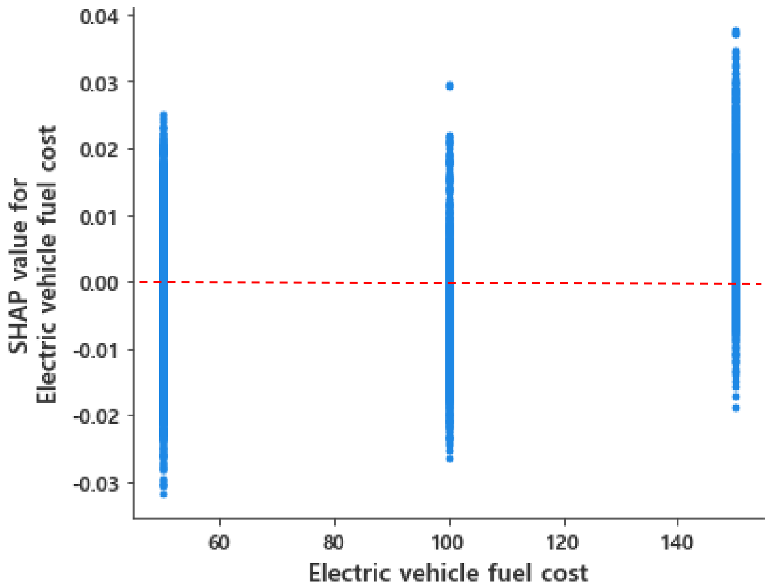

Figure 41. Partial dependence plot of shapley value of recharging cost of battery electric vehicle fuel cost on preference towards battery electric vehicle

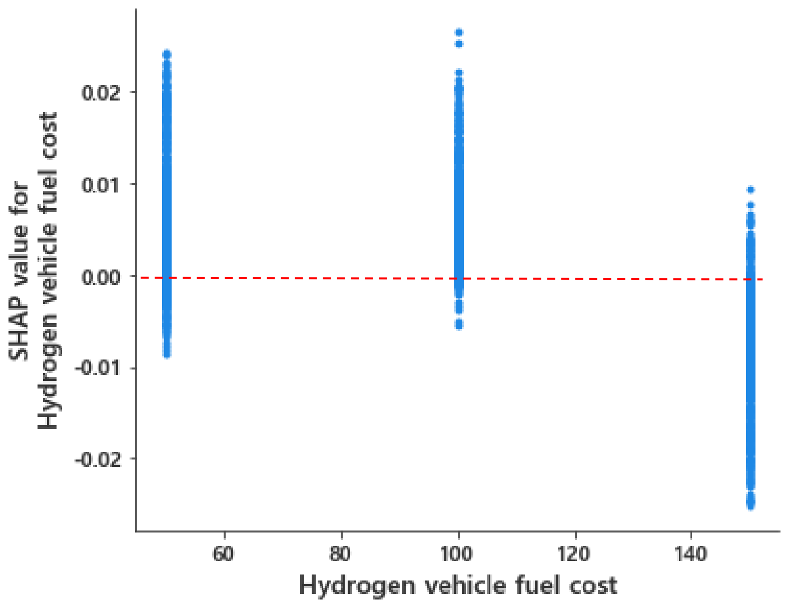

Figure 42. Partial dependence plot of shapley value of fuel cost of hydrogen fuel-cell electric vehicle fuel cost on preference towards hydrogen

The x-axis of Figure 39, Figure 41, Figure 40, and Figure 42 display the discrete fuel cost given in the conjoint survey option cards its unit is KRW/km. From Figure 39 and Figure 41, it could be implied that fuel cost of ICEV does not have strong relatedness with the preference to the according fuel-type. On the other hand, from Figure 40 and Figure 42, as the fuel cost increases for ZEV, the preference towards it decreased. One interesting point was that the decreasing tendency was not linear. As the fuel cost increased from 100 KRW/km to 140 KRW/km, the decrease gap was larger than that from 60 KRW/km to 100 KRW/km.

4.3.6 Policy Understanding and Technology Understanding

The understanding of ZEV subsidy policy and the technological mechanism of ZEV are not considered as one of the critical elements for the preference towards each vehicle fuel-type. The highest ranking was for the preference towards BEV, understanding of ZEV as ranking in the third place.

However, these are analyzed because for two reasons. First, they showed correlation with the level of education of the drivers. Second, both of these elements are able to and also most effectively enhanced through advertisements or campaigns. If they are proved to be fundamental for ZEV induction, governments and ZEV manufacturers must invest these as part of the marketing strategies.

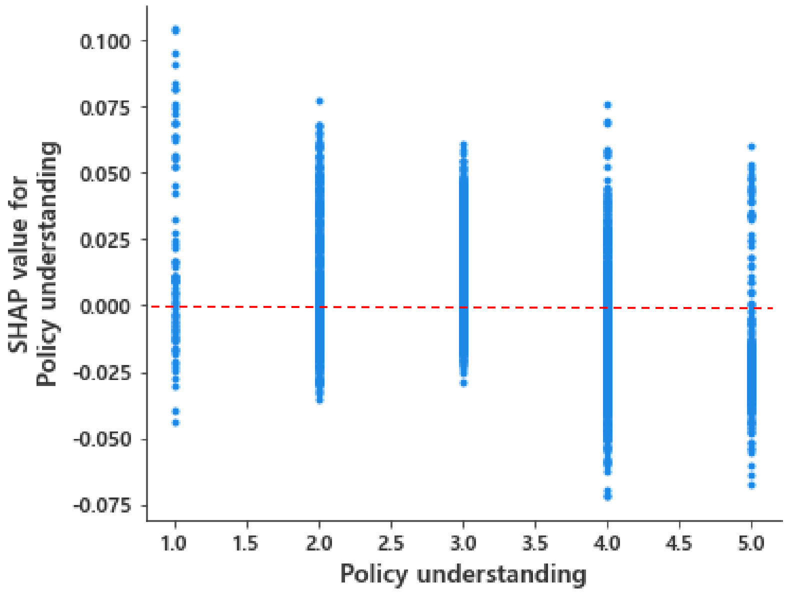

Figure 43. Partial dependence plot of shapley value of level of ZEV subsidy policy understanding on preference towards gasoline

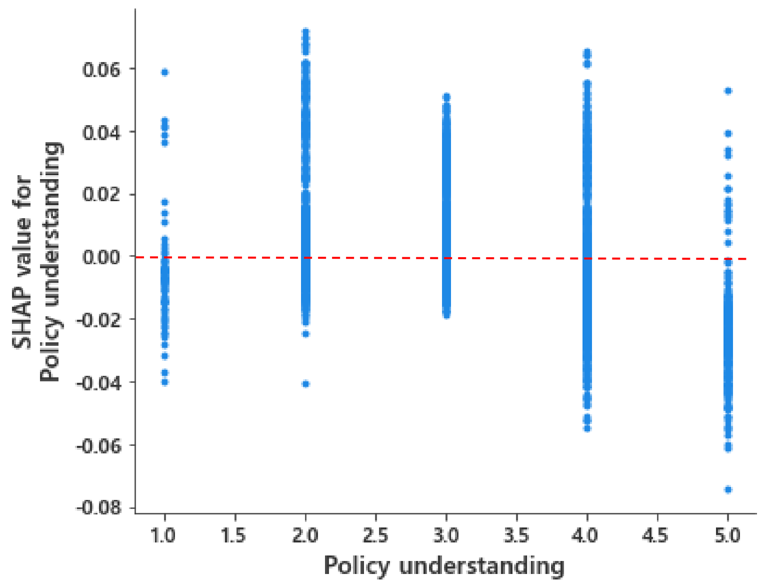

Figure 44. Partial dependence plot of shapley value of level of ZEV subsidy policy understanding on preference towards diesel

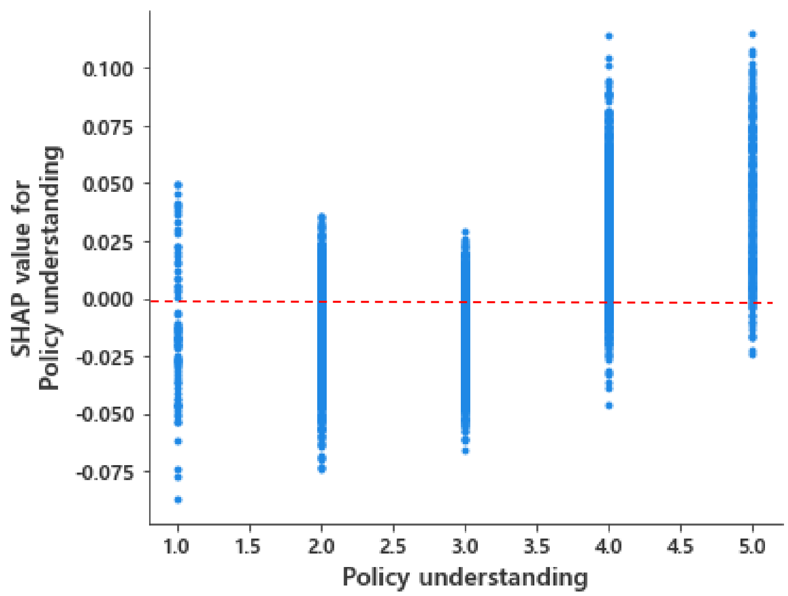

Figure 45. Partial dependence plot of shapley value of level of ZEV subsidy policy understanding on preference towards battery electric vehicle

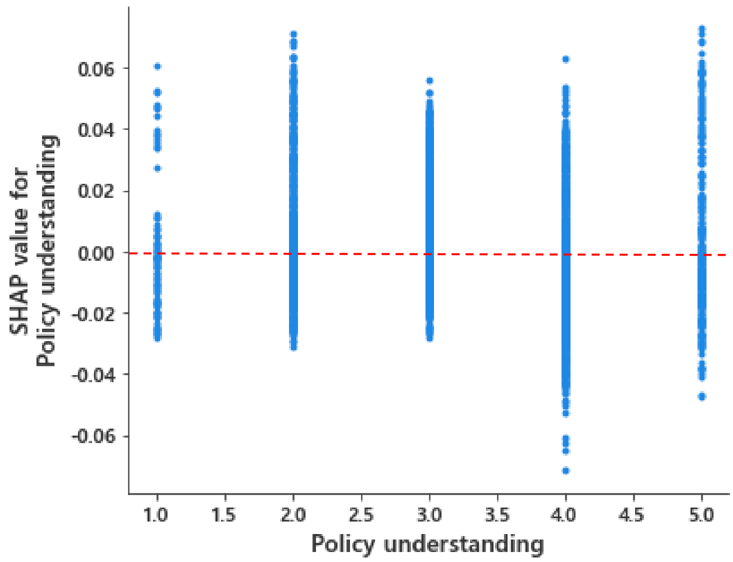

Figure 46. Partial dependence plot of shapley value of level of ZEV subsidy policy understanding on preference towards hydrogen fuel-cell electric vehicle

Figure 43, Figure 44, Figure 45, and Figure 46 are partial dependence plots of shapley value of ZEV policy understanding level on preference towards each vehicle fuel-type, gasoline, diesel, electric, and hydrogen. The x-axis shows the discrete Likert scale given to the respondents. "1" stands for "Very little understanding", "2" stands for "Little understanding", "3" stands for "Moderate understanding", "4" stands for "Well understanding", and "5" stands for "Very well understanding".

Except for the Figure 45, the others showed relatively small relatedness between the preference towards according fuel-type and the understanding level of the ZEV subsidy policy. Figure 45 portrayed non-linear increase of preference towards BEV as the understanding level increased. Since this Likert scale responses are subjective, enhancing the understanding of subsidy policy would not result as increase of preference towards BEV. However, it is critical to make people feel as if they do know more than “moderate” level.

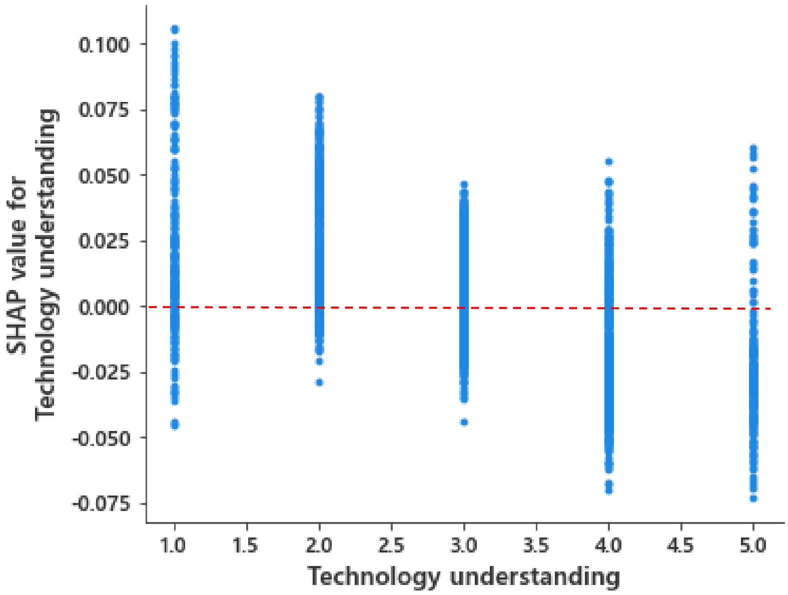

Figure 47. Partial dependence plot of shapley value of level of ZEV technology understanding on preference towards gasoline vehicle

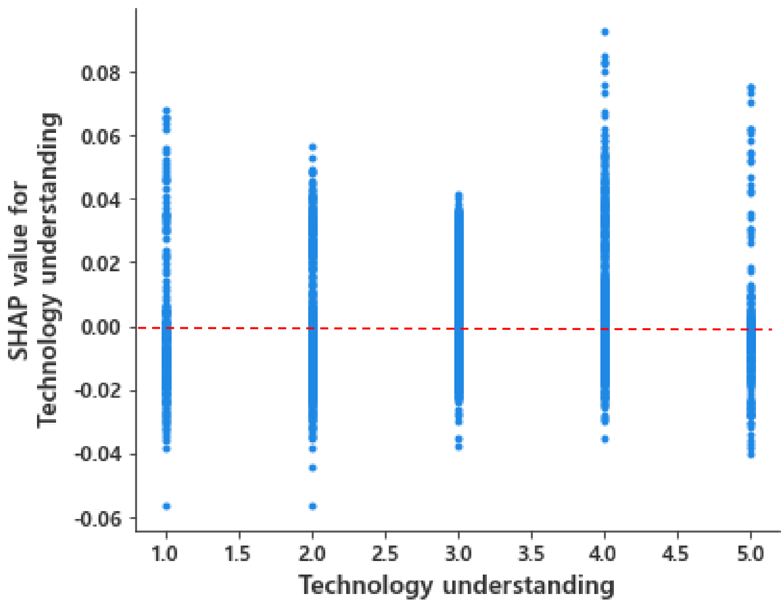

Figure 48. Partial dependence plot of shapley value of level of ZEV technology understanding on preference towards diesel vehicle

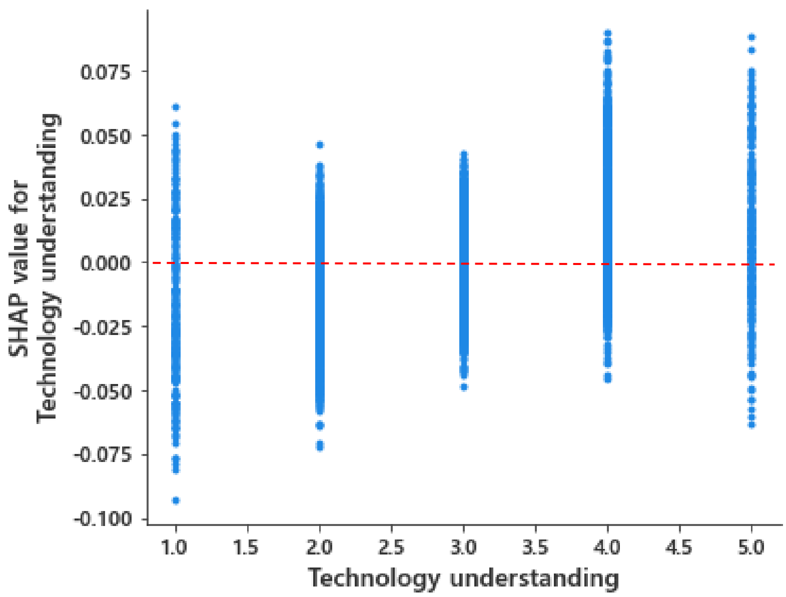

Figure 49. Partial dependence plot of shapley value of level of ZEV technology understanding on preference towards battery electric vehicle

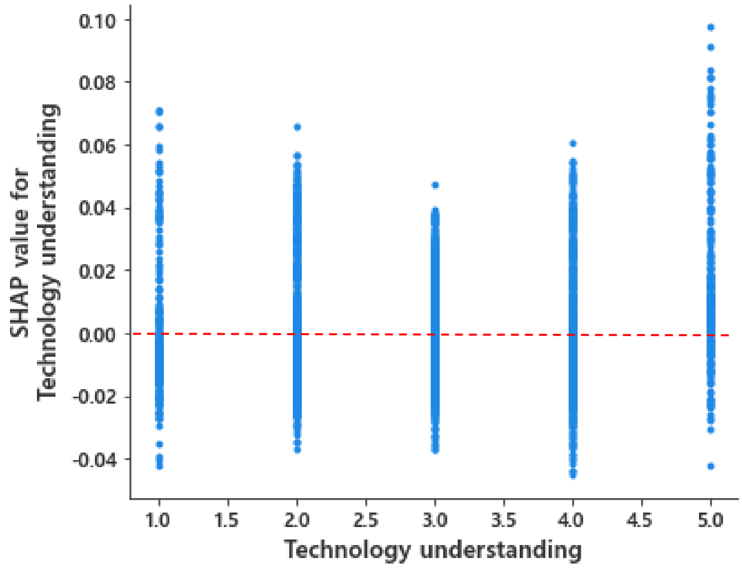

Figure 50. Partial dependence plot of shapley value of level of ZEV technology understanding on preference towards hydrogen fuel-cell electric vehicle

Figure 47, Figure 48, Figure 49, and Figure 50 are the partial dependence plots of shapley value of ZEV technology understanding level for preference towards each vehicle fuel-type, gasoline, diesel, electric, and hydrogen. Just as the previous figures, the x-axis shows the discrete Likert scale given to the respondents. “1” stands for “Very little understanding”, “2” stands for “Little understanding”, “3” stands for “Moderate understanding”, “4” stands for “Well understanding”, and “5” stands for “Very well understanding”.

Except for the gasoline, the other fuel-types showed little relatedness between ZEV technology understanding level and their preferences. In case of gasoline, as the understanding level increased, the preference towards gasoline decreased.

Chapter 5. Discussion

Chapter 5 is for the discussion of this overall study. It is divided into two sections. Section 5.1 is to discuss the key findings and contribution of the research. The following section, Section 5.2 explains the limitations of this study and future research topics to fulfill these limitations.

5.1 Key Findings and Contribution of the Research

The contribution of this research is mainly two points. The first one is the inclusion of status quo, current vehicle status information of drivers. The status quo of vehicles shows the choices the specific drivers made in the previous process of vehicle purchasing. Therefore, it includes the undercovered psychology of customers that they unintentionally consider during the purchase.

However, the relatedness between the currently owning vehicles and the preference towards vehicle options is insufficiently studied. There are two main reasons for this scarcity. The first one is the difficulty of securing the relevant data. In order to pursue this study, both the vehicle status quo of each driver and their preferences towards vehicle features, including the trade-offs, and other personal data were required. The initial data can be found from automobile insurance company; however, it is without the survey data. Therefore, the conjoint survey including the questions checking the vehicle status quo of each respondent is required, as this study did.

In addition, even when the data is secure, the analysis is challenging. As shown in the Appendix 1, because of the collinearity with other features they are easily omitted in order to achieve a solid model from conventional logit regression methods. Or else, some other essential features must be excluded in order to study the importance and impact of status quo information. This study contributed with a new method to overcome this difficulty applying machine learning model, the random forest classifier. SHAP applied to a machine learning model trained with conjoint survey data can provide additional information, along with regression-based models, including the multinomial logit model. Value-oriented feature importance of Shapley provides a better understanding of the impact of each element on each option. In particular, for complex problems, such as policy structuring, a more vivid portrait is worse in rare cases.

The second contribution of this research is from the results of this study, is the effective policy segmentation criteria. With the knowledge of element importance ranking on preference towards each vehicle fuel-type, more efficient and accurate policy and strategy can be built. For instance, Figure 27 and Figure 15 show the partial dependence plots of the two most relevant elements on preference for diesel, the car owner's age, and the average price of currently owned vehicles. From Figure 27, it can be observed that from 30 to 50, the preference for diesel vehicles decreases as the age of car owner increases. However, the 20s and 50s show a relatively nonlinear pattern; people in their 20s generally prefer gasoline vehicles, while the 50-year-olds do not. Figure 15 shows a similar but steeper pattern to that of Figure 15. The preference towards diesel vehicles decreases as the average price of currently owned vehicles reaches 30 million KRW (28,000 USD).

The understanding of diesel PREs leads to the efficient segmentation target of the ZEV induction for those in their 20s to 30s and owning relatively cheaper vehicles because this is the group of car owners with the highest preference for diesel. These segmentation criteria are more potent than others, such as gender or education level, because their mean Shapley absolute values are lower than those of the car owner's age or the average price of status quo vehicles, as shown in Figure 27. In addition, the policy of ZEV induction would be better when differentiating the target segment, depending on whether it is to convince those driving diesel or gasoline vehicles.

The additional major contribution of this study was that it showed diverse non-linear patterns of relationships between the features and the preference towards fuel-types. From Section 4.3, diverse non-linear relationships between each feature and preference towards each fuel-type were revealed. With these plots, two notable implications for policy. The first is a powerful and efficient segmentation criterion for the ZEV induction policy targets. To decrease gasoline preference most effectively, those who currently drive gasoline vehicles under 30 million KRW (28,000 USD) as their most frequently used vehicle shall be the target of induction. However, in the case of gasoline, targeting people in their 20s to 30s who own vehicles under 20 million KRW (18,500 USD) would be a better target.

The second implication was for ZEV infrastructure investment, recharging, and hydrogen refueling. To introduce the FCEV, infrastructure achieving at least 50% of the current gasoline refueling comfort level would be sufficient. However, the preference for BEVs requires a comparable level of recharging comfort to current gasoline refueling. In addition, fuel cell electric vehicles showed a more substitutional relationship with BEVs than with ICEVs.

5.2 Limitations and Future Research Topics

This study has some limitations. First, the discretion of the conjoint survey data may have resulted in incomplete uncovering of the tendencies; in other words, the possibility of tendencies hidden between the revealed data points remains.

In addition, applying machine learning models assuming the conjoint cards' features as additional input variables might not reflect the actual human mechanism of decisions. People specifically compare the levels of same features depending on different conjoint card. However, the machine learning model do not differentiate the input variables into groups or have layered training mechanism. Even though the comparison would be included in the tree building process, whether if the same features were directly compared is insecure.

Finally, implications drawn from the machine learning models must be used cautiously as they are not statistically concrete. The implications would be used for gaining further understanding of the customers regarding vehicle fuel-types, for building efficient policy and strategy. The partial dependence plots can not reveal the exact separating level but only reveal the overall tendencies.

Therefore, further studies to overcome these limitations must be pursued. Developing the un-rigid model, however comparing the level of same features of conjoint cards would bring more accuracy to the prediction and implications as well. In addition, from the conjoint survey data, only the most preferred vehicle choice was analyzed. Combining other responses such as whether if to purchase it within a year, and in case of purchasing whether if to remove a current owning vehicle and replace it or to simply add as a new one. It would give more rigid implications to understand the customer's vehicle selection.

Bibliography

Ahmed, U., & Roorda, M. J. (2022). Modeling freight vehicle type choice using machine learning and discrete choice methods. *Transportation Research Record*, *2676*(2), 541-552.

Al-Alawi, B. M., & Bradley, T. H. (2013). Review of hybrid, plug-in hybrid, and electric vehicle market modeling studies. *Renewable and Sustainable Energy Reviews*, *21*, 190-203.

Berigai Ramaiaha, A., Maurya, R., & Arya, S. R. (2018). Bidirectional converter for electric vehicle battery charging with power quality features. *International Transactions on Electrical Energy Systems*, *28*(9), e2589.

Bhat, C. R., & Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, *41*(5), 506-526.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Byun, H., Shin, J., & Lee, C.-Y. (2018). Using a discrete choice experiment to predict the penetration possibility of environmentally friendly vehicles. *Energy*, *144*, 312-321.

Choi, H., & Koo, Y. (2019). Do I have to buy it now? A vehicle replacement model considering strategic consumer behavior. *Transportation Research Part D: Transport and Environment*, *73*, 318-337.

Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal*, *42*(1), 30-57.

Contestabile, M., Alajaji, M., & Almubarak, B. (2017). Will current electric vehicle policy lead to cost-effective electrification of passenger car transport? *Energy Policy*, *110*, 20-30.

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised learning. In *Machine learning techniques for multimedia* (pp. 21-49). Springer.

Ewing, G. O., & Sarigöllü, E. (1998). Car fuel-type choice under travel demand management and economic incentives. *Transportation Research Part D: Transport and Environment*, *3*(6), 429-444.

Gota, S., Huizenga, C., Peet, K., Medimorec, N., & Bakker, S. (2018). Decarbonising transport to achieve Paris Agreement targets. *Energy Efficiency*, *12*(2), 363-386. https://doi.org/10.1007/s12053-018-9671-3

Green, P. E., & Rao, V. R. (1971). Conjoint measurement-for quantifying judgmental data. *Journal of Marketing research*, *8*(3), 355-363.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research*, *5*(2), 103-123.

Greene, D. L., Ogden, J. M., & Lin, Z. (2020). Challenges in the designing, planning and deployment of hydrogen refueling infrastructure for fuel cell electric vehicles. *ETransportation*, *6*, 100086.

Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, *78*, 273-282.

Hoen, A., & Koetse, M. J. (2014). A choice experiment on alternative fuel vehicle preferences of private car owners in the Netherlands. *Transportation Research Part A: Policy and Practice*, *61*, 199-215.

Jaiswal, D., Deshmukh, A. K., & Thaichon, P. (2022). Who will adopt electric vehicles? Segmenting and exemplifying potential buyer heterogeneity and forthcoming research. *Journal of Retailing and Consumer Services*, *67*, 102969.

Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E., & Sun, Z. (2022). Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: A non-linear and interaction effect analysis. *Journal of Transport Geography*, *103*, 103414.

Junquera, B., Moreno, B., & Álvarez, R. (2016). Analyzing consumer attitudes towards electric vehicle purchasing intentions in Spain: Technological limitations and vehicle confidence. *Technological Forecasting and Social Change*, *109*, 6-14.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review, 105*(5), 491-495.

Knez, M., Jereb, B., Jadraque Gago, E., Rosak-Szyrocka, J., & Obrecht, M. (2021). Features influencing policy recommendations for the promotion of zero-emission vehicles in Slovenia, Spain, and Poland. *Clean technologies and environmental policy, 23*(3), 749-764.

Köhler, J., Wietschel, M., Whitmarsh, L., Keles, D., & Schade, W. (2010). Infrastructure investment for a transition to hydrogen automobiles. *Technological Forecasting and Social Change*, *77*(8), 1237-1248.

Kurani, K. S. (2022). 2021 Zero Emission Vehicle Market Study: Volume 2: Intra-California Regions Defined by Air Districts.

Li, C., Chen, Y., Zhao, Y., Lung, D. C., Ye, Z., Song, W., Liu, F.-F., Cai, J.-P., Wong, W.-M., & Yip, C. C.-Y. (2022). Intravenous injection of coronavirus disease 2019 (COVID-19) mRNA vaccine can induce acute myopericarditis in mouse model. *Clinical Infectious Diseases*, *74*(11), 1933-1950.

Logan, K. G., Nelson, J. D., Lu, X., & Hastings, A. (2020). UK and China: Will electric vehicle integration meet Paris Agreement Targets? *Transportation Research Interdisciplinary Perspectives*, *8*, 100245.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature machine intelligence*, *2*(1), 56-67.

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

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

Mills, M. K. (2008). Environmentally-active consumers' preferences for zero-emission vehicles: public sector and marketing implications. *Journal of Nonprofit & Public Sector Marketing*, *19*(1), 1-33.

Molnar, C., Casalicchio, G., & Bischl, B. (2018). iml: An R package for interpretable machine learning. *Journal of Open Source Software*, *3*(26), 786.

Nastjuk, I., Herrenkind, B., Marrone, M., Brendel, A. B., & Kolbe, L. M. (2020). What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective. *Technological Forecasting and Social Change*, *161*, 120319.

Paredes, M., Hemberg, E., O'Reilly, U.-M., & Zegras, C. (2017). Machine learning or discrete choice models for car ownership demand estimation and prediction? 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS),

Penna, C. C., & Geels, F. W. (2012). Multi-dimensional struggles in the greening of industry: A dialectic issue lifecycle model and case study. *Technological Forecasting and Social Change*, *79*(6), 999-1020.

Polikar, R. (2012). Ensemble learning. In *Ensemble machine learning* (pp. 1-34). Springer.

Saarela, M., & Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, *3*(2), 1-12.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC bioinformatics*, *9*(1), 1-11.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Appendix 1: Mixed Logit Model

*Tool: Stata/SE 17.0

*Status Quo omitted because of collinearity

choice	Coefficient	std. err.	z	P>|z|
option_fuel_type	*0.1691178*	*0.037761*	*4.48*	*0*
2.option_car_class	*0.0170189*	*0.00087*	*19.57*	*0*
option_infra	*-0.0063249*	*0.000753*	*-8.4*	*0*
option_fuel_cost	*-0.0004387*	*4.09E-05*	*-10.74*	*0*
option_price	*0.1691178*	*0.037761*	*4.48*	*0*
Gasoline				
Gender (female)	*-0.0297492*	*0.181395*	*-0.16*	*0.87*
Age	*-0.0088068*	*0.010891*	*-0.81*	*0.419*
Number of car owned	*1.703222*	*0.968743*	*1.76*	*0.079*
Mileage	*-4.72E-07*	*1.94E-06*	*-0.24*	*0.808*
Students	*-0.026653*	*0.106765*	*-0.25*	*0.803*
Policy Understanding	*-0.0817365*	*0.10356*	*-0.79*	*0.43*
Tech Understanding	*-0.164381*	*0.094012*	*-1.75*	*0.08*
Education Level	*-0.1838487*	*0.16148*	*-1.14*	*0.255*
Household Income	*0.0000155*	*0.000387*	*0.04*	*0.968*
Political Orientation	*0.3322118*	*0.104086*	*3.19*	*0.001*
SQ Average Price	*0.0000109*	*2.12E-05*	*0.51*	*0.607*
Diesel				
Gender (female)	*0.2635908*	*0.190521*	*1.38*	*0.167*
Age	*-0.0266786*	*0.011411*	*-2.34*	*0.019*
Number of car owned	*1.294459*	*0.78553*	*1.65*	*0.099*
Mileage	*2.33E-06*	*1.73E-06*	*1.35*	*0.178*
Students	*0.04773*	*0.111498*	*0.43*	*0.669*
Policy Understanding	*-0.2162048*	*0.100325*	*-2.16*	*0.031*
Tech Understanding	*0.0588504*	*0.088354*	*0.67*	*0.505*

Education Level	*0.1345539*	*0.164789*	*0.82*	*0.414*
Household Income	*-0.0005668*	*0.000425*	*-1.33*	*0.182*
Political Orientation	*0.1044727*	*0.108081*	*0.97*	*0.334*
SQ Average Price	*0.0000207*	*2.42E-05*	*0.85*	*0.393*
Electric	*Base alternative*			
Hydrogen				
Gender (female)	*0.0308063*	*0.176151*	*0.17*	*0.861*
Age	*0.013859*	*0.010979*	*1.26*	*0.207*
Number of car owned	*0.8449395*	*0.801231*	*1.05*	*0.292*
Mileage	*1.58E-08*	*1.90E-06*	*0.01*	*0.993*
Students	*-0.0083761*	*0.099665*	*-0.08*	*0.933*
Policy Understanding	*-0.1616189*	*0.094425*	*-1.71*	*0.087*
Tech Understanding	*0.0350541*	*0.08436*	*0.42*	*0.678*
Education Level	*-0.0456217*	*0.161745*	*-0.28*	*0.778*
Household Income	*-0.0005106*	*0.000426*	*-1.2*	*0.23*
Political Orientation	*-0.0051292*	*0.096332*	*-0.05*	*0.958*
SQ Average Price	*-6.53E-06*	*2.86E-05*	*-0.23*	*0.82*

Abstract (Korean)

자동차 연료유형에 따라 그 선호도에 영향을 미치는 요인은 다르고, 이에 대한 이해를 통해 효과적인 정책 대상 도출이 가능하다. 본 연구는 친환경차 유도정책의 대상을 효율적으로 도출하기 위해 4개의 주요 자동차 연료유형(가솔린, 디젤, 전기, 수소) 각각의 선호도와 관련된 요소 차이들이 연료유형에 따라 가지는 차이를 파악하는 새로운 방법을 제시하는 것을 목적으로 한다. 선호도에 미치는 요인들의 목록은 현재 보유 중인 차량 정보를 포함하는 차량 소유자의 속성과 차량 자체의 속성을 모두 포함했다. 연료유형 별 선호도에 미치는 요인들의 중요도 비교는 컨조인트 설문조사 결과 얻은 데이터를 학습시킨 random forest classifier 모델에 게임이론을 기반으로 한 SHAP의 적용을 통해 가능했다. 각 연료유형의 요인별 섀플리 값과 이에 대한 partial dependence plot을 비교한 결과, 효과적인 정책 대상을 도출할 수 있었다. 또한, 전기차 충전 및 수소차의 수소 충전 기반시설에 대한 투자 정책도 보다 효과적으로 도출할 수 있었다. 마지막으로, SHAP 등 XAI 도구를 이용함으로써 상호연관성이 높아 기존의 방식으로 분석하기 어려웠거나, 데이터 종류가 많아 기존 모델로 정확한 분석 결과를 얻기 어려웠던 현재 보유 중인 차량 정보 등의 요인들을 동시에 고려할 수 있었다.

주요어 : 컨조인트 설문, 머신 러닝, 섀플리 값, 자동차 연료유형, 정책 대상

도출, 친환경차 기반시설

학 번 : 2021-21881