Ph. D. Dissertation in Economics

Development of a Demand Forecasting Methodology with Aggregate Market Data:

Focusing on Diffusion and Choice Model

시장 데이터를 이용한 수요예측 방법론의 개발

August 2014

Graduate School of Seoul National University

Technology Management, Economics, and Policy Program

Yeunjoong Kim

Development of a Demand Forecasting Methodology with Aggregate Market Data:

Focusing on Diffusion and Choice Model

지도교수 이 종 수

이 논문을 경제학박사학위 논문으로 제출함

2014 년 8 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

김 연 중

김연중의 경제학박사학위 논문을 인준함

2014 년 8 월

위 원 장 _____(인)

부위원장 _____(인)

위 원 _____(인)

위 원 _____(인)

위 원 _____(인)

Abstract

Development of a Demand Forecasting Methodology with Aggregate Market Data:

Focusing on Diffusion and Choice Model

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

A variety of demand forecasting methods are used, and each of method has its own advantages and disadvantages. The diffusion model is used to predict the demands of a new product or service with only a few observations; however, it is sometimes inappropriate when political intervention exists. On the other hand, to estimate heterogeneous consumer preference, most studies have used individual-level data because the estimation could not be conducted properly with aggregate-level data, which make it difficult to overcome the implicit problems in the models. The purpose of this dissertation is to propose demand forecasting methodologies to overcome the limitations of aggregate market data and to demonstrate their applications through two essays.

First, it is shown that the classical diffusion model is not suitable to predict the renewable energy technology market, which is highly influenced by policy intervention. To resolve this issue, the adjusted logistic model is proposed by focusing on the policy's role as a change agent and a facilitator. A politically planned goal is introduced in the model as an explanatory variable, and the gap between the actual diffusion pattern and the politically planned goal is explained by the learning-by-doing effect. The estimated result of the adjusted logistic model is statistically significant, and the model shows more accurate forecasts and market potential than do other models. In addition, based on the choice model that considers a decision maker's utility, the diffusion pattern of renewable energy technology is reinterpreted to measure the effect of the government subsidy, and the proper subsidy portfolio is suggested through dynamic programming for the successive diffusion of the renewable technology.

Second, the BLP model with hierarchical preference is introduced to identify the heterogeneous consumer preference and predict a market share with aggregate market data only. The hierarchical preference divides the consumer utility into homogeneous and heterogeneous parts. With multiple aggregate market data, the two different preferences are measured and incorporated with a new product in a new market for forecasting. Based on the estimated consumer preference and socio-demographic information, it is possible to identify the ideal consumer group and to establish the profit-maximizing marketing strategy. As an empirical study, the consumer preference for automobiles in four U.S. states is analyzed and the forecasted and actual market shares in a new market are

compared. The suggested methodology shows high predictability and the target market for profit maximization is suitably identified based on consumer socio-demographic information and environmental conditions.

In summary, this dissertation presents how aggregate market data are incorporated into demand forecasting, and its applications are provided through two essays.

Keywords: Demand forecasting, Diffusion model, Choice model, Aggregate market data
Student Number: 2009-21082

Contents

Contents vi

List of Tables ix

List of Figures.....xi

Chapter 1. Introduction 1

1.1 Research Background 1

1.2 Objectives and Outline 5

Chapter 2. Overview of Diffusion Model and Choice Models.....10

2.1 Limited Observations on the Initial Market: The Diffusion Model Approach.....10

2.1.1 Key Factors Affecting the Diffusion Process..... 10

2.1.2 Diffusion Models..... 13

2.1.3 Motivation..... 21

2.2 Choice Models with Individual-level and Aggregate-level Data 27

2.2.1 Logit Model and its Extensions..... 28

2.2.2 Discrete Choice Experiments with Individual-level Data 33

2.2.3 Motivation..... 36

Chapter 3. Diffusion Forecasting for Renewable Energy Technology in South Korea..... 39

3.1 Introduction.....40

3.1.1 Renewable Energy Technology (RET) and its Barriers..... 40

3.1.2 RET Diffusion Studies Considering Policy..... 45

3.2 Adjusted Logistic Model Considering Policy Goals 50

3.2.1 Model Specification 50

3.2.2 Comparison of Forecasting Accuracy 56

3.3 Diffusion Pattern Forecasting Based on the Aggregate Market Data 64

3.4 Re-interpretation by Choice Model 71

3.4.1 Model and Data 71

3.4.2 Results 75

3.5 Cost Minimization for Policy Makers 83

3.6 Summary and Discussion.....88

Chapter 4. Automobile Demand Forecasting in North America.....91

4.1 Introduction.....92

4.2 Demand Forecasting Methodology Based on the BLP model 94

4.2.1 BLP model 94

4.2.2 Endogeneity and Instrument Variables..... 98

4.2.3 Model Specifications 102

4.3 Analysis of Consumer Preference 108

4.3.1 Model and Data 108

4.3.2 Results 112

4.4 Forecasting..... 118

4.5 Market Segmentation.....123

4.6 Summary and Discussion.....128

Chapter 5. Conclusion 131

5.1 Summary 131

5.2 Contributions and Limitations 135

Bibliography 140

Abstract (Korean) 160

List of Tables

Table 1. Extended diffusion models 20
Table 2. Estimated parameter p in previous literature 24
Table 3. Estimation results of the Bass diffusion model and the logistic model 25
Table 4. Previous studies using DCE 35
Table 5. Summary of previous literature using diffusion model 49
Table 6. Estimation results of the three models 58
Table 7. Calculated forecasts of the three models (unit: kW) 59
Table 8. Comparison of forecasting accuracies between the three models 60
Table 9. RETs plant capacity in South Korea 67
Table 10. Estimation results of the proposed model 68
Table 11. Estimation result based on the binary logit demand model 76
Table 12. The South Korean government's planned subsidy expenditure for RET plant diffusion 77
Table 13. Actual subsidy expenditure for RETs by the South Korean government 78
Table 14. Subsidy expenditure portfolios for RET diffusion (unit: KRW billion) 80
Table 15. Estimation results for *Model 1* 113
Table 16. Estimation results for *Model 2* 115
Table 17. A sample of the average of the own and cross-price elasticity for midsize sedans in 2005, California (*Model 1*) 117
Table 18. A sample of the average of own and cross price elasticity for compact sedan in August, Texas (*Model 2*) 118
Table 19. The percentage of actual market shares in the prediction interval 122

Table 20. The optimal consumer group for market share maximization (*Model 1*) 126
Table 21. The optimal region for market share maximization (*Model 2*). 127

List of Figures

Figure 1. Outline of dissertation 7
Figure 2. Key factors affecting the rate of diffusion 11
Figure 3. Installed capacity of RET plants using the Bass model and the logistic model (unit: kW) 26
Figure 4. Research framework for essay 1 40
Figure 5. Barriers to RET diffusion in Asia 42
Figure 6. Planned changes in RET plant capacity by policy plans (unit: kW) 51
Figure 7. Actual and planned RET plant capacity in South Korea (unit: kW) 53
Figure 8. Capacity data in 2012 and estimation results (unit: kW) 56
Figure 9. Diffusion patterns of the three models (unit: kW) 59
Figure 10. Diffusion patterns of the three models until 2027 (unit: kW) ... 61
Figure 11. RET diffusion patterns by estimated results (unit: kW) 70
Figure 12. Diffusion patterns by expenditure portfolios (unit: kW) 82
Figure 13. The cost-minimized portfolio and its diffusion pattern 86
Figure 14. Research framework for essay 2 92
Figure 15. The forecasted market shares in Ontario based on *Model 2* (January 2005) 123

Chapter 1. Introduction

1.1 Research Background

Forecasting is important in various fields. At the individual level, people want to know tomorrow's weather or predict how many guests will come to a birthday party. At the institutional level, forecasts are significant for new product supply planning, logistics, and investment. For governments, economic forecasts play a critical role in establishing new policy. Accurate forecasting supports the execution of marketing strategy and policy, while improper forecasts result in the wasteful use of resources.

Various forecasting studies have been conducted in the last half century. In a broader context, demand forecasting studies can be divided into two categories, qualitative and quantitative analyses.

Commonly used qualitative analyses include judgmental bootstrapping, the Delphi method, conjoint analysis, and index analysis. Judgmental bootstrapping and Delphi employ expert opinions when data collection is not available. Judgmental bootstrapping is based on a virtual scenario to predict professional expectations, while the Delphi method seeks to reach a consensus on forecasting through repeated discussions (Ashton et al., 1994; Milkovich et al., 1972; Lawrence et al., 2000; Rowe and Wright, 1999; Bosun and

Modrak, 2014; Kauko and Palmroos, 2014). Conjoint analysis was developed in the early 1970s to make a series of trade-offs among the product attributes (Hauser and Rao, 2004; Eggers and Eggers, 2011). Index analysis focuses on a small number of alternatives to predict the choice. This method is occasionally used to analyze presidential elections (Armstrong and Graefe, 2011; Cuzán and Bundrick, 2009). Qualitative analysis is useful when there are data shortcomings or rapid changes in demand; however, it is highly subjective.

Quantitative analysis overcomes this limitation of qualitative analysis by employing numerical or mathematical data. Regression analysis, time series analysis, and diffusion models are well-known quantitative analysis methods. Regression analysis determines the relationships among variables, and it is used not only for forecasting but also for sensitivity analysis in many fields (Bianco et al., 2009; Pankraz, 2012). While regression analysis is used to examine a hypothesis with independent time series data, time series analysis is a statistical model for forecasting based on long-term data. From macro-economics to the financial industry, time series analysis is widely used to forecast (Louviere and Hensher, 1983; Kulendran and King, 1997; Ediger and Akar, 2007; Lim and McAleer, 2002; Saab et al., 2001). Diffusion models are used to forecast demand for a product or service with the assumption that the diffusing pattern follows an S-curve. Communication channels, including both external and internal influences, are the key factors in investigating future demand.

Both qualitative and quantitative analyses have pros and cons, and the forecasting

methods for each case are determined according to the objectives of demand forecasting and data availability. The methods are frequently adjusted to complement each other. An advanced approach to overcoming the limitations of quantitative analysis is to combine it with another mathematical method with a theoretical background (Clemen, 1989).

One of the commonly used mathematical methods is the choice model, which was originally developed by Thurstone (1927) for cognitive psychology and which explains consumer behavior with respect to comparative judgment. Following Thurstone (1927), Luce (1959) and Marschak (1960) show that the discrete choice model is consistent with utility maximization with the logit formula. McFadden (1974), who received a Nobel Prize for economics for choice modeling theory, coordinates econometric theory and its empirical application for the discrete choice model. However, the choice model can provide only choice probability information about consumer preferences. For complete demand forecasting, the estimate for the entire market potential is required. The diffusion model can be an appropriate starting point to derive the market potential since it is able to forecast the diffusing pattern only at the aggregate level. By integrating the market potential and choice probability, better demand forecasting estimates can be obtained, and synergy between qualitative analysis and econometric theory can be realized.

Another well-known integrated approach is the discrete choice experiment (DCE hereafter), which combines conjoint analysis and choice model analysis. DCE is a popular analysis tool because it inherits great advantages from both analyses (Zwerina,

1997). First, conjoint analysis provides the stated preference data from hypothetical scenarios and thus increases the ease of data acquisition. Second, the DCE is able to simply calculate the market share based on the choice model simulation. Third, unique effects can be easily introduced by adding an attribute or dummy, and willingness to pay (WTP hereafter) can be indirectly estimated based on consumer characteristics. With these advantages, the DCE has been widely applied in various fields. In the marketing field, estimated consumer preference is helpful to develop a marketing strategy and design a new product (Louviere and Woodworth, 1983). In health economic, the DCE has contributed to the evaluation of healthcare service benefit assessments with patients' preferences (Ryan and Gerard, 2003). Recently, the policy perspective has been analyzed with DCE as the consumer preference for policy support to ensure legitimacy.

Although combined demand forecasting methods have outstanding advantages, there are some barriers to overcome. In the case of diffusion, the model depends highly on the communication factor. This property explains the diffusion of new durable goods and innovative services under free market competition; however, when diffusion is driven by artificial influence, such as the promotion efforts of the change agents and facilitators, the underlying assumptions should be modified for better forecasting. A common problem in DCE analysis is that estimation requires a large number of conjoint observations to identify the consumer preference. Surveys are mandatory to fulfill the requirement, and surveys might cause the hypothetical bias problem that respondents always indicate a higher willingness to pay in a virtual situation than in a real market situation. The cost for

surveys is also needed.

In this dissertation, two essays that use aggregate market data based on quantitative demand forecasting analysis are provided. First, political intervention is considered using the diffusion model. On the basis of the classical diffusion models, political intervention is introduced to explain the induced diffusion pattern by the efforts of change agents and a comparison of accuracy is conducted. The estimated results are then used to recommend a subsidy plan. Second, a new demand forecasting platform that does not need individual survey data is described, and its application by marketers is discussed. Specifically, an ideal consumer group and region are found in the U.S. automobile market.

1.2 Objectives and Outline

The purpose of this research is to provide an effective demand forecasting methodology with aggregate-level market data. Two essays will be provided based on the diffusion and choice models, respectively.

In the first essay, a new diffusion model is developed to explain the induced diffusion, focusing on the efforts of the change agent and the facilitator. Since the 1960s, the diffusion theory has been expanded to various kinds of innovation diffusion. However, induced diffusion that is driven by the efforts of the change agent and the facilitator has received little development. With the emergence of Renewable Energy Technology (RET hereafter), a policy plan that can be interpreted as an effort of the change agent has

become an important factor affecting diffusion. Therefore, the first essay proposes a new diffusion model to forecast the induced diffusion, focusing on South Korean RET markets. In addition, it is shown that the additionally introduced policy plan in the estimation supplements the data availability and properly calibrates the market potential.

Second, a platform is suggested that can measure the consumer preference as accurately with only aggregate-level data as with individual-level data. For the most part, missing individual-level data can be collected through a survey with stated preference data; in contrast, aggregate-level revealed data are relatively easy to acquire. To avoid the burden and intrinsic problems of surveys, a forecasting platform that can measure the heterogeneous consumer preference with aggregate-level data is necessary. The identified preference structure provides accurate demand forecasts and predicts the future market situation when a new product is released. The U.S. automobile market is empirically studied in the second essay.

This dissertation consists of five chapters. Figure 1 shows the outline of the dissertation.

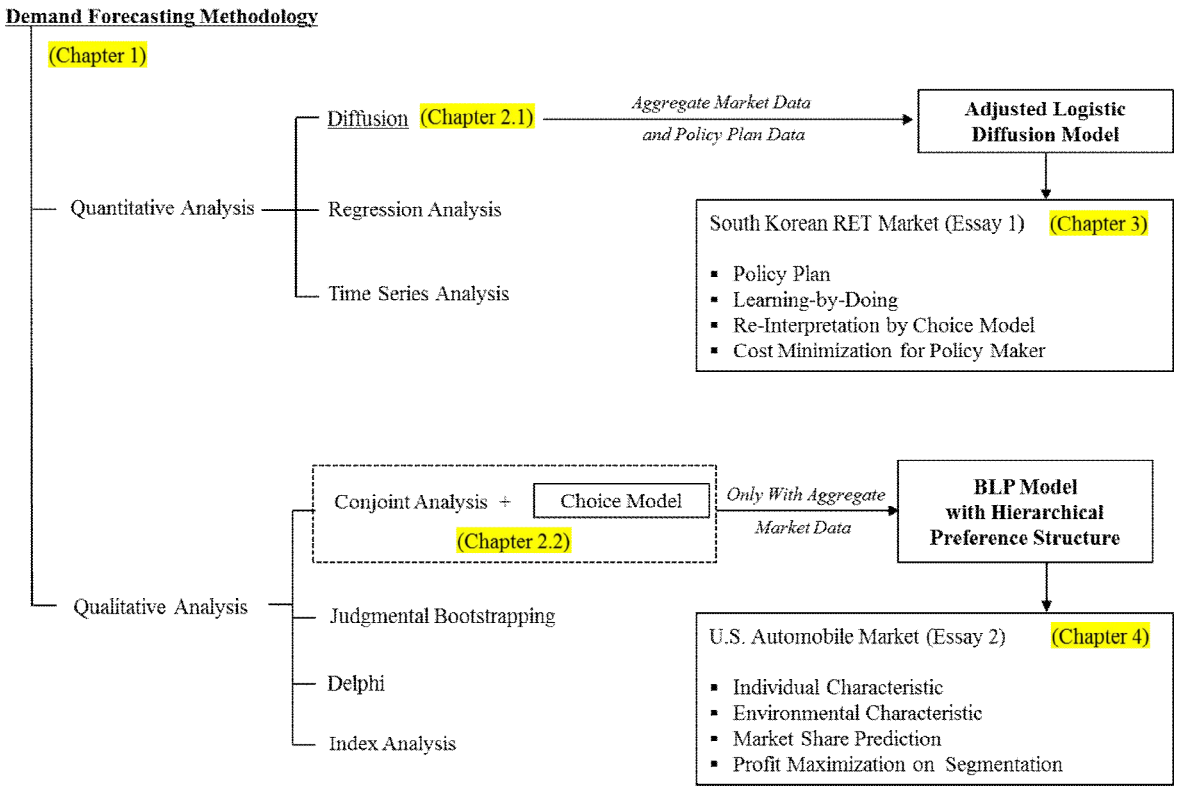

Figure 1. Outline of dissertation

Chapter 2 reviews the previous literature on diffusion theory and choice models. The key factors that affect diffusion are introduced and reviewed. The mathematical specifications of the Bass model and the logistic model are explained as representative diffusion models, and expanded diffusion models are briefly discussed. The limitations of previous studies are also described, focusing on RET diffusion. An overview of basic choice models is provided, and the DCE studies using individual-level data are introduced. The dependence of DCE analysis on surveys is discussed, and the intrinsic problems with stated preference data are explained.

Chapter 3, the first essay, analyzes the diffusion of South Korean RET plant capacity. The adjusted logistic model is developed by introducing the policy plan and applying it to South Korean RET diffusion. The adequacy of the estimated diffusion pattern is reviewed with the planned diffusion pattern. Then, the forecasted diffusion pattern is re-interpreted in combination with the choice model to derive the cost minimization problem of policy makers. Finally, the cost minimization problem is solved through dynamic programming, providing political implications.

Chapter 4, the second essay, is a consumer preference analysis of the U.S. automobile market. The forecasting methodology using a random coefficient logit model is described. It is then extended to market segmentation from the perspective of companies. Individual and environmental characteristics are introduced to investigate the heterogeneity, and the estimated results are used to forecast market shares in a new market. Additionally, market segmentation is conducted to maximize the market share. The segmentation result

provides evidence that heterogeneity assists in identifying the ideal consumer group according to socio-demographics.

Finally, *Chapter 5* summarizes the two essays, and this dissertation's main contributions and limitations are discussed.

Chapter 2. Overview of Diffusion Model and Choice Models

2.1 Limited Observations on the Initial Market: The Diffusion Model Approach

2.1.1 Key Factors Affecting the Diffusion Process

The most commonly used definition of diffusion in both economics and marketing comes from Rogers (2003)[1]. He defines diffusion as the process by which innovation is communicated among people in a social system through certain channels over time. Four elements can be found in the definition: innovation, the communication channel, time, and the social system. These are the key elements of the diffusion process; they provide the basis of the research direction. Innovation includes ideas, practices, and objects that are newly recognized by individuals or groups. The communication channel is the path the message goes through, such as mass media and interpersonal channels. Time describes the process from recognition to the adoption or rejection of the innovation by individuals. The social system is the aggregate of individuals or groups that adopt the

[1] Rogers (2003) is the revised fifth edition; the first edition was published in 1962.

innovation.

Rogers (2003) suggested four key factors that influence the diffusion of innovation: the perceived attributes of the innovation, the type of innovation decision, the nature of the social system and the communication channel, and the extent of change in agents' promotion efforts (Figure 2).

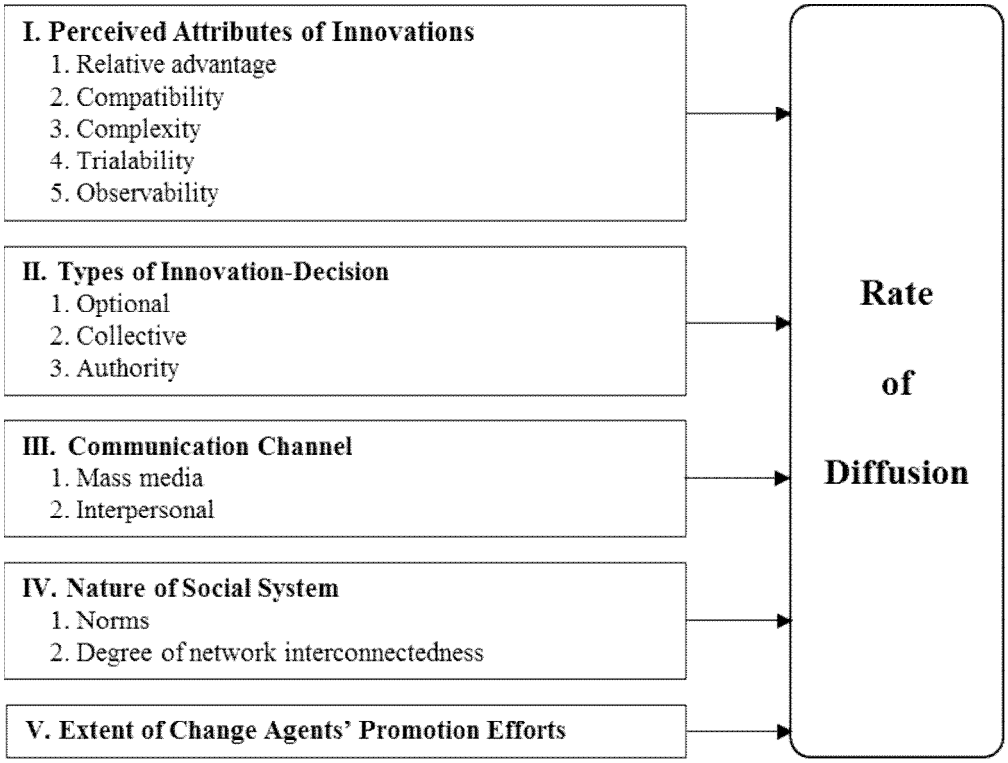

Figure 2. Key factors affecting the rate of diffusion

The perceived attributes are connected with consumers' utilities directly; however, each factor has a complex effect (Gerrard and Cunningham, 2003). For example, internet

banking provides great mobility as a relative advantage, but it also increases the complexity risk and compatibility problems. If the consumer uses a PC, his or her utility is higher than others who do not use a PC. Therefore, to measure the influence of the perceived attributes, the characteristics of innovation should be categorized properly.

The innovation decision is sometimes made by a group or organization. When the decision is made by consensus among the members, all the members follow the decision, and it is called a collective innovation decision. Examples include a hospital adopting medical technologies such as radioisotopes and steel manufacturers adopting basic oxygen furnaces (Rapoport, 1978; Oster, 1982). While optional innovation decisions are made by individuals, an authority innovation decision is made by a relatively small number of individuals who have power or expertise.

Communication channels are classified into mass media channels and interpersonal channels. Mass media channels can immediately transmit messages to many social members; they include newspapers, television, and radio. Interpersonal channels are most effective if they enable face-to-face exchanges between people with similar socioeconomic backgrounds. These two channels represent both external and internal influences and are the constituents of the basic model. Since Bass (1969) suggested the diffusion model, communication channels have been widely used to develop mathematical models. A more extensive discussion is provided in *Chapter 2.1.2*.

The social system can affect the diffusion of innovation even when individuals have the same background characteristics, such as education level, age, and income. Rogers

and Kincaid (1981) show through a South Korean case study that different diffusion patterns of family planning programs exist because differing system structures and village norms affect decision making.

Opinion leaders and change agents play an important role when they have high levels of influencing power and network interconnectedness. In the past, the level of networking power was not high, but the influence of opinion leaders has increased as communication networks have become larger (Rogers, 2003). Kwon et al. (2009) find support for this in a case study showing that 85% of information diffusion occurs among non-explicit relationships in blog networking.

The five key factors that facilitate the diffusion of innovation are reviewed. In the next chapter, various diffusion models are introduced in terms of communication channels.

2.1.2 **Diffusion Models**

Research on the diffusion of an innovation has been performed to forecast future demand based on the definition of diffusion by Rogers (2003). As previously reviewed, communication channels are the one of key factors in explaining the diffusion pattern. Well-known basic diffusion models that generally assume first purchase of new products or services by a population are conceived by Bass (1969) and Mansfield (1961). Bass considers only two segments, market potential and current market size, and his research

focuses on both external and internal influences so that it is a mixed influence model. The external influences that promote the adoption of an innovation through mass media, such as advertising, are defined as *innovation factors*. Some of the early adopters encourage others to adopt the innovation through word-of-mouth, and these internal influences are known as *imitation factors*. In contrast, the logistic model derived by Mansfield (1961) depends only on internal influences. The modeling approach of Mansfield is quite different from Bass (1969), and it will be discussed after the Bass (1969). The general form of the diffusion model is described in the following equation (Mahajan and Peterson, 1985):

$$\frac{dN(t)}{dt} = g(t)\left[\bar{N} - N(t)\right], \quad \cdots\cdots \text{Eq. (1)}$$

where $dN(t)/dt$ is the rate of diffusion at time t, $N(t)$ is the cumulative number of adopters at time t, $\bar{N}$ is the number of potential adopters in the social system, and $g(t)$ is the coefficient of diffusion or adoption probability at time t. $g(t)$ determines the rate of diffusion according to its functional forms.

External Influence Diffusion Model: $g(t) = p$; $\cdots\cdots$ Eq. (2)

Internal Influence Diffusion Model: $g(t) = qN(t)$; $\cdots\cdots$ Eq. (3)

Mixed Influence Diffusion Model: $g(t) = p + qN(t)$, $\cdots\cdots$ Eq. (4)

where p and q are commonly called the "coefficient of innovation" and the "coefficient of imitation," respectively. Since Eq. (1) is a differential equation, the solved Bass model is expressed by integration as

$$N(t) = \bar{N}\left[\frac{p(p+q)^2 e^{-(p+q)t}}{(p+qe^{-(p+q)t})^2}\right]. \quad \text{Eq. (5)}$$

Following the same logic used to derive Equation (3), the logistic model that introduces internal influence only can be simplified as

$$N(t) = \bar{N}\left(\frac{1}{1+e^{-(a+qt)}}\right), \quad \text{Eq. (6)}$$

where a is a constant. However, Mansfield (1961) uses a different process to describe Eq. (6). The diffusion model proposed in this study is based on the logistic model, so it is explained in detail with reference to Mansfield (1961).

Mansfield (1961) first defines the deterministic model for the diffusion rate. Let N_{ij} be the total number of firms[2] that are able to adopt innovation j in industry i, and let

[2] This is also called "market potential" in diffusion theory.

F_{ij} be the number of firms that have adopted innovation j in industry i.[3] The diffusion rate, $\lambda_{ij}(t)$, is defined as the proportion of firms that have not adopted innovation j at time t but have adopted it by time $t+1$.

$$\lambda_{ij}(t) = \frac{F_{ij}(t+1) - F_{ij}(t)}{N_{ij} - F_{ij}(t)}. \quad \cdots\cdots \text{Eq. (7)}$$

In addition, Mansfield (1961) introduces the function of the diffusion rate with some variables to explain the diffusion rate.

$$\lambda_{ij}(t) = f_i\left(\frac{F_{ij}(t)}{N_{ij}}, \pi_{ij}, Q_{ij}\right), \quad \cdots\cdots \text{Eq. (8)}$$

where π_{ij} is the profitability of adopting this innovation and Q_{ij} is the investment required to adopt this innovation as a percentage of the total assets of firms.

In Eq. (8)[4], the first term indicates the network effect; this is the same effect as that of the imitation factor in the Bass model. Mansfield (1961) assumes that the innovation becomes less risky as innovation information and accumulated experience increase. The

---

[3] The notation differs from that in Mansfield (1961) to facilitate discussion related to the choice model in the following chapters.

[4] Precisely, Mansfield (1961) tests the function with four other variables: the equipment life cycle for the innovation, the annual growth of industry sales during the period, the year when the innovation was introduced, and the phase of the business cycle during which the innovation was introduced. However, all the variables are found to be non-significant. In this study, only three variables are considered in Eq. (8) for simplicity.

second term is introduced because profitable firms tend to be able to afford to take a risk with the innovation. The third term indicates that for the same profitable innovation, relatively large investment disrupts innovation adoption. Finally, the function is dependent on the industry because the diffusion rate differs among industries—for example, the rate is high in competitive and financially stable markets.

Mansfield (1961) assumes based on Eq. (8) that the diffusion rate can be expressed within an appropriate range by Taylor's expansion, which omits the third and higher terms and sets the effect of $\left(\frac{F_{ij}(t)}{N_{ij}}\right)^2$ as zero. The solved differential equation is as follows:

$$F_{ij}(t) = \frac{N_{ij}\left[e^{l_{ij}+(P_{ij}+\phi_{ij})t} - (P_{ij}/\phi_{ij})\right]}{1+e^{\left[l_{ij}+(P_{ij}+\phi_{ij})t\right]}}, \quad \text{Eq. (9)}$$

where l_{ij} is a constant of integration, P_{ij} is the sum of all terms in the Taylor expansion not containing $\frac{F_{ij}(t)}{N_{ij}}$, and ϕ_{ij} is the coefficients of $\frac{F_{ij}(t)}{N_{ij}}$ in the Taylor expansion.

Additionally, the assumption that the number of firms is zero when the innovation is first introduced is applied as

$$\lim_{t\to-\infty} F_{ij}(t) = 0 . \quad \text{Eq. (10)}$$

The above condition makes P_{ij} zero, so the equation is summarized as

$$F_{ij}(t) = \frac{N_{ij}}{1 + e^{-(l_{ij} + \phi_{ij} \cdot t)}} \quad \text{Eq. (11)}$$

$$\phi_{ij} = a_0 + a_1 \pi_{ij} + a_2 Q_{ij} \quad \text{Eq. (12)}$$

Therefore, the diffusion pattern is governed by the imitation factor ϕ_{ij}. The imitation factor is explained by the profitability and the required investment; on the other hand, the imitation factor in the Bass model is simply assumed to be a constant as in Eq. (6). Mansfield (1961) adds a random variable with zero expected value into Eq. (12) and estimates the parameters a_0, a_1, and a_2 based on the estimated value of ϕ_{ij}. Data from twelve innovations in four industries show statistical significance[5].

Although the Bass and logistic models are derived from different processes, both yield S-shaped cumulative adopter distributions and share some assumptions, such as single-product adoption and fixed market potential. These assumptions are sometimes unrealistic but increase estimating performance even with limited data. To overcome

[5] The four industries are bituminous coal, iron and steel, brewing, and railroads. The twelve innovations are the shuttle car, the trackless mobile loader, and the continuous mining machine (in bituminous coal); the byproduct coke oven, the continuous wide strip mill, and the continuous annealing line for tin plate (in iron and steel); the pallet-loading machine, the tin container, and the high-speed bottle filler (in brewing); and the diesel locomotive, centralized traffic control, and car retarders (in railroads).

these assumptions, various extensions, such as the multi-innovation diffusion model (Mahajan and Peterson, 1978) and the dynamic diffusion model (Bayus, 1987), have been proposed.

Multi-innovation diffusion models can be applied to situations in which the diffusion of innovation is influenced by other competitive innovations or incumbent technologies. If an innovation is diffused under the competition environment, a substitute or complementary relationship should be considered. Bayus et al. (2000) review previous diffusion studies and summarize the conceptual framework for multi-product interactions.

Many studies have expanded the Bass and logistic models and have provided meaningful results (Mahajan et al., 1990; Mahajan et al., 1995; Sultan et al., 1990; Teotia and Raju, 1986; Frank, 2004; Mahajan et al., 1993; Bass et al., 1994; Krishnan et al., 2000). The details are provided in Table 1.

Table 1. Extended diffusion models

Model	Released assumption	Examples
Dynamic diffusion model	The saturation level can be changed by exogenous or endogenous factors.	Mahajan and Peterson (1978), Lackman (1978), Chow (1967)
Flexible diffusion model	The diffusion pattern may have an asymmetric S-shape.	Von Bertalanffy (1957), Floyd (1962), Sharif and Kabir (1976), Jeuland (1981), Easingwood et al. (1981), Easingwood et al. (1983)
Generalized Bass model	The diffusion pattern incorporates the effects of marketing-mix variables in the likelihood of adoption.	Bass et al. (1994)
Multi-adoption model	Innovation is adopted repeatedly.	Lilien et al. (1981)
Multi-innovation model	An innovation is contingent on other innovations; for example, it can be complementary or a substitute.	Peterson and Mahajan (1978), Bayus (1987), Givon et al. (1995), Givon et al. (1997), Prasad and Mahajan (2003)
Multi-stage diffusion model	Innovation adoption is generated through multiple stages.	Dodson and Muller (1978), Mahajan et al. (1984)
Successive generation diffusion model	The nature of innovation can change over time and be multi-generational.	Norton and Bass (1987), Mahajan and Muller (1996), Islam and Meade (1997), Jun and Park (1999)
Space and time diffusion model	The dimensions of time and space should be considered simultaneously.	Mahajan and Peterson (1979)

2.1.3 **Motivation**

Although the Bass and logistic models have helped to develop diffusion theory since the 1960s, these models have several limitations when applied to RETs. First, there are strong policy support measures and regulations in RET markets. RETs have quite different characteristics from the consumer durables that Bass (1969) mainly discusses. In general, new technology is introduced when profits are expected to be greater than the costs of the introduction (Hall and Khan, 2003). However, to be selected as a power generating method, RETs are not price competitive enough to compete with existing energy technologies. According to Rao and Kishore (2010), RETs have a considerably slower diffusion rate compared to other commercial products, even though governments and public institutions provide support programs or even subsidies for RETs. In addition, Purohit and Michaelowa (2008) show that diesel irrigation water pumps are preferred to Solar Photovoltaic (SPV hereafter) pumps in India even if the government were to maintain subsidy support programs at the same level as it had operated up until then. RETs are still highly dependent on government subsidies and incentives, and many researchers claim that political support and regulations affect technology choice (Gray and Shadbegian, 1998; Jaffe et al., 2005; Kemp and Volpi, 2008; Orr, 1976; Stoneman and Diederen, 1994). Without such support, RETs would hardly be adopted owing to high associated energy costs. Rao and Kishore (2010) point out that the patterns of diffusion of RETs are very complex so that previous diffusion models are proper to apply to

commercial products rather than renewable energy technologies.

The second reason the traditional models have limited application to RETs is that when the Bass model is introduced to forecast diffusion of RETs, the innovation parameter p is not interpreted properly. The most innovative buyers among categorized consumer groups[6] are affected by the innovation parameter p, and they become adopters at $T = 0$ (Bass, 1969). However, companies cannot be simply defined and classified according to the innovative tendencies like consumers. Companies do not adopt new technology to fulfill their satisfaction for innovation, but we can say that their innovation actions are taken to reach their business goals. Owing to the heterogeneity of individuals, they gain completely different utilities from the adoption of an innovation. In contrast, all companies follow the basic mechanism of profit maximization, even though their circumstances differ. In other words, for companies, profit maximization is more important than fulfilling innovation. Guidolin and Mortarino (2010), who deal with SPV technology diffusion in 11 countries using a generalized Bass model, estimate a very low value for parameter p, reflecting that innovation rarely affects the SPV energy market. Dalla Valle and Furlan (2011) and Purohit and Kandpal (2005) report similar results for the cumulative capacity of wind power in Europe and the United States and for irrigation water pumping using RETs in India, respectively. A very low value of parameter p limits the acceleration of diffusion compared to normal levels of p. In particular, when

[6] Rogers (2003) classifies consumers into five classes: innovators, early adopters, early majorities, late majorities, and laggards.

parameter p is not significant, it means that innovation does not affect RETs' diffusion. Many studies have shown an insignificant coefficient or a very low value for innovation (Table 2).

Table 2. Estimated parameter p in previous literature

Thesis	Data Period	Energy and Region	Coefficient of Innovation
Purohit and Kandpal (2005)	1998–2002	SPV power in India	0.000466 (N/A)
	1995–2002	Windmill power in India	0.000203 (N/A)
	1982–2002	Biogas driven dual fuel engine in India	0.000032 (N/A)
	1995–2002	Producer gas driven dual fuel engine in India	0.000014 (N/A)
Guidolin and Mortarino (2010)	1992–2006	SPV power in Japan	0.000123 (0.000355)
		SPV power in UK	0.003472 (0.010425)
		SPV power in Germany	0.000202 (0.000046)
		SPV power in Australia	0.000165 (0.000157)
		SPV power in Canada	0.000086 (0.000253)
		SPV power in France	0.000047 (0.0000098)
		SPV power in Austria	0.000049 (0.0000201)
		SPV power in Netherlands	0.000091 (0.010314)
		SPV power in US	0.000015 (0.000658)
		SPV power in Italy	0.001694 (0.019708)
		SPV power in Spain	0.000007 (0.000756)
Dalla Valle and Furlan (2011)	1990–2008	Wind power in Europe	0.0009 (0.0006)
	1981–2008	Wind power in US	0.00003 (0.00004)
	1987–2008	Wind power in Germany	0.0006 (0.0002)
	1991–2008	Wind power in Spain	0.0006 (0.0003)
	1980–2008	Wind power in Denmark	0.002 (0.001)
	1995–2008	Wind power in Italy	0.0006 (0.0002)

Note: Numbers in parentheses are estimated standard error

In addition, the diffusion patterns of RETs are estimated to evaluate how well classical diffusion models explain the patterns in the South Korean RET market data. Installed renewable energy plant capacity data in South Korea between 2003 and 2012 are provided by the Electric Power Statistics Information System (EPSIS)[7], and Equations (2) and (3) are used to make these estimates for the Bass and logistic models, respectively. Table 3 summarizes the estimation results, and Figure 3 shows the pattern of diffusion until 2033 based on the estimation results.

Table 3. Estimation results of the Bass diffusion model and the logistic model

	Parameters	Estimate	Std. Error	t value	Pr(>|t|)
Bass Model	*m*	2.248e+06**	4.569e+05	4.920	0.00171
	p	1.171e-03	1.724e-03	0.680	0.51864
	q	8.787e-01**	2.322e-01	3.785	0.00685
Logistic Model	*m*	2.416e+06**	1.569e+05	15.398	1.18e-06
	a	-5.389e+00**	4.887e-01	-11.026	1.12e-05
	b	7.622e-01**	8.742e-02	8.719	5.24e-05

Note: ** indicates statistical significance at the 5% level

[7] Hydroelectric technologies are excluded in this study among RETs because hydroelectric generation has reached saturation under geographical limitations and will account for only 1% of planned capacity volume. In addition, hydroelectric technologies accounted for the bulk of RETs in the 2000s, therefore, causing problems for parameter estimation.

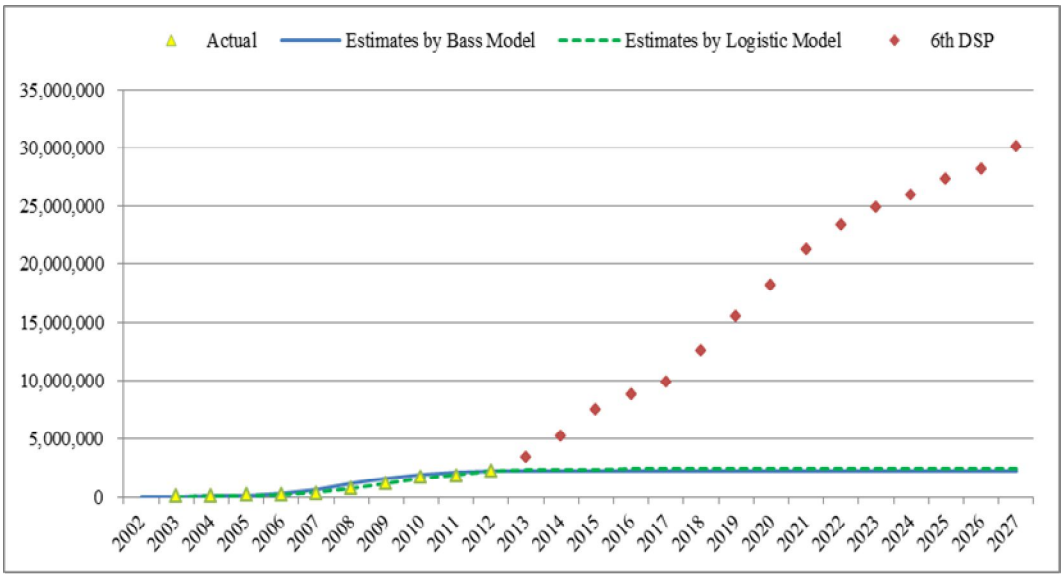

Figure 3. Installed capacity of RET plants using the Bass model and the logistic model (unit: kW)

All the parameters are significant excluding the external influence parameter p in the Bass diffusion model (Table 3), and can be interpreted as having no external effect. The market potential parameters of the two models are significant at the 1% level; however, their forecasting does not take into account politically planned RET capacities, which South Korea's 6th Demand Supply Program (DSP hereafter) stipulates (Figure 3). The Bass diffusion model and the logistic model estimate that total installed capacity would be saturated at 2,246,349 kW and 2,415,849 kW, respectively, in 2014 whereas the Korean government aims to install 30,148,000 kW until 2027. This huge difference occurs because the models fail to reflect the political aspects. Environmental regulations such as a mandatory policy that enforces to introduce environmental friendly energy

sources are designed to guide the trajectories of technological change and to guard against negative externalities, which occur as companies freely consume air and water. These externalities are a central difference between environmental technologies and other technologies. Therefore, environmental regulations and policies play an important role in technology diffusion and development (Kemp and Soete, 1990). Söderholm and Klaassen (2007) argue that induced diffusion is quite different from conventional diffusion under policy intervention, such as, feed-in tariffs.

Thus, a new model should be developed to explain the RET diffusion. Since RET diffusion is induced by policies and regulations, the diffusion pattern cannot be suitably forecasted from the classical perspective using communication factors. The change agents or facilitators will be the key to explaining an induced diffusion. In *Chapter 3*, a new diffusion model will be proposed based on the role of policy as a change agent or a facilitator, and an empirical study will be conducted.

2.2 Choice Models with Individual-level and Aggregate-level Data

Choice models are widely used for demand forecasting based on the analysis of consumer behavior and preference. In choice models, it is assumed that consumers decide to maximize their utility; these are called random utility models. To establish new demand forecasting methodology with heterogeneous consumer preferences, the discrete choice

model is used. In the following sub-chapters, various discrete choice models are introduced, and how previous DCE studies incorporate the individual-level data is described. The random coefficient logit model, which is the basis of the BLP model (Berry et al., 1995, hereafter BLP), is then discussed in detail. Finally, the research motivation for the choice model with aggregate-level data is stated.

2.2.1 **Logit Model and its Extensions**

A decision maker n chooses an alternative from among J alternatives, and the decision maker has a certain level of utility U_{nj}, j=1,…J from choosing alternative j. He or she would choose the alternative that maximizes his or her utility. The decision maker chooses alternative j rather than the other alternatives if and only if $U_{ni} > U_{nj} \quad \forall j \neq i$. The maximized utility consists of a deterministic part V_{nj} and a stochastic part ε_{nj}.

$$U_{nj} = V_{nj} + \varepsilon_{nj} . \qquad \text{Eq. (13)}$$

The deterministic part can be observed by researchers, and it can be denoted as a function V_{nj}. The deterministic part, which is also called the representative utility, is normally specified as a linear function—that is, $V_{nj} = \beta x_{nj}$, where x_{nj} indicates the observable attributes of alternative j and β indicates the coefficients to be estimated.

The stochastic part ε_{nj} can be observed by decision makers but not by researchers. Instead, researchers handle this part as the joint density of the random vector, which is denoted by $f(\varepsilon_n)$ (Train, 2003). Therefore, the choice probability P_{ni} that decision maker n chooses alternative i from among alternatives J can be written as

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i) \\ &= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall j \neq i) \\ &= \text{Prob}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \quad \forall j \neq i). \end{aligned} \quad \cdots\cdots \text{Eq. (14)}$$

The logit model can be achieved by assuming that the stochastic part is the independently and identically distributed (iid hereafter) extreme value, which is known as the type-I extreme value. Following McFadden (1974), the choice probability of the logit model is

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} = \frac{e^{\beta x_{ni}}}{\sum_j e^{\beta x_{nj}}}. \quad \cdots\cdots \text{Eq. (15)}$$

When there are only two alternatives (to adopt or not to adopt), the model is called a binary logit model. On the other hand, when there are three or more alternatives, it is called a multinomial logit model. The logit model is the easiest and most widely used of the discrete choice models because it uses the closed form of choice probability. However,

the logit model exhibits *independence from irrelevant alternatives* (IIA hereafter). This property comes from the iid assumption, and it leads to an unrealistic substitution pattern since the ratio of choice probabilities is independent of the other alternatives, as discussed below.

$$\frac{P_{ni}}{P_{nk}} = \frac{e^{V_{ni}} / \sum_j e^{V_{nj}}}{e^{V_{nk}} / \sum_j e^{V_{nj}}} = \frac{e^{V_{ni}}}{e^{V_{nk}}} = e^{V_{ni} - V_{nk}} . \quad \text{Eq. (16)}$$

To avoid the IIA property in the logit model, the Generalized Extreme Value (GEV hereafter) model was developed. The GEV model allows various substitution patterns to introduce correlations over alternatives. The best-known GEV model is the nested logit model, which is adequate when alternatives can be divided into subsets. If two alternatives are in the same nest, the ratio of choice probabilities is independent of the other alternatives, and the IIA property still exists within the nest. If two alternatives are in different nests, the ratio of choice probabilities depends on the other alternatives in the two nests. In this case, the IIA property does not hold in different nests (Train, 2003).

The random coefficient model, which is also called the mixed logit model, releases the restrictions with regard to the substitution pattern. Moreover, the random coefficient model allows random taste variation and correlation in unobserved factors (McFadden and Train, 2000). It differs from the standard logit model in that the random coefficient model incorporates an attribute's coefficient as a distribution, which means that the

marginal utility varies over individuals.

$$U_{nj} = \beta_n x_{nj} + \varepsilon_{nj}. \quad \cdots\cdots \text{Eq. (17)}$$

As in logit model, the stochastic terms are iid extreme values, and by assuming the distribution of β_n, the choice probability is derived through the integration of the density function of β_n, $f(\beta)$.

$$P_{ni} = \int \left(\frac{e^{\beta xi}}{\sum_j e^{\beta x_{nj}}} \right) f(\beta) d\beta. \quad \cdots\cdots \text{Eq. (18)}$$

The random coefficient model avoids the restrictive substitution pattern of the logit model. The ratio of choice probability is not independent of other alternatives; the percentage change in alternative i's probability by changing the mth attribute of alternative k is given below (Train, 2003).

$$\begin{aligned} \eta_{nix_{nk}^m} &= -\frac{x_{nk}^m}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nk}(\beta) f(\beta) d\beta \\ &= -x_{nk}^m \int \beta^m L_{ni}(\beta) \left[\frac{L_{nk}(\beta)}{P_{ni}} \right] f(\beta) d\beta \end{aligned}, \quad \cdots\cdots \text{Eq. (19)}$$

where $L_{ni}(\beta) = \frac{e^{\beta x_{ni}}}{\sum_j e^{\beta x_{nj}}}$ and $L_{nk}(\beta) = \frac{e^{\beta x_{nk}}}{\sum_j e^{\beta x_{nj}}}$, and β^m is the m th element of β.

The logit and the random coefficient logit models assume that the stochastic part follows the type-I extreme value distribution; on the other hand, the stochastic part of the utility can be assumed to be a normal distribution with a mean vector of zero and covariance matrix Ω; this is known as a probit model. Under this assumption, the choice probability is expressed as

$$P_{ni} = \int I(V_{ni} + e_{ni} > V_{nj} + e_{nj} \forall j \neq i)\phi(\varepsilon_n)\, d\varepsilon_n, \quad \text{Eq. (20)}$$

where $\phi(\varepsilon_n)$ denotes the density function of ε_n and is an indicator of whether the statement in parentheses holds (Train, 2003). The assumption that the stochastic part is normally distributed is proper for many general cases. However, it is sometimes unrealistic—for example, people always have a negative coefficient for price. In addition, the choice probability is not a closed form, and it should be evaluated numerically by simulation.

Although researchers choose the logit model or the probit model according to their intuition, both models have been used in many studies. In the next chapter, various applications in combination with conjoint analysis are discussed.

2.2.2 Discrete Choice Experiments with Individual-level Data

As introduced in *Chapter 1.1*, the DCE is a forecasting method that combines conjoint analysis and the choice model; the choice model was briefly introduced in the previous chapter. Conjoint analysis was developed in mathematical psychology: Luce and Tukey (1964) triggered the wide use of conjoint analysis by providing a method to measure stated-preference data. The conjoint method shows the consumer preference for each product attribute based on the product rankings that consumers select (Green and Srinivasan, 1978). This property is useful in the marketing field, and conjoint analysis has been used extensively since Green and Rao (1969) introduced it into the marketing field. Another strength of conjoint analysis is its high reliability. Anderson et al. (1992) compare the reliability of consumer preference surveys and find conjoint analysis to be the most reliable survey method.

The DCE, which is also known as choice-based conjoint analysis, was introduced into the marketing field by Louviere and Woodworth (1983). There are not substantial differences between the conventional conjoint analysis and DCE for the choice task; however, the DCE can provide more realistic responses because it uses choice data rather than rating data. Another difference is that the DCE suggests several choice sets by dividing the total choice sets, and respondents then choose the most preferred alternative in each choice set; in contrast, conventional approaches present all the choice sets at once

(Haaijer and Wedel, 2001).

Since it has many advantages of conjoint analysis and the choice model, the DCE has become one of the most popular methodologies for consumer preference study. The marketing literature includes a variety of DCE applications. The previous studies that utilize the DCE are summarized in Table 4.

Table 4. Previous studies using DCE

Study	Applied model	Remark
Dellaert et al. (1997)	Logit / Nested logit / Probit	The logit, probit, and nested logit models were applied to investigate Dutch tourists' preferences; the nested logit model was found to be most suitable.
Dow and Endersby (2004)	Multinomial logit / Multinomial probit	Multinomial probit and logit models were compared for voter choice in multi-party elections; the simpler logit model was found to be often preferable to the more complex probit model.
Jeong and Lee (2010)	Mixed logit	In the online music service, methods to prevent illegal file sharing were suggested based on consumer preference.
Lee and Cho (2009)	Rank-ordered logit	Consumer demand was predicted for diesel cars; a 42% market share was predicted in South Korea under the government's proposed pricing system.
Lee et al. (2006)	Rank-ordered logit	The demand for large-screen televisions was predicted by combining the dynamic utility function and the Bass diffusion model. Televisions with 40-inch screens were preferred to televisions with larger screens.
Potoglou and Kanaroglou (2007)	Nested logit	Consumers' extra willingness to pay for a clean vehicle was examined in Hamilton, Canada.

The most significant improvement achieved by combining the conjoint and choice models is that the individual-level stated preference can be used to identify the heterogeneous consumer preference within the choice model. Thus, the DCE allows researchers to analyze the preference structure without revealed data.

However, the DCE has several limitations. First, the DCE might have a hypothetical bias problem in the stated-preference. This is an essential problem with survey data; however, it can be alleviated by including revealed-preference data or by using a practical method to reduce the bias, such as cheap talk. Second, consumers' answers are not accurate when the subject is unfamiliar or the choice sets include many alternatives and attributes. Specifically, when consumers are not interested in the survey, they tend to choose the no-choice option, which brings a bias. Third, the biggest problem is that the DCE is always requires stated-preference data. Obtaining this data is a burden for companies and organizations because it takes time and financial support. An ideal solution is using revealed-preference data instead of stated-preference data; however, it is very difficult to obtain revealed-preference data at the individual level. Therefore, aggregate revealed market data can be used if it is possible to analyze consumer preference at the same level as when using individual stated-preference data.

2.2.3 **Motivation**

Although the choice model can be used with aggregate market data, still there is an

ignored endogeneity problem when aggregate data are used with the choice model. BLP 1995) attempted to solve this problem in terms of industrial organization. BLP (1995) discuss the endogeneity problem between the price and unobserved factors and suggest using instrument variables for proper estimation. Furthermore, they propose a method to estimate consumer preference with only market share data and consumer characteristic distribution information. The proposed method provides estimators that are as reliable those using actual consumer socio-demographic information. This approach is very useful because researchers and marketers often have access to aggregate market data.

Three advantages are obtained when heterogeneous consumer preference is analyzed with only aggregate-level data.

First, the forecasting method using aggregate market data overcomes the problems from stated-preference data. A hypothetical bias is necessarily produced when the survey is conducted (Zwerina, 1997). Because the answers from survey respondents are obtained for a virtual situation, the stated preference cannot be the same as the revealed preference. This difference is an unavoidable problem of using survey data. Moreover, insincere responses can be given by respondents who are not interested in the research subject (Haaijer et al., 2001). These responses are useless; however, researchers cannot identify these responses among the answers. The aggregate market data are revealed data, so no hypothetical bias and no meaningless responses exist in the data. Therefore, researchers and marketers can obtain more accurate heterogeneous consumer preference when using aggregate market data than with DCE.

Second, researchers and marketers do not need to perform a survey to conduct a market analysis. Surveys cost money and time. In contrast, aggregate market data are obtainable from marketers, and other consumer information and environmental conditions can be collected from the national database system. It is easier to collect aggregate market data in terms of cost and time than to conduct a survey.

Lastly, if a heterogeneous consumer preference is identified, it is possible to predict a market share for a given market condition only with consumer information and environmental market conditions. Generally, consumers' socio-demographic information is used to explain the heterogeneity. When a company develops a new product, marketers can predict the demand with the identified preference and consumers' socio-demographic information and can test the substitution pattern under the market situation. In addition, the company can determine how favorable a new market is by introducing the new market's information. This predictive ability is very useful for marketers, and can be used to develop marketing strategies for various situations.

Therefore, it is worth considering demand forecasting methodology using aggregate market data. Although BLP (1995) already provided the mathematical specifications, complicated modeling characteristics prevent use in the marketing field. In *Chapter 4*, the demand forecasting methodology based on BLP (1995) is described in detail, and an application to the U.S. automobile market is provided. Additionally, how the results are used by marketers is discussed.

Chapter 3. Diffusion Forecasting for Renewable Energy Technology in South Korea

In this essay, a new diffusion model will be devised by focusing on the political effects in renewable energy technologies (RETs). Classical diffusion models have strengthened diffusion forecasting in terms of innovation and imitation. Nevertheless, it is important to note that classical models are limited in their application to innovations that are not self-sustainable, such as RETs. Exogenous government interventions clearly exist and seriously influence diffusion patterns through policies. The proposed adjusted logistic model overcomes this limitation and supplements conservative diffusion models. In addition, the estimated diffusion pattern provides an appropriate direction for subsidy plan establishment from the perspective of policy makers. The research framework of this essay is shown in Figure 4.

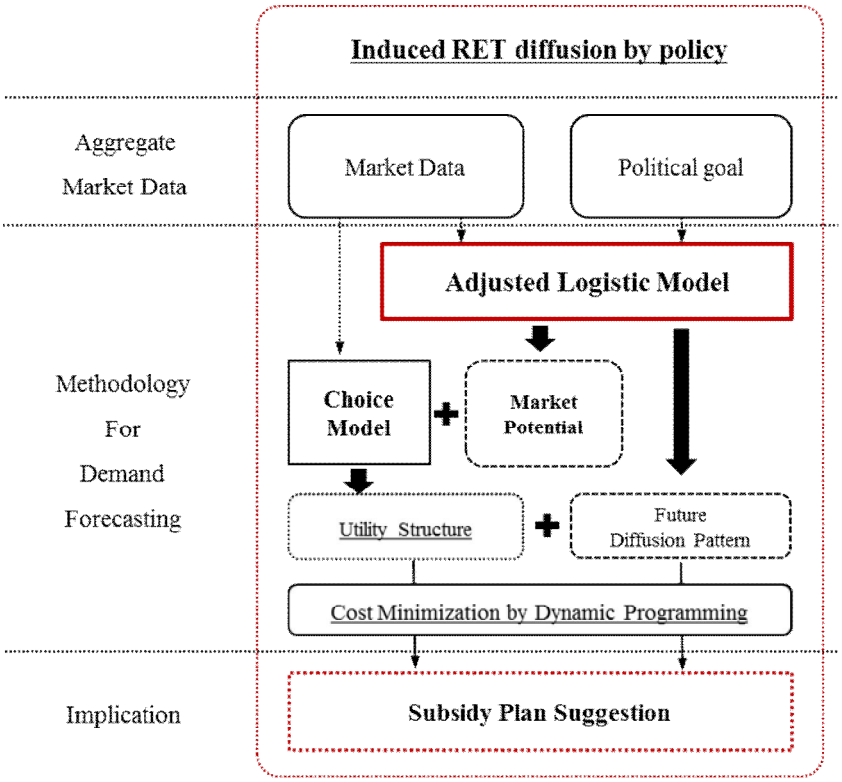

Figure 4. Research framework for essay 1

3.1 Introduction

3.1.1 Renewable Energy Technology (RET) and its Barriers

In South Korea, renewable energy is defined by the Act on the Promotion of the Development, Use and Diffusion of New and Renewable Energy as energy that uses existing fossil fuels after conversion or uses energy from sunlight, water, precipitation,

biological organisms, or geothermal sources by conversion. It also designates specific eleven energy types: solar, bio, wind, hydro, fuel cells, coal liquefaction and gasification, heavy residue gasification, marine, waste, geothermal, and hydrogen. There are minor differences in the definitions of renewable energy between countries—for instance, the United States government defines renewable energy more comprehensively. According to Armstrong and Hamrin (2000), the term "renewable" means "non-depletable" or "naturally replenishable," so renewable resources can include solar energy, wind, falling water, the heat of the earth (geothermal), plant materials (biomass), waves, ocean currents, temperature differences in the oceans, and tidal energy. Differences in the definition of renewable energy between countries are related to the scope of political support for RETs. Armstrong and Hamrin (2000) state in the *Renewable Energy Policy Manual* that commercial technologies and other technologies in the research or early commercial stage, as well as non-electric renewable energy technologies, are outside the scope of the manual because policy makers concerned with the development of the national grid system will focus on resources that have established themselves commercially and are cost effective for on-grid applications. Therefore, it can be said that the definition and scope of RETs are very closely related to political issues, which also indicates that RETs require policy support to overcome certain barriers.

Various studies have focused on barriers to the diffusion of RETs. Dulal et al. (2013) categorize the barriers in the Asian market into five groups: Market distortions, Economics, Policy and institutional, Technical, and Socio-cultural/Political (Figure 5).

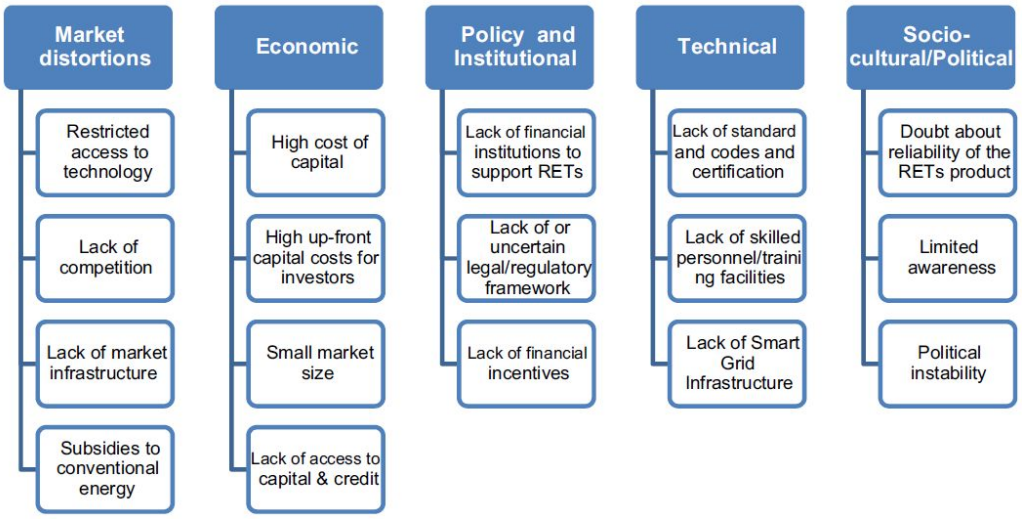

Figure 5. Barriers to RET diffusion in Asia

Source: Dulal et al. (2013)

High dependency on fossil fuel in Asian countries causes market distortions, and the relatively low economics of RETs hinder investment. Natural large-scale RET diffusion is impossible without government support because of high costs and small market sizes. Most Asian countries lack the necessary policy and institutional supports, and technical supports such as standards and training facilities are insufficient to induce RET development. Moreover, consumers are not informed about RET products to overcome this limitation, five strategies are provided as follows:

1. Fiscal incentives for RETs

2. Institutional strengthening

3. Institutional investment

4. Renewable energy certificates (REC)

5. Cross-subsidization

Dulal et al. (2013) suggest that these five strategies should be applied differently according to local conditions and environments, and they emphasize the importance of energy policy through various cases with the above five strategies.

Another strand of research addresses the problem of slow RET diffusion in terms of the innovation system. In this point of view, the diffusion of innovation depends highly on the environment in which innovations are developed. Negro et al. (2012) review 50 case studies and categorize systemic problems. They focus on the characteristics of the incumbent system and the emerging RET system and attempt to determine the relationship between the two. Their analysis identifies seven problems[8], some of which are deeply connected with policies. For instance, hard institutional problems, the most common type of problem, include inconsistent policy and regulation. Announced subsidies periods are frequently delayed or reduced, and the implemented tariffs are sometimes lower than planned. This type of uncertain policy implementation leads to a lack of trust in the government. Furthermore, policy makers' attention shifts to

[8] The seven systemic problems are hard institutional problems, market structures, soft institutional problems, capabilities, knowledge infrastructure, too-weak and too-strong interaction problems, and physical infrastructure.

technologies, and misalignment between policies increases this mistrust. In terms of capability problems, policy makers' lack of technological knowledge leads to the introduction of the wrong technologies. Even in the innovation system approach, it is shown that incumbent dominance obstructs RET diffusion, so market failure is also a systemic problem. Although the applied methodologies for RETs barriers are different, Negro et al. (2012) also suggest policy recommendations to overcome hard institutional failure. Using the innovation system approach, Tsoutsos and Stamboulis (2005) also understand the unsustainable trajectories of RET diffusion in terms of systemic innovation processes. They argue that RET policy should focus on the integration of the innovation dimension for renewables to facilitate self-sustainability, and they identify seven types of barriers based on Kemp et al. (1998), who discuss regime shifts through processes of niche formation of policies. They identify seven barriers to technological regime change, which are shown below.

1. Technological factors
2. Government policy and regulatory framework
3. Cultural and psychological factors
4. Demand factors
5. Production factors
6. Infrastructure and maintenance
7. Undesirable social and environmental effects of new technologies

Tsoutsos and Stamboulis (2005) propose that to overcome the barriers, user innovation-oriented policy should be implemented because it accelerates the diffusion process in three ways. First, concrete and specific attention encourages the development of learning mechanisms across the value chain from developers to consumers. Second, new players participate in the framework as structural elements that mobilize public opinion and resources. Last, financing supports boost the mechanism to balance competition between the RETs and incumbents. These three policy directions would provide competitiveness for the RETs.

Previous studies observe that RETs cannot be diffused and developed by themselves, as other technologies can, and suggest various methods to overcome the barriers; all these solutions are based on policy.

3.1.2 RET Diffusion Studies Considering Policy

Although the classical diffusion model does not provide a proper interpretation for South Korean RETs, there are various studies on RET diffusion based on the classical models.

Based on the Bass model, Rao and Kishore (2009) analyze wind power technology diffusion in selected states in India. While the Indian central government has adopted a leadership position on wind power technology diffusion, individual states have different

policies to encourage diffusion. Their Bass model estimation results show the policy impacts by compositing policy indexes such as preferential tariffs, wheeling charges, banking, the availability of adequate transmission facilities, additional state incentives, third party sales, and land availability. The composite policy index shows a strong correlation with the point of inflection in the diffusion. In addition, even with a higher value of p, diffusion is sustainable only for short periods unless the imitation factor is high enough to support it because the subsequent rate of diffusion mainly depends on the imitation factor q.

Davies and Diaz-Rainey (2011) explore the possibility that political intervention and incentives can stimulate the diffusion of RETs. They apply the Bass model to the induced diffusion model and assume that the political effect is equal to the external effect, which also represents an innovation parameter p. They test four propositions to determine whether diffusion patterns of wind energy are different under political support in 25 OECD countries. Conventional logistic diffusion processes are observed for 20 of 25 countries: the diffusion curves of these 20 countries are symmetric S-shaped and have slow diffusion speeds. The other five countries, which accept the induced model, have strong policy interventions such as feed-in tariff systems.

A similar approach can be found in Giaccaria and Dalmazzone (2012), who define the external factor as a function of political impacts to confirm that political interventions such as the decentralization of energy policy planning and authorization procedures influence RET diffusion. They test Italian wind and photovoltaic energy technologies

from 1999 to 2010, and the results show specifically which intervention instruments had an influence in the early stage and how large those impacts were. Their study also shows that subsidies on capital costs have a significant effect on RET diffusion. They conclude that the incentive scheme is the result of a complicated process of institutional learning and interactions among policy instruments.

Several studies use the logistic model. Chen et al. (2012) employ the logistic model to measure policy impacts in Taiwan. According to this study, mixed policies that combine environmental standards and economic instruments have positive impacts on RET diffusion, whereas negative impacts occur when only environmental standards are enforced. This is because newly specified environmental regulations focus on existing technologies, and clean technology development has already been completed in the industry (Mickwitz et al., 2008).

Purohit and Kandpal (2005) test various diffusion models for four RETs for Indian irrigation: water pumping systems, solar photovoltaic pumps, windmill pumps, and biogas/producer gas-driven dual fuel engine pumps. The logistic model shows the most positive diffusion pattern, while the Bass and Pearl models show comparatively moderate diffusion patterns. The slowest and least effective diffusion is shown by the Gompertz model. All four models report that RETs for irrigation water pumping systems cannot reach the market potential in the next 25 years. By extension, the authors assume learning effects based on previous studies and present the corresponding capital investment in RET irrigation water pumping systems.

As shown earlier, many studies note the importance of policy roles. A summary of the previous literature is provided in Table 5.

Table 5. Summary of previous literature using diffusion model

Study	Applied model	Remark
Chen et al. (2012)	Logistic model	Emission standards and economic instruments in Taiwan are tested to determine whether they influence clean technology diffusion using the logistic model; these two interventions have positive effects on technology diffusion.
Davies and Diaz-Rainey (2011)	Induced model based on Bass model	Twenty-five OECD countries have conservative logistic diffusion patterns for wind energy without political support; the Bass curve adequately explains the patterns of diffusion under strong policy intervention.
Giaccaria and Dalmazzone (2012)	Logistic model and Bass model	Energy policy design influences RET diffusion patterns for Italian wind and photovoltaic energy. The incentive package is explained as a combination of the institutional learning process and interaction among policy instruments.
Purohit and Kandpal (2005)	Bass model, Gompertz model, logistic model, Pearl model	Various diffusion models are used to examine the diffusion patterns of four RETs for irrigation water pumping in India. The projected dissemination levels of diffusion for the four RETs are provided.
Rao and Kishore (2009)	Bass model	Wind power technology in four Indian states is used for diffusion analysis based on the Bass model. The diffusion patterns of the four states differ according to policy composition, and the addition to external value can bring a higher penetration level even though the internal effect prevails in the diffusion.

Rao and Kishore (2010) provided an overview of RET diffusion analysis using diffusion models. They emphasize that RETs are different from commercial and consumer products, so classical diffusion models thus have limitations to RET application. They conclude that the linkage to government policies should be considered to pursue the parameters or coefficients of the RET diffusion model.

3.2 Adjusted Logistic Model Considering Policy Goals

3.2.1 Model Specification

This chapter devises an adjusted logistic model—rather than a Bass model—in order to avoid the limited interpretation of innovation factors, as mentioned in *Chapter 2.1.3*. The estimation result, whose parameter p of the classical Bass diffusion model is not significant, supports the introduction of the logistic model. The Bass diffusion model, although it is widely used, is not always suitable for forecasting diffusion patterns (Kamakura and Balasubramanian, 1988). The Bass model assumes the market to be static; however, the market potential is changeable in the case of RET plant capacity diffusion. According to the 6th DSP, electricity consumption increased 5.6% annually between 2002 and 2011, and the total plant capacity increased by 52% from 2002 to 2011 following the consumption increase. Given this trend, the actual RET plant capacity will

increase over time. In addition, planned updates to RET plant capacity support the increased market potential. Figure 6 shows the planned changes in RET capacity over time; the planned capacity has also been increased.

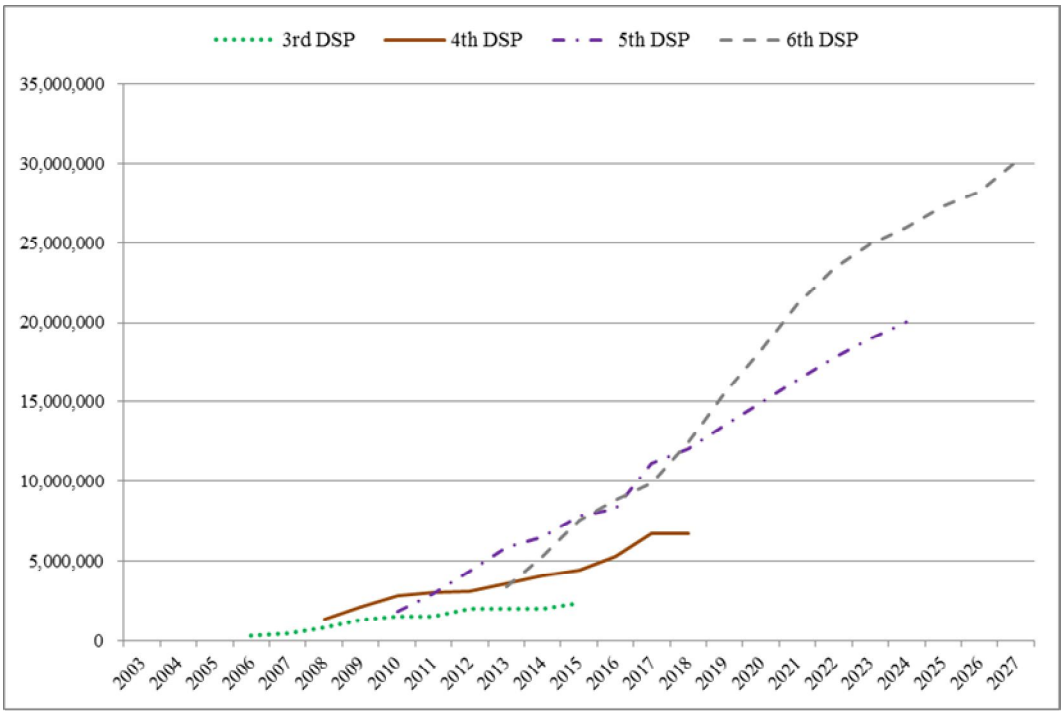

Figure 6. Planned changes in RET plant capacity by policy plans (unit: kW)

Moreover, the Bass model is not able to explain multi-adoption (Norton and Bass, 1987). While the Bass model properly explains the first adoption, companies frequently adopt multiple units of RET plants or a large amount of RET plant capacity at once. Therefore, forecasting RET plant capacity with the Bass model is not suitable.

It is known that the logistic model has been used to explain the diffusion of industrial

innovations in technological substitution studies (Mahajan and Muller, 1979). The introduction of RETs by companies is not the beginning of rapid diffusion[9] but is an a priori test, because companies sometimes release conceptualized models as test beds for potential profit, and these are quite different from those used for commercialization. For instance, the preemption of patents within markets, and the showcasing of technology and an environmentally friendly image are not designed to reap short-term profits. This kind of behavior is essentially for future profits rather than for the sake of innovation. In contrast, parameter q, the imitation factor, is significant. Sultan et al. (1990) conduct a meta-analysis of 213 applications and show that the imitation coefficients of industrial and medical innovation are much higher than other imitation coefficients for innovation. Moreover, some subjects can be explained by an internal effect only, such as, administrative technology (Mansfield, 1961; Teece, 1980).

However, the classical logistic model does not properly reflect political inducement. Political institutions and policy design have substantial influence on industry development (Conrad, 1996; Feess and Muehlheusser, 2002) but there is a mismatch in the relationship between political goals and actual installed capacity for RET power plants. Specifically, actual installed solar, ocean, and geothermal energy plant capacities follow political goals with some shortfall (Figure 7). RET diffusion does not happen spontaneously but rather follows political intentions. For RETs with relatively weak

[9] Figure 3 shows that both the Bass diffusion model and the logistic model saturate near the last observation point, which means that RET diffusion is not rapid.

competiveness, there needs to be clear political intentions to increase diffusion, such as supporting programs and subsidies. Efforts up to now have achieved only some parts of the political goals of the government to disseminate RETs while at the same time some shortfalls have occurred.

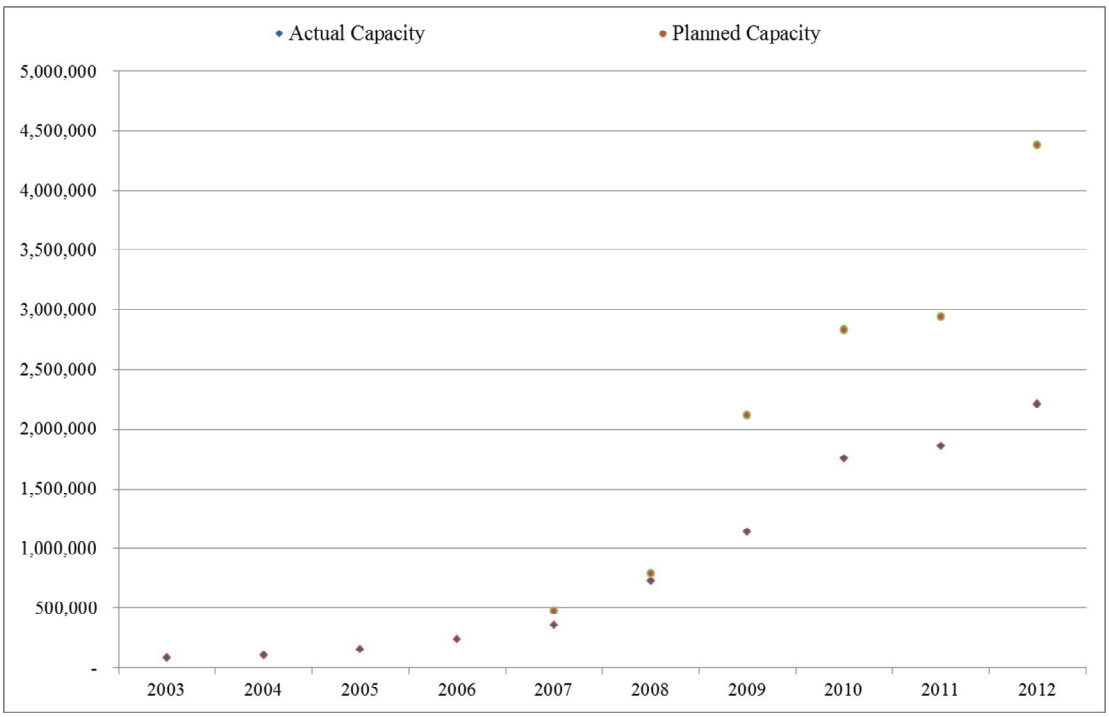

Figure 7. Actual and planned RET plant capacity in South Korea (unit: kW)

Source: EPSIS and the 3rd through 5th DSPs

As mentioned, the gap between planned capacity and actual capacity should be considered alongside political influences. This study assumes that the planned capacity and actual capacity follow S-shaped curves and that the gap between planned and actual

capacity gradually decreases through the learning effect. The learning effect has been widely applied to technology improvement and price reduction (Feess and Muehlheusser, 2002; McDonald and Schrattenholzer, 2001; Nemet, 2006; Van Benthem et al., 2008). An adjusted logistic model is proposed with the assumption that the shortfall in politically determined capacity decreases over time because of the learning effect. Therefore, the suggested model is defined as

$$N(t) = m\left[\frac{1}{1+e^{-(a+(b-(I(d)\times e^{-kt})\times t))}}\right], \quad \text{Eq. (21)}$$

where a, b, k, and m are parameters and $N(t)$ refers to the cumulative installed capacities (kW), which comprise both the actual installed RET plant capacity and the politically planned RET plant capacity. The meanings of the following three parameters are the same as those used in the classical logistic diffusion model: m is market potential, b is the diffusion rate, and a is a constant. k represents the learning effect that explains the decreased gap between planned and actual capacity over time. $I(d)$ identifies two kinds of capacities as an indicator function. When the capacity is the actual installed RET plant capacity, $I(d)$ is one; otherwise, $I(d)$ is zero, which means that the capacity is the politically planned RET plant capacity. As the politically planned RET plant capacity data are used with Eq. (21), the pattern of politically planned RET

diffusion is estimated with the original logistic model. On the other hand, when the actual installed RET plant capacity is used ($I(d)=1$), the diffusion curve of the actual installed RET plant capacity has a different imitation parameter, $b-e^{-kt}$. The actual RET plant capacity is the result of efforts to reach the politically planned RET plant capacity, and the capacity gap in reaching the goal is the shortfall rate. Therefore, the realized diffusion pattern of politically planned RET adoption, which is equal to the actual RET adoption, is estimated by considering the learning effect, e^{-kt}. In our proposed model, however, political RET plant capacity is introduced through an indicator function, ($I(d)$), so that political intervention to overcome RET's competitive disadvantages can be considered.

In terms of the politically planned RET plant capacity data, multiple policy plans can be introduced since the applicable planned capacity data changes over time. The DSP is biennially updated, so different planned capacity data are used every two years. Figure 8 shows an example when the forecasting is performed for 2012.

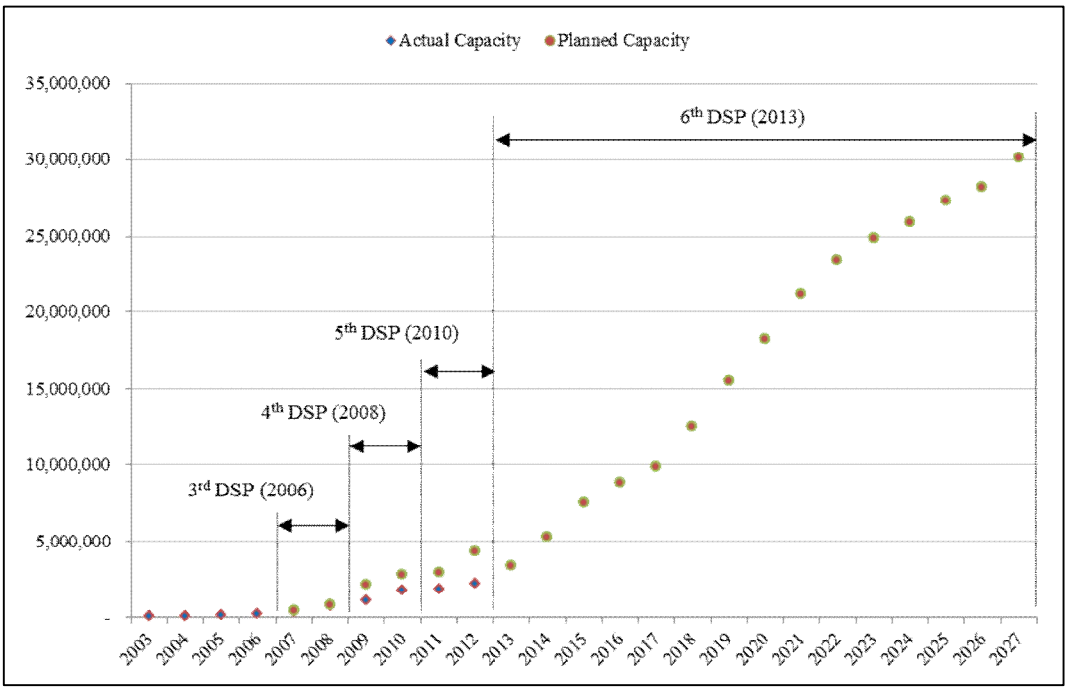

Figure 8. Capacity data in 2012 and estimation results (unit: kW)

3.2.2 Comparison of Forecasting Accuracy

In this chapter, the forecasting accuracies of the proposed model, the classical Bass model, and the logistic model are assessed through an out-of-sample test. Two tests, fixed origin evaluation and rolling origin evaluation, are used to evaluate the forecasting accuracies. Both tests divide the data series into a fit period and a test period, and set the fit period to generate forecasts for the test period (Tashman, 2000). The final time in the fit period is called the origin. Fixed origin evaluation generates as many forecasts as the

test period. Forecasting errors can be calculated by comparing forecasts to the actual historical data. In rolling-origin evaluation, the origin T is updated every unit period to generate forecasts for the remaining periods of a test period. Standing at origin T, N forecasts are generated. At the next period, another $N-1$ forecasts are generated based on origin $T+1$. This step is repeated until N becomes one.

In this study, fixed origin evaluation is applied to minimize the time updating that leads to the introduction of biennially updated capacity data. Updating different planned data sets causes large variation of the estimation results, especially for market potential. The time is fixed at the end of 2009 with 3 reserve years from 2010 to 2012. In this case, actual installed capacity data are available from 2003 to 2009, and the planned capacity data are obtained only from the 3rd and 4th DSPs because the 5th DSP was announced in December 2010. The estimated results of the three models and forecasts of each model are shown in Table 6 and Table 7, respectively.

Table 6. Estimation results of the three models

Parameters	Bass Model	Logistic Model	Proposed Model
m (10^3)	2,594.10**	6,582.31	9,969.66**
	(1,222.85)	(4,093.87)	(2,680.99)
p	-.0194		
	(.03434)		
q	.9752**		
	(.3356)		
a		-5.4990**	-3.3791**
		(.3822)	(.2520)
b		.5607**	.25784**
		(.0640)	(.5436)
k			.3490**
			(.07952)

Note: ** indicates statistical significance at the 5% and numbers in parentheses are estimated standard error

Table 7. Calculated forecasts of the three models (unit: kW)

Year	Actual Observation	Bass Model	Logistic Model	Proposed Model
2003	83,315	83,315	46,840	112,542
2004	107,905	113,147	81,623	206,001
2005	155,622	170,447	141,677	250,472
2006	240,422	278,630	244,254	341,849
2007	350,609	476,149	416,299	489,727
2008	728,443	814,081	696,217	708,011
2009	1,136,352	1,324,217	1,129,860	1,010,281
2010	*1,748,980*	*1,931,678*	*1,753,154*	*1,406,072*
2011	*1,858,660*	*2,399,819*	*2,558,920*	*1,897,870*
2012	*2,210,773*	*2,571,304*	*3,468,994*	*2,479,279*

Note: The calculated forecasts are marked in italics

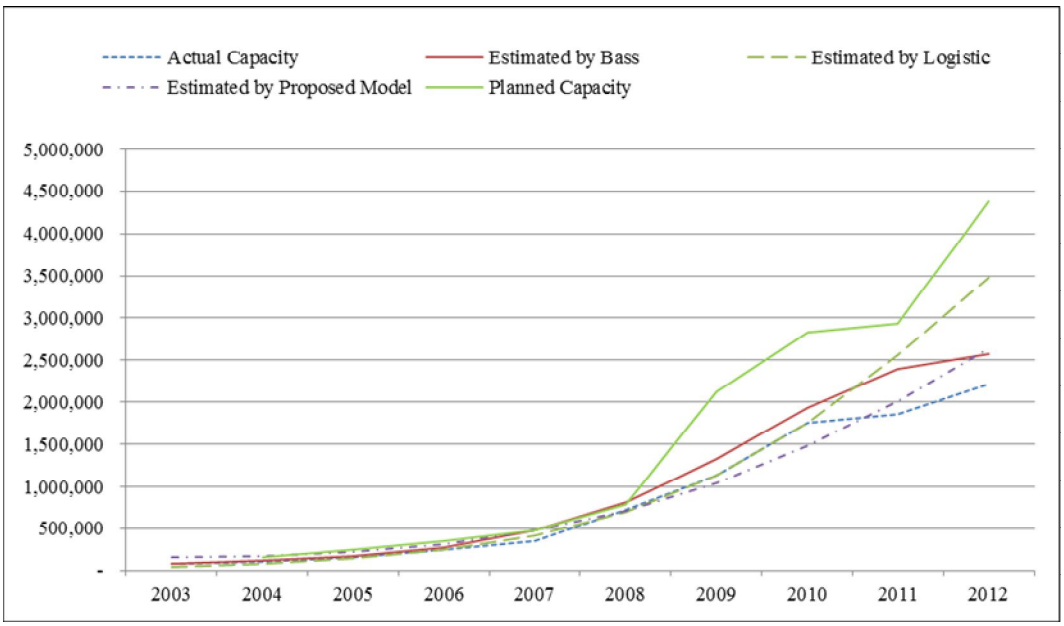

Figure 9. Diffusion patterns of the three models (unit: kW)

Based on the estimation results, three forecasts are calculated for 2010, 2011, and 2012. To compare the accuracy of the forecasts, root mean square error (RMSE)[10] and mean absolute percentage error (MAPE)[11] are used, and the results are shown in Table 8. In terms of both RMSE and MAPE, the proposed model shows better prediction than the two classical models.

Table 8. Comparison of forecasting accuracies between the three models

RMSE			MAPE		
Bass Model	Logistic Model	Proposed Model	Bass Model	Logistic Model	Proposed Model
389,964	831,365	252,467	18.62	31.61	11.29

The proposed model has saturated nearly 9,670,000 kW, which is higher than the politically planned capacity of 6,719,500 kW introduced in 2022. The estimated market potential of the proposed model is highly dependent on the planned capacity data. To be precise, planned capacity is estimated with the logistic model, and actual capacity is

---

[10] $RMSE = \sqrt{\frac{\sum_{t=1}^{N}\left(\hat{F}_t - F_t\right)^2}{N}}$

[11] $MAPE = 100 \times \frac{\sum_{t=1}^{N}\left|\frac{\hat{F}_t - F_t}{F_t}\right|}{N}$

estimated with an adjusted logistic model that simultaneously considers the learning effect. Planned capacity is treated as a normal observation through the logistic model, which makes the planned capacity always lower than the market potential. This is acceptable since governments set up realistic planned capacity targets and usually abandon the enforcement or support policies when RET industries have matured. Through this mechanism, market potential is updated automatically when the political planning data are renewed (Figure 10).

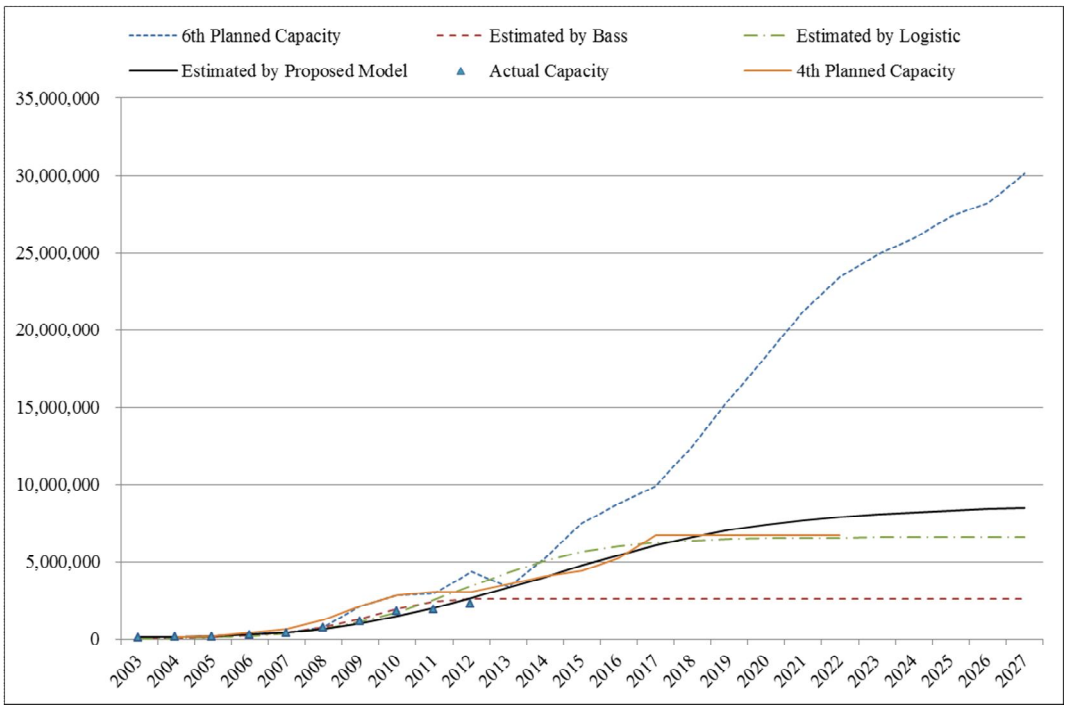

Figure 10. Diffusion patterns of the three models until 2027 (unit: kW)

One of the common assumptions that we deal with in this study is that market potential is static. Under this assumption, the Bass diffusion model, the logistic model, and our proposed model estimate the market potential parameter m as a constant. This assumption makes the models unrealistic since per capita energy consumption is increasing globally, even under circumstances of suppressed energy production from fossil fuels. If the diffusion pattern has improper market potential limits, the estimates might not match with actual adoption at the initial stage (Purohit and Kandpal, 2005). In reality, market potential is dynamic. A dynamic model overcomes this condition as the market potential is set up as a function, for example, by introducing variables like population growth rates and income growth rates. The proposed model incorporates political goals, therefore, making it a dynamic model. The goal of politically planned capacity that is updated periodically and published by the government calibrates the market potential because the proposed model assumes the political goal is less than the market potential and the political goal usually increases over time. Therefore, more ambitious political goals raise the market potential while the learning effect reduces the gap between political goals and actual installed capacity of RETs. This mechanism of updating data works like a dynamic model and improves forecasting accuracies.

Finally, the advantages of the proposed model over the Bass model are demonstrated. In *Chapters 2* and *3*, the limitations of Bass model were pointed out; these are summarized as below.

1. Diffusion models that assume the free competitive market condition are improper to explain the diffusion of RET because the RET lacks competitiveness (Rao and Kishore, 2010).
2. The logistic model is more appropriate for diffusion in technological substitution, especially for industrial innovation (Mahajan and Muller, 1979).
3. The policy effects cannot be measured when the innovativeness coefficient is not significant in the Bass model, and many countries have an insignificant innovativeness coefficient (Davies and Diaz-Rainey, 2011).
4. The RET plant capacity is gradually increasing; however, the Bass model assumes the potential market size to be static.

In contrast, the proposed model focuses on change agents' promotion efforts rather than spontaneous diffusion through communication channels. Moving away from the traditional perspective, this study tries to provide a new perspective by assuming that RET diffusion is induced by policy. Moreover, using the logistic model as a basis to develop the new diffusion model can avoid the situation in which the policy effect cannot be measured when the innovativeness coefficient of the Bass model is insignificant. In the proposed model, the market potential is estimated as a fixed value; however, the market potential can be updated by introducing a new policy plan.

Therefore, the advantages of the proposed model over the Bass model are summarized as follows.

1. The proposed model explains the induced diffusion more realistically with policy intervention than does spontaneous diffusion through the communication channel.
2. As the proposed model is based on the logistic model, it is more appropriate in terms of technology substitution.
3. The explanation of the policy effects does not depend on the innovativeness coefficient of the Bass model.
4. More realistic market potential is estimated based on policy updating.

3.3 Diffusion Pattern Forecasting Based on the Aggregate Market Data

In 2012, the South Korean government introduced the Renewable Portfolio Standard (RPS hereafter), which mandates that the generation of electricity from renewable energy sources increase from 2% in 2012 to 10% in 2022 and that all the power generators that operate more than 500 MW of power producing facilities must supply the applicable percentage of electricity using renewable energy technology[12]. One important characteristic of the South Korean RPS is that it regulates the total ratio of electricity

[12] Korea Energy Management Corporation Renewable Energy Center, Retrieved June 22, 2014, from http://www.energy.or.kr/

generation from RET plants. This means that the RPS policy only mandates the expansion of the entire RET market; the government does not intervene in the competition among the RETs. In other words, free market competition based on cost and technology competitiveness is induced among the RETs in the economic and social environments. Therefore, the proposed model only considers the total RET plant capacity.

Eight specific renewable energy sources are used following the designation of the South Korean RET policy: wind power, ocean energy, solar power, biomass, waste incineration, waste gas, fuel cells, and integrated gasification combined cycle–clean coal technology (IGCC–CCT). We assume that the sum of these sources represents the RET diffusion pattern. To introduce the proposed model, we use aggregated renewable energy sources data for both actual installed capacity data and political capacity goals. First, the actual cumulative data from 2003 to 2012 are provided by The Electric Power Statistics Information System (EPSIS). Second, politically planned capacity data from 2007 to 2027 are obtained from the 3rd DSP (Ministry of Commerce, Industry and Energy, 2006), the 4th DSP (Ministry of Knowledge Economy, 2008a), the 5th DSP (Ministry of Knowledge Economy, 2010), and the 6th DSP (Ministry of Knowledge Economy, 2013)[13]. The DSP is South Korea's national power supply master plan over a 15-year time horizon and is updated every 2 years[14]. In the case of the 6th DSP, data up to 2027 are used since

[13] The 1st (Ministry of Commerce, Industry and Energy, 2002)and 2nd DSP (Ministry of Commerce, Industry and Energy, 2006) are excluded since they have different categories dividing renewable energy sources. However, even though the 1st and 2nd DSP data is included in the model, estimation result shows almost the same values and significance as the case excluding the data.

[14] The 6th DSP was to be released in 2012; however, it was announced only in 2013 after a one-year delay.

long-term planned capacity is required to estimate more accurate market potential based on the country's energy policy. The data used in this study are summarized in Table 9.

Table 9. RETs plant capacity in South Korea

Year	Actual Installed Capacity (kW) (I(d)=1)	Politically Planned Capacity (kW) (I(d)=0)
2003	83,315	
2004	107,905	
2005	155,622	
2006	240,422	
2007	350,609	473,796
2008	728,443	789,496
2009	1,136,352	2,119,500
2010	1,748,980	2,831,600
2011	1,858,660	2,934,900
2012	2,210,773	4,377,500
2013		3,396,000
2014		5,279,000
2015		7,517,000
2016		8,807,000
2017		9,880,000
2018		12,510,000
2019		15,515,000
2020		18,276,000
2021		21,181,000
2022		23,420,000
2023		24,888,000
2024		25,953,000
2025		27,342,000
2026		28,203,000
2027		30,148,000

Source: EPSIS and the 3rd through 6th DSPs

To estimate diffusion models, there are two widely used methods, maximum likelihood estimation (MLE) and non-linear least squares (NLS). Ordinary least squares (OLS) estimation is used rarely since time interval bias is generated during the estimation of a continuous time model with discrete time series data. Another problem with OLS is that it is difficult to measure significance as parameters are estimated indirectly. On account of these limitations, OLS is used only to provide the starting points for MLE or NLS. Usually, MLE and NLS are superior to OLS in terms of forecasting and goodness of fit. Although Schmittlein and Mahajan (1982) describe the advantages of MLE, standard errors are highly likely to be underestimated in the case of new product diffusion estimation using MLE (Srinivasan and Mason, 1986). On the other hand, NLS shows better performance when only a few observations are available (Meade and Islam, 2006). Therefore, we select the NLS method for estimation. Estimation results using R (i386 3.0.1) and TSP (Time Series Processor) 1.15 are shown in Table 10.

Table 10. Estimation results of the proposed model

Parameters	Estimate	Std. Error	t value	Pr(>|t|)
a	-5.1940**	1.676e-01	-30.992	< 2e-16
b	0.3004**	1.254e-02	23.955	< 2e-16
k	0.2583**	3.857e-02	6.697	3.45e-7
m	3.268e+07**	8.098e+05	40.360	< 2e-16

Note: ** indicates statistical significance at the 5% level

All the parameters are statistically significant at the 5% level and have the expected signs. Parameter b, the diffusion rate, is comparatively slow. Sultan et al. (1990) report that the average imitation coefficient is 0.38 for their study's 213 applications, and Mahajan et al. (1995) place the imitation coefficient between 0.3 and 0.5. This seems to be reasonable because South Korea is a latecomer in the RET market, and RET development started only in the early 2000s. Parameter k is significant so it could be said that the gap between politically planned capacity and actual capacity has been reduced by the learning effect. The shortfall rate was about 54.6% in 2009, and has since decreased continuously, to 32.6% in 2011 and 19.4% in 2013. After 2025, there is a shortfall of less than 1%. Specifically, South Korea's market potential for RETs is forecast to be 32,680,000 kW, a marginally larger volume than the politically planned capacity in 2027 of 30,148,000 kW. The reason the politically planned capacity in 2027 is close to the market potential is that long-term political planning from 2008 to 2027 is introduced into our proposed model via a logistic model, which already has an S-shaped diffusion pattern. Therefore, the proposed model assuming the S-shaped curve is fit to the characteristics of the capacity plan and has a higher saturated ceiling than the planned capacity on account of the intervention of the learning effect.

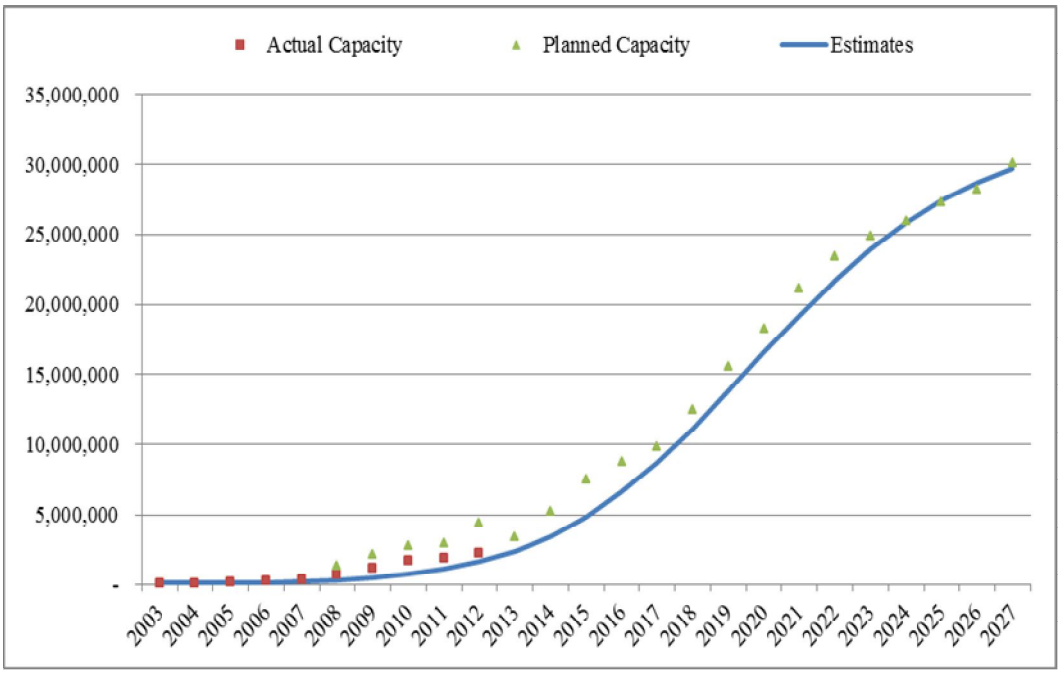

Figure 11. RET diffusion patterns by estimated results (unit: kW)

Lastly, it is worth to considering whether the estimated diffusion rate presents the overall policy effect properly. If the policy effect is the same across countries, insufficient data can be supplemented by introducing other countries' diffusion observations and policy plans, similar to how Mansfield (1961) uses multiple industry data to explain the diffusion rate. However, this assumption is unrealistic since the effects of the specific policy plans cannot be equal across countries. More plant capacity data and policy data within a country should be collected to confirm that the political effects are properly measured with the policy goals. When sufficient data are collected, the estimated diffusion rate as the imitation coefficient can be explained by introducing individual

policy plans into a regression equation—Chen et al. (2012) demonstrate this approach through an empirical study of clean technology in Taiwan. Therefore, it is difficult to investigate how well the proposed model explains the policy effects with current South Korean RET market data; however, this problem can be solved when more data are collected.

3.4 Re-interpretation by Choice Model

This chapter describes how the estimated market potential and diffusion pattern are used together to obtain additional implications within the choice model. The estimates of the diffusion model demonstrate the diffusion pattern at the aggregate level, and aggregate-level data, such as market shares, can be understood as the cumulative adoption results of individual decision makers (Train, 2003). Therefore, in this chapter, the effects of the network externality and subsidy are examined through the choice model. The identified effects of the network externality and subsidy will show a way for the government to minimize the subsidy cost.

3.4.1 Model and Data

Several assumptions are required to connect the adjusted logistic model and the choice model. First, only the first adoption is considered. After a decision maker adopts

the use of an RET plant, he or she is out of the market. Second, a decision maker does not act as a forward-looking consumer. Potential adopters do not consider the strategic timing decision to adopt the RET. Third, adoption costs, including installation and operation costs, are homogeneous across decision makers. Based on this assumption, the network effect and subsidy effect can be easily compared. Fourth, the plant capacity unit, 1 kW, is treated as the adoption unit. Therefore, unfortunately, economies of scale are not allowed.

When consumer i adopts 1 kW from RET plants at time t, the utility is given by

$$U_{it} = V_t(F_{t-1}, r_t) + \varepsilon_{it}, \quad \text{Eq. (22)}$$

where F_{t-1} is the cumulative number of adopters at time $t-1$[15] and r_t is the subsidy level at time t. The stochastic term ε_{it} should be assumed to be uncorrelated across time periods for estimation since the data are actually panel data rather than cross-sectional data. As the stochastic term follows a type-I extreme value distribution, the adoption probability is as follows:

$$P_t = \frac{e^{V_t(F_{t-1}, r_t)}}{1 + e^{V_t(F_{t-1}, r_t)}}. \quad \text{Eq. (23)}$$

Then, demand at time t is defined by subtracting the cumulative number of adopters

[15] The number of adopters is the same as the number of installed RET plants estimated by the adjusted diffusion model in *Chapter 3.3*.

at time $t-1$ from the cumulative number of adopters at time t; demand is then derived by multiplying the number of potential consumers at time t by the adoption probability above. This approach is suggested by Lobel and Perakis (2011).

$$F_t - F_{t-1} = (M - F_{t-1})\frac{e^{V_t(F_{t-1}, r_t)}}{1 + e^{V_t(F_{t-1}, r_t)}},$$
$$\frac{F_t - F_{t-1}}{M - F_{t-1}} = \frac{e^{V_t(F_{t-1}, r_t)}}{1 + e^{V_t(F_{t-1}, r_t)}}. \quad \cdots\cdots \text{Eq. (24)}$$

The left-hand side is the same as Eq. (7) in *Chapter 2.1.2.* Mansfield (1961) identifies this term as the diffusion rate and assumes that various factors, such as the network effect and profitability, affect the diffusion rate. On the other hand, Eq. (24) explains the diffusion rate with the utility structure by incorporating the network effect and the subsidy level. Let the diffusion rate, the left term on the second line of Eq. (24), be λ_t for convenience as in Mansfield (1961). Then, the demand model is expressed as

$$\ln\left(\frac{\lambda_t}{1-\lambda_t}\right) = V_t(F_{t-1}, r_t). \quad \cdots\cdots \text{Eq. (25)}$$

The left-hand side is calculated simply by using the estimated results in *Chapter 3.3*; therefore, the simple regression remains if the observed part has a linear function form. To identify the network externality and the subsidy effect, the observed part is defined as

$$V_t(F_{t-1}, r_t) = \beta_1 \ln\left(\frac{F_{t-1}}{M}\right) + \beta_2 r_t + C + \xi_t, \cdots\cdots \text{Eq. (26)}$$

where C is the alternative specific constant (ASC hereafter), which indicates the baseline utility to adopt 1 kW from the RET plant. As a network externality, a logarithm form is introduced. Without the logarithm, the value of the network explanatory variable is extremely small because the market potential is too large compared to the number of early cumulative adoptions. A more important reason to use the logarithm form is that the marginal effect of the network externality on potential consumers should decrease over time since the number of consumers that have not adopted decreases over time. This means that the effect of the network externality follows a concave functional form in terms of the number of adoptions (Vives, 1997). Regarding the market potential, it is assumed that significant estimates are given by the adjusted logistic model. The error term with respect to time t still remains. In order to estimate Eq. (26), two issues should be considered. First, the error term ξ_t might have a correlation with the endogenous variable, the network externality variable. This can be solved by introducing instrument variables as in BLP (1995). Second, it should be confirmed that ξ has no autocorrelation. Therefore, in the next chapter, the autocorrelation of the error term ξ is tested using the Durbin–Watson statistic (Durbin and Watson, 1951).

r_t represents the subsidy level from the government. The government's budget for RETs is composed of investment in diffusion and investment in technology development.

Investment in diffusion includes expenditures for diffusion assistance, loans for RET plant installation, and expenditures for feed-in tariffs (FIT). In contrast, investment in technology development provides utilities to developers rather than to adopters; this budget does not directly increase the adopter's utility. Therefore, only government investment in RET plant diffusion is considered. The information for total government investment in RETs from 2003 to 2012 was collected through Statistics Korea. Investment in RET plant diffusion comprises 89.7% of the total government investment in RETs (Ministry of Knowledge Economy, 2008b). For the estimation, investment in RET plant diffusion is divided by the capacity of the installed RET plants so that r_t denotes the amount of the subsidy for 1 kW from a RET plant at time t.

3.4.2 **Results**

As stated previously, the estimation should consider the endogeneity. The two-stage least squares (2SLS hereafter) method is applied to estimate Eq. (26) with instrument variables. The instrument variables include the market potential, total plant capacity (including non-RETs plants), and the square value of the network externality variable. The estimation results are shown in Table 11.

Table 11. Estimation result based on the binary logit demand model

Parameters	Estimate	Std. Error	t-ratio	Pr(>|t|)
C	-0.9519**	0.4646	-2.049	0.041
β_1	1.4727**	0.1275	11.544	0.000
β_2	2.9507**	1.454	2.029	0.042
Adjusted R^2	0.9493			

Note: ** indicates statistical significance at the 5% level and the unit of variable r_t is million KRW.

All the parameters are statistically significant at the 5% level. The baseline utility, which is gained by adopting 1 kW from an RET plant, is negative, as expected, while the marginal effects of the network externality and subsidy are positive. With regard to the autocorrelation problem, the critical interval of 1% significance for two explanatory variables and 9 observations to accept the null hypothesis of no autocorrelation is between 1.389 (d_U) and 2.611 ($4-d_U$) according to Savin and White (1977). The Durbin–Watson value of the estimated model is 1.505, so the null hypothesis is accepted.

Even using the estimation results, the future diffusion pattern cannot be forecasted unless the future subsidy levels are provided. Instead, the required subsidies for a certain level of diffusion can be identified by using the estimates. Two diffusion levels are worth considering. *Case 1* uses the estimated actual diffusion pattern from the forecasts of the adjusted logistic model in *Chapter 3.3*, which has not been achieved on the political goal but overcomes the shortfall through learning by doing. The required subsidies for each year are the requirement to follow forecasted pattern of the proposed diffusion model

(like business as usual); however, it is not enough to reach the government's target level for RET plant diffusion. *Case 2* traces the politically planned RET plant capacity. The South Korean government sets up the planned RET capacity each year from 2014 to 2027 in the 6th DSP, and it is easy to reverse-calculate the required subsidy expenditure to achieve the planning goal for all periods.

Furthermore, in *Case 3*, the government's planned subsidy expenditure is considered. In *Case 3*, the planned subsidies can be compared with the requirement levels of *Case 1* and *Case 2*. The planned subsidy expenditure for RETs is examined in comparison with the latter to accomplish the RET plant diffusion plan. The *Third Renewable Energy Technology Development, Use and Supply Master Plan* (The Ministry of Knowledge and Economy, 2008b) describes the planned subsidy expenditure from 2009 to 2030 as below.

Table 12. The South Korean government's planned subsidy expenditure for RET plant diffusion[16]

Year	2008	2010	2015	2020	2030
Planned subsidy expenditure (unit: KRW billion)	707.7	1,312.6	2062.7	2,045.4	1,073.1

Source: The Ministry of Knowledge and Economy (2008b)

[16] The planned subsidy expenditure is calculated by multiplying the ratio of the investment in RET plant diffusion (89.7%) by the total planned government investment in RETs (Ministry of Knowledge Economy, 2008b).

However, the actual subsidy expenditure has not matched the planned subsidy expenditure in the past. According to Statistics Korea, the actual subsidy expenditure for RETs has been much less than planned, especially since 2010.

Table 13. Actual subsidy expenditure for RETs by the South Korean government

Year	2008	2009	2010	2011	2012	2013
Executed subsidy (unit: KRW billion)	784.4	797.5	876.6	1,003.5	998.2	851.2
Execution Ratio (%)	110.8	-	66.8	61.5	55.5	43.3

Source: Statistics Korea (http://kostat.go.kr/) and The Ministry of Knowledge and Economy (2008b)

As shown in Table 13, subsidy expenditure for RET diffusion is very uncertain. It is improper to assume the planned subsidy expenditure to be the actual subsidy expenditure for RETs under these circumstances. Instead, the planned subsidy expenditure should be calibrated based on the past expenditure ratio. Therefore, the average expenditure ratio for the last three years (53.5%) is applied to calibrate the planned subsidy expenditure, and it is assumed that the subsidy expenditure of *Case 3* increases or decreases linearly for the periods not mentioned in Table 12.

The three cases are summarized below.

Case 1. The required subsidy expenditure to meet the estimated actual RET plant capacity from the adjusted logistic diffusion model

Case 2. The required subsidy expenditure to meet the planned RET plant capacity in the 6th DSP

Case 3. The planned subsidy expenditure by the third Renewable Energy Master Plan after calibration based on the average of the last three-year expenditure ratio

The calculated subsidy expenditure portfolios of *Case 1*, *Case 2*, and *Case 3* are shown below.

Table 14. Subsidy expenditure portfolios for RET diffusion (unit: KRW billion)

Year	*Case 1*	*Case 2*	*Case 3*
2014	1,524.66	3,126.14	1,262.38
2015	1,817.28	3,396.62	1,361.41
2016	2,060.97	1,024.56	1,359.13
2017	2,210.99	- 89.04	1,356.84
2018	2,228.06	3,197.17	1,354.55
2019	2,088.84	3,579.03	1,352.27
2020	1,792.65	2,737.70	1,349.99
2021	1,361.26	3,328.52	1,285.80
2022	832.22	1,555.59	1,221.63
2023	249.01	- 1,722.29	1,157.46
2024	- 347.78	- 4,133.07	1,093.28
2025	- 925.89	- 384.04	1,029.11
2026	- 1,463.30	- 4,031.66	964.93
2027	- 1,947.43	9,303.51	900.76
Total present value	11,200.68	18,228.77	14,279.39

Note: 3% of the discount rate is used as the present value.

In *Case 1*, the subsidy expenditure increases until 2018. After 2023, no more expenditure is required; rather, tax income is expected. The Ministry of Knowledge and Economy (2008) expects that the RETs ensure affordability around 2020; however, this cannot be achieved in the case of business as usual. *Case 2* reflects the irregular expenditure plan to meet the target level. Early in the period, a large amount of expenditure is required to catch up to the politically planned capacity, and a tax gain is generated temporarily in 2017. This is because the network externality is sufficient to

achieve the planned capacity level and the surplus impact is collected as tax income. A large amount of expenditure from 2018 to 2021 brings tax income one year earlier than in *Case 1*; however, a much larger expenditure amount is required in 2027 since a sudden increase of capacity in 2027 is planned in the 6th DSP. This sudden increase near the saturated market cannot be produced by the network externality since the remaining potential adopters are few. It should be overcome through the government expenditure, so a huge expenditure is required in 2027. In *Case 3*, the overall expenditure level is not sufficient to fulfill the politically planned capacity, especially early in the period. As the expenditure is not heavy early in the period, tax income is not produced until 2027. The subsidy expenditure ratio should also be improved to meet the target level. The diffusion patterns from 2014 to 2027 for each case are shown in Figure 12.

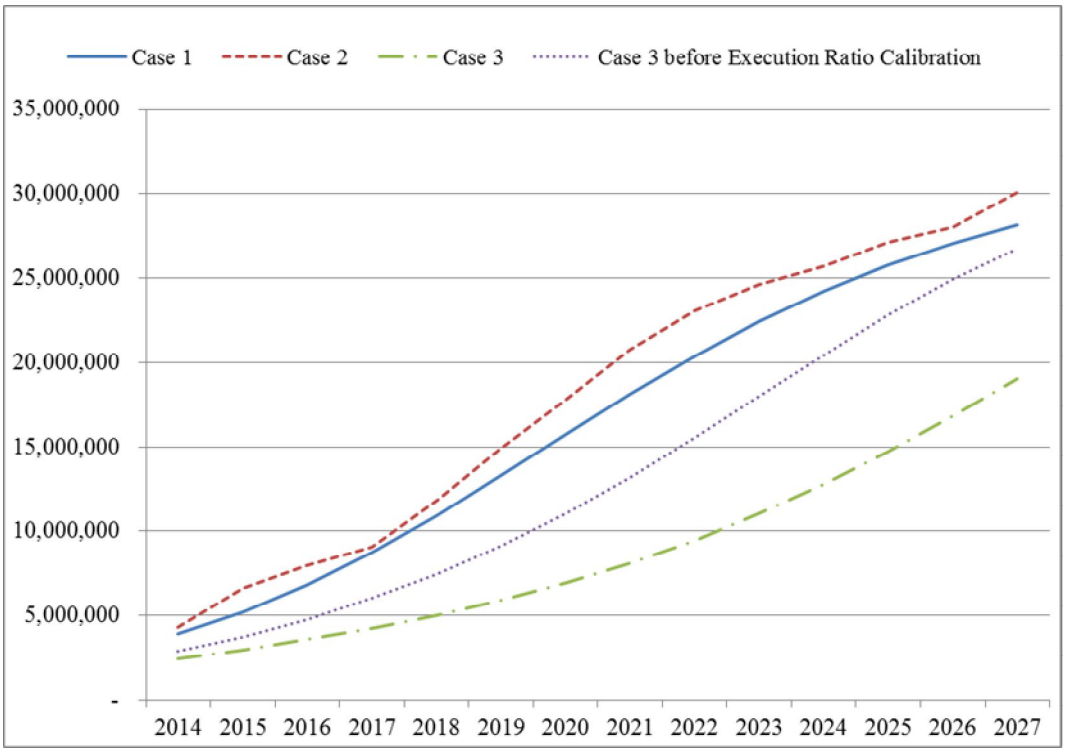

Figure 12. Diffusion patterns by expenditure portfolios (unit: kW)

The politically planned goal for RET capacity cannot be reached even with 100% of s ubsidy expenditure execution (*Case 3* before Execution Ratio Calibration in Figure 12). This means that the political goal for RET diffusion cannot be fulfilled only by improving the expenditure rate. Another point worth considering is the expenditure efficiency. The difference in the total present value between *Case 1* and *Case 2* is immense, whereas the gap between the totally installed capacities is small. In addition, the totally installed capacity of *Case 1* is 148.1% larger than that of *Case 3* even though the expenditure in *Case 3* is 127.5% higher than that in *Case 1*. This shows that annual expenditure

allocation is very important to decrease the total expenditure and improve expenditure efficiency. In the next chapter, a method by which policy makers minimize the total expenditure is discussed.

3.5 Cost Minimization for Policy Makers

Based on the results of the choice model estimation, the policy maker's problem can be derived. The subsidy expenditure should be minimized during diffusion and support should be provided to achieve national target levels for successful policy implementation. Several attempts have been made to estimate the dynamic incentives for RETs. Jaffe and Robert (1995) suggest a framework to empirically test the impacts of policies for RET diffusion. They define the benefit and cost functions and determine the profit-maximizing situation at the point that the marginal cost is equal to the marginal revenue. In another study of incentive policy, Söderholm and Klaassen (2007) focus on the cost function and the expected total benefit of wind power to explain the cumulative capacity with various regulations and policies. The common objective of these studies is to determine the ideal policy. In this chapter, the cost-minimized subsidizing portfolio is suggested via dynamic programming (DP hereafter).

The goal of policy makers is to minimize RET support expenditure while meeting the target capacity of RETs. More precisely, the total present value of the support expenditure should be minimized by annual expenditure allocation. The target capacity level of RET

plants is set as 30,148,000 kW in 2027, as is given in the 6th DSP, and the planning horizon is from 2014 to 2027. The objective function and constraints are given as

$$Cost(F_{2014}, F_{2027\_target}) = \min_{r_{2014},\ldots,r_{2027}} \sum_{t=2014}^{2027} \delta^{t-2014} r_t F_t$$

$$\begin{aligned}
s.t. \quad & F_t = F_{t-1} + f_t(F_{t-1}, r_t), && \forall t = 2014,\ldots,2027 \\
& f_t(F_{t-1}, r_t) = (M - F_{t-1}) \frac{\exp\left(\beta_1 \ln\left(\frac{F_{t-1}}{M}\right) + \beta_2 r_t + C\right)}{1 + \exp\left(\beta_1 \ln\left(\frac{F_{t-1}}{M}\right) + \beta_2 r_t + C\right)}, && \cdots\cdots \text{Eq. (27)} \\
& && \forall t = 2014,\ldots,2027 \\
& F_t \geq \textit{estimated actual diffusion level as Case } 1, && \\
& && \forall t = 2014,\ldots,2027 \\
& r_t F_t \leq \textit{planned budgets at time } t, && \forall t = 2014,\ldots,2027.
\end{aligned}$$

f_t represents the number of adopters at time t and δ represents the discount rate. For simplicity, the discount rate is assumed to be 3%.

The upper two constraints are defined for coding convenience; however, the lower two constraints are newly added. If the impact of government expenditure is extremely large, concentrating all the expenditure in the first year or the last year might be a solution. However, this not feasible since long-term government subsidy expenditures cannot be spent in a lump sum. Thus, the third constraint, which guarantees that the normal diffusion level, is introduced (the normal diffusion level is set as the estimated actual diffusion level in *Chapter 3.3*). Even under this constraint, the subsidy support level can

be executed intermittently but still achieve the minimum diffusion level, as business as usual does. The last constraint limits the expenditure amount for each year based on the *Third RET Development and Supply Master Plan* (The Ministry of Knowledge and Economy, 2008b); in other words, the 100% expenditure rate is the upper limit of the planned expenditure.

To reflect real circumstances, it is easy to add further constraints such as a minimum or maximum annual expenditure. For instance, Lobel and Perakis (2011) introduce the decreasing subsidy constraint ($r_{t-1} \geq r_t$) to prevent the timing behavior of customers.

The support expenditure in a particular year increases the adoption rate, and the increased adoption rate leads to a larger network effect in the next year. Depending on the amount of support expenditure in the previous year, the required expenditure is higher or lower than the planned expenditure to fulfill the third constraint. Therefore, it is possible that policy makers allocate the expenditure each year to minimize the total present value of the support expenditure to achieve the target diffusion level.

The policy maker problem including the objective function and constraints is solved by DP. The free DP software, LINGO 14.0, is used to calculate the cost-minimized expenditure portfolio.

Year	Cost minimized expenditure
2014	2,359.59
2015	2,468.35
2016	2,390.29
2017	2,314.68
2018	2,241.45
2019	2,170.54
2020	2,101.88
2021	1,941.88
2022	1,789.62
2023	1,644.74
2024	1,506.94
2025	1,375.93
2026	1,251.42
2027	- 10,347.41
Total present value	15,209.88

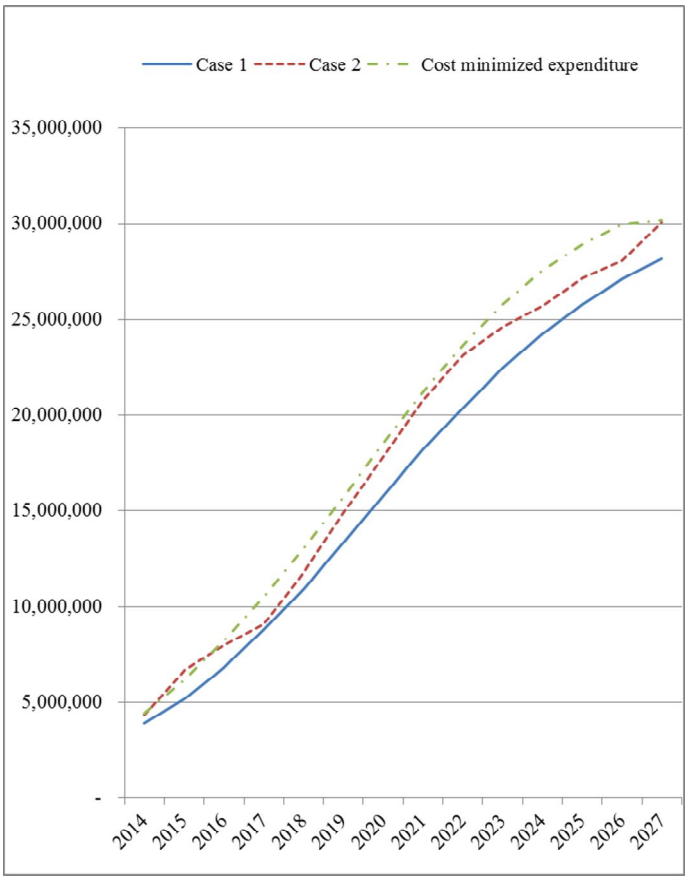

Figure 13. The cost-minimized portfolio and its diffusion pattern

(unit: billion KRW and kW)

Case 2 is artificially generated to satisfy the target levels for all periods; however, if the condition that mandates achievement in all periods is excluded, the subsidy expenditure allocation can be different for the same final target level. *Case 3*, which is the cost-minimized expenditure portfolio, releases this condition and considers only the final target level. By comparing the cost-minimized expenditure portfolio and *Case 2*, 3,019.89 billion KRW can be saved with the fulfillment of the target level. The cost-minimized

expenditure portfolio allocates a large portion of the expenditure to the early years, and the expenditure lasts until 2026. This is to maximize the tax income in the last planning year.

It should be noted that the planned subsidy expenditure in the *Third Renewable Energy Technology Development, Use and Supply Master Plan* (The Ministry of Knowledge and Economy, 2008b) is sufficient to achieve the target level if it is executed as planned. The surplus of the planned subsidy expenditure can be utilized for other purposes. However, although the cost-minimized expenditure portfolio is satisfied with the given constraints, it does not necessarily maximize social welfare. Prior to the cost minimization, the optimal diffusion level should be determined based on the social benefit from RETs plants. Unfortunately, it is difficult to identify or measure the social benefit from RETs plants because some values, such as those for environmental friendliness and reduced greenhouse gases, are not tangible.

Meanwhile, it is worth considering whether concentrated investment in the early stage is appropriate for cost minimization under the circumstance that the unit cost for RET diffusion is decreasing. The required investment in the RET plant diffusion might be reduced over time because of technology improvements. However, RET diffusion investment in the initial market can accelerate technology improvement like R&D investment can. In a PV case study, the World Bank Group (1996) concludes that early investment that includes both diffusion and R&D investment is necessary because it spurs infrastructure distribution and cost reduction. Nemet (2006) claims that policy and

investment should be provided at the early stage because of the learning-by-doing effect and its associated cost reduction. Another reason for early investment is that it carries the advantage of occupying the market ahead of other countries. The South Korean government aims to grow the RET industry as a new leading industry for future exports (Ministry of Knowledge Economy, 2008b). Late investment will delay market penetration, and the Korean RET industry would then be a late mover in the market. Thus, simultaneous early investment in diffusion and R&D investment is recommended to create synergy that will accelerate technology improvement.

3.6 Summary and Discussion

In *Chapter 3*, the factors impeding the RET plant were investigated through a review of the previous literature, and diffusion models that consider the political effects on diffusion were introduced. The common finding from previous studies is that there are barriers to RET diffusion that should be overcome by political efforts.

In addition, the diffusion of South Korean RET plants is analyzed with the proposed adjusted logistic model. The model has two notable features. First, it introduces politically planned data. There are huge gaps between the market potential forecast of classical diffusion models and the politically planned market size, and this difference has made forecasting unreliable up to now. The adjusted logistic model generates more accurate forecasts based on the national energy supply plan. It is shown that the estimated

market potential follows the planned market size, albeit with shortfalls, and provides appropriate forecasts for RET diffusion patterns. Second, the proposed model explains how the actual diffusion pattern follows the planned diffusion pattern through the learning effect. In the early stage, RET industries might be dependent on outside support to develop and sustain themselves; however, the industries could become independent by technology improvement and price reduction, which increase their industrial efficiency. According to the same principle, a reduced difference between planned and actual capacity could be achieved.

Practically, the proposed model improved forecasting accuracies by introducing planned capacity data into the model. Better market potential estimates were gained under the assumption that the latest planned capacity reflects future market size. Even with a small number of observations, the planned capacity fulfilled the lack of data and guided proper estimation. Normally, governments announce their long-term energy plans regularly, and this effectively helps the estimation. Furthermore, it is possible to forecast more accurate diffusion patterns if additional energy plans are made available.

Additional findings are obtained from the re-interpretation of the diffusion pattern using the choice model. The estimated diffusion pattern is combined with the government subsidy plan, and the combined data set provides the marginal effects of the network externality and subsidy. The current subsidy plan is compared to the cost-minimized subsidy plan and evaluated. The planned subsidy expenditure is found to be sufficient to achieve the target diffusion level; however, the allocation is not suitable for cost saving.

Lastly, the cost minimization problem is presented and solved via DP. The key point of the suggested cost minimization problem is that efforts for the network externality effect should occur in the early stage to induce potential adopters. Adequate initial expenditures are necessary to achieve a successful diffusion pattern that satisfies the politically planned goal.

However, there are limitations to the proposed model. First, it does not work well when only a small number of observations exist because the catch-up pattern, in which actual capacity traces planned capacity, should be explained based on the actual capacity data. Second, the proposed model is not suitable for situations in which actual capacity exceeds planned capacity. These limitations provide direction for future research.

Chapter 4. Automobile Demand Forecasting in North America

In *Chapter 4*, new demand forecasting methodology based on the random coefficient logit demand model is suggested. The utility structure that includes consumer heterogeneity can be presented well using the BLP model. Using the utility structure, the market share for a new market can be calculated if the aggregate consumer data in the new market are available. Moreover, consumer groups that are most willing to purchase the product can be identified with the estimated utility structure. The research framework of this essay is shown in Figure 14.

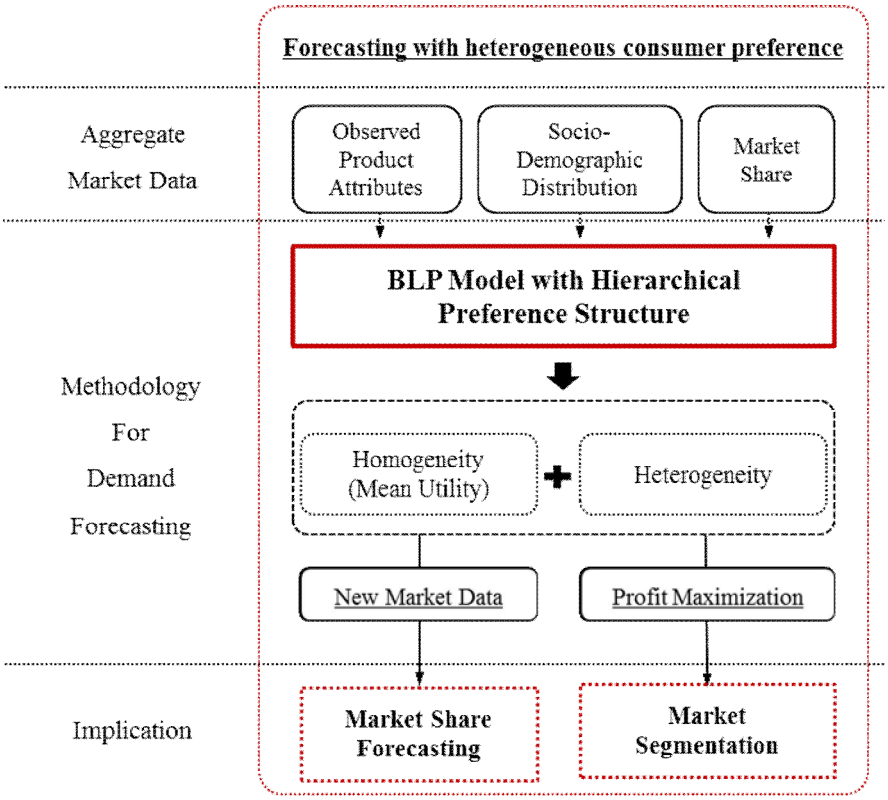

Figure 14. Research framework for essay 2

4.1 Introduction

There are various previous studies of the automobile industry (Boyd and Mellman, 1980; Bajic, 1988; Levinsohn, 1988; Trandel, 1991; McCarthy, 1996; McCarthy et al., 1992). Boyd and Mellman (1980) analyze market share changes by changing the gasoline price based on the hedonic demand model and confirm that fuel-efficient automobiles are preferred when the gasoline price increases. Bajic (1988) uses the hedonic function to

forecast market share changes by changing the automobile price and quality and identifies the key attributes that consumers mainly consider. Similarly, Levinsohn (1988) and Trandel (1991) show that the physical characteristics of automobiles are not key factors of automobile demand. Following Trandel (1991), McCarthy (1996) analyzes the demand elasticity of new cars by income and price. In contrast to these studies, which focus on automobile attributes and environmental factors, McCarthy et al. (1992) focus on consumers. They classify consumers into loyal consumers and new consumers and identify the ratio of loyal consumers in the automobile market based on the profit from loyal consumers. Their findings explain how important companies create loyal consumers for company sustainability.

Other studies based on discrete choice models have been conducted. Since individual-level data are very useful to explain consumer preference, especially when a product has not been released in the market, some studies use stated preference data to examine consumer preference (Ewing and Sarigöllü, 2000; Brownstone and Train, 1999; Kim and Lee, 2007; Ahn and Kim, 2008; Lee and Cho, 2009; Bhat and Sen, 2006). Other studies test consumer preference using both stated data and revealed data for the automobile market (Bhat and Castelar, 2002; Brownstone and Train, 2000).

Whereas automobile industry studies use only market-level data, BLP (1995) explain more realistic substitution patterns among alternatives. Endogeneity caused by price generates bias for the price response parameters in discrete choice theory (Besanko et al., 1998); however, BLP (1995) solve the endogeneity problem between price and

unobserved factors by using instrument variables and explain the choice substitution pattern among alternatives even when individual-level data are available as distribution information. As an expansion of BLP (1995), Sudhir (2001) shows that the competitive relation of the automobile industry discriminatively influences the mini-compact, sub-compact, compact, mid-size, and full-size car categories. In addition, Train and Winston (2007) show that consumers are more sensitive to car attributes other than brand loyalty or dealership and support BLP's (1995) finding that the substitution patterns among alternatives are more realistic.

4.2 Demand Forecasting Methodology Based on the BLP model

4.2.1 BLP model

In this chapter, the random coefficient logit demand model proposed by BLP (1995) is introduced with the mathematical specification of Nevo (2000).

The utility is composed of a deterministic part V_{nj} and a stochastic part ε_{nj}; however, in the BLP model, a more precise definition is introduced to explicitly address the unobserved part. When consumer i chooses product j in market t, the utility is defined as

$$U_{ijt} = (\mathbf{x}_{jt}, p_{jt}, \xi_{jt}, \mathbf{D}_i, \mathbf{v}_i, \varepsilon_{ijt}), \quad \text{Eq. (28)}$$

where $\mathbf{x}_{jt}$ and p_{jt} are the observed product characteristics and price, respectively. ξ_{jt} is the product characteristics unobserved by researchers but recognized by consumers, and $\mathbf{D}_i$ is the individual characteristics. $\mathbf{v}_i$ is consumer i's unobserved characteristics, which are assumed to be random taste variables. ε_{ijt} represents a homoscedastic identically and independently distributed disturbance with zero mean.

In the random coefficient logit model, the parameters that are the marginal utilities of product characteristics should reflect consumer heterogeneity. Then, the parameters have a hierarchical structure and are influenced by individual characteristics as an intersection term with product characteristics and unobserved individual characteristics.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Pi}\mathbf{D}_i + \boldsymbol{\Sigma}\mathbf{v}_i, \quad \mathbf{v}_i \sim \mathbf{P}_{\mathbf{v}}(\mathbf{v}). \quad \cdots\cdots \text{Eq. (29)}$$

It is assumed that consumers have the same fixed parameters, $\boldsymbol{\beta}$, across individuals, and other fixed parameters, $\boldsymbol{\Pi}$, depending on the observed characteristics obtained by multiplying the observed product characteristics and individual characteristics. In addition, in the last term on the right-hand side, $\mathbf{v}_i$ represents random taste, and its effect depends on the product's observed characteristics and the unobserved individual characteristics. In general, the distribution of $\mathbf{v}_i$ is assumed to have a standard multivariate normal distribution. If the data for observed individual characteristics are available as a distribution, socio-demographic information can be introduced as a

distribution.

$$\mathbf{D}_i \sim \hat{\mathbf{P}}_{\mathbf{D}}(\mathbf{D}). \quad \cdots\cdots \text{Eq. (30)}$$

The utility function u_{ijt} can be shown by combining Eqs. (28) and (29). For convenience, henceforth, the parameter of the product price is expressed as α_i separate from $\boldsymbol{\beta}_i$.[17]

$$\begin{aligned} u_{ijt} &= -\alpha_i p_{jt} + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt} + \varepsilon_{ijt} \\ &= -p_{jt}(\alpha + \boldsymbol{\Pi}_\alpha \mathbf{D}_i + \boldsymbol{\Sigma}_\alpha \mathbf{v}_{i\alpha}) + \mathbf{x}_{jt}(\boldsymbol{\beta} + \boldsymbol{\Pi}_{\boldsymbol{\beta}} \mathbf{D}_i + \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \mathbf{v}_{i\boldsymbol{\beta}}) + \xi_{jt} + \varepsilon_{ijt}. \end{aligned} \quad \cdots\cdots \text{Eq. (31)}$$

Eq. (31) can be arranged according to whether the component of utility is the same across individuals.

$$\begin{aligned} u_{ijt} &= -(\alpha p_{jt} + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}) - (\boldsymbol{\Pi}_\alpha \mathbf{D}_i + \boldsymbol{\Sigma}_\alpha \mathbf{v}_{i\alpha}) p_{jt} + \mathbf{x}_{jt}(\boldsymbol{\Pi}_{\boldsymbol{\beta}} \mathbf{D}_i + \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \mathbf{v}_{i\boldsymbol{\beta}}) + \varepsilon_{ijt} \\ &= -(\alpha p_{jt} + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}) + (-p_{jt}, \mathbf{x}_{jt})(\boldsymbol{\Pi}\mathbf{D}_i + \boldsymbol{\Sigma}\mathbf{v}_i) + \varepsilon_{ijt} \\ &= \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \end{aligned} \quad \cdots\cdots \text{Eq. (32)}$$

Now, the utility function is divided into two parts, mean utility and heterogeneous

---

[17] According to the changed notation, $\boldsymbol{\Pi}$, $\boldsymbol{\Sigma}$, and $\mathbf{v}_i$ would be split into $(\boldsymbol{\Pi}_\alpha, \boldsymbol{\Pi}_{\boldsymbol{\beta}})$, $(\boldsymbol{\Sigma}_\alpha, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, and $(\mathbf{v}_{i\alpha}, \mathbf{v}_{i\boldsymbol{\beta}})$, respectively.

utility. Mean utility is the same across individuals, while heterogeneous utility varies among individuals. Some of heterogeneous utility is explained by observed individual characteristics but the others are explained by unobserved individual characteristics. This division between price and unobserved product characteristics is very important in addressing the endogeneity problem because estimating all parameters at once is difficult. The estimation algorithm will be introduced in the following chapter.

The market share of product j in market t is derived by integrating the distribution of observed and unobserved individual characteristics; it does not have a simple closed form like a logit model.

$$s_{jt} = \int_{\mathbf{v}} \int_{\mathbf{D}} \left[\frac{e^{\delta_{jt} + \mu_{ijt}}}{1 + \sum_K e^{\delta_{kt} + \mu_{ikt}}} \right] \mathrm{d}\hat{\mathbf{P}}_{\mathbf{D}}(\mathbf{D})\, \mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}). \quad \text{Eq. (33)}$$

It should be noted that the above market share is the average choice probability considering the individual characteristics when K alternatives exist in the choice set. The utility of outside goods is normalized to zero in the model so that the total sum of market share is not equal to 1. This specification requires calibration when a new product is released or when forecasting is conducted for a different market based on the estimated consumer preference (Train, 2003).

In addition, the price elasticity of market shares is derived based on the aggregate market share. The market share changes according to the price changes of product j are

as follows:

$$\eta_{jkt} = \frac{\partial s_{jt}}{\partial p_{kt}} \cdot \frac{p_{kt}}{s_{jt}} = \begin{cases} -\dfrac{p_{jt}}{s_{jt}} \int_{\mathbf{v}} \int_{\mathbf{D}} \alpha_i s_{ijt}(1 - s_{ijt})\, \mathrm{d}\hat{\mathbf{P}}_{\mathbf{D}}(\mathbf{D})\, \mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}) & if\ \ j = k, \\ \dfrac{p_{kt}}{s_{jt}} \int_{\mathbf{v}} \int_{\mathbf{D}} \alpha_i s_{ijt} s_{ikt}\, \mathrm{d}\hat{\mathbf{P}}_{\mathbf{D}}(\mathbf{D})\, \mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}) & \text{otherwise}. \end{cases} \quad \cdots\cdots\cdots\cdots\text{Eq. (34)}$$

In the next chapter, the specifics of the endogeneity problem are discussed and estimation algorithm using instrument variables is introduced.

4.2.2 Endogeneity and Instrument Variables

Berry (1994) shows that the price has endogeneity with error term ε by considering the supply side. Specifically, producers consider all attributes to fix the price on the Nash equilibrium; however, researchers can only observe some of the attributes. The price, the most commonly observed attribute, is normally selected by researchers to explain producers' behavior. This leads to endogeneity between price and the error term and the introduction of instrument variables.

According to BLP (1995), macro-data trigger the correlation between price and unobserved factors, and a product-specific dummy might solve this correlation. However, this approach has an efficiency problem with handling real market situations in which many alternatives exist, such as the automobile market. Therefore, using instrument

variables is unavoidable. Proper instrument variables are correlated with the endogenous observed factor—the price, in this case—but not with the disturbance terms. In BLP (1995), three kinds of instrument variables are used: the attribute level of the competitor's product, the attribute level sum of all products (excluding the own product level), and the sum of the attribute levels of all the competitor's products. Furthermore, BLP (1995) consider the supply side and argue that unobserved factors affecting the marginal cost should be considered to improve estimation efficiency. In contrast, Nevo (2000) argues that considering the supply side is necessary to introduce the instrument variables. Considering both the demand and supply sides can improve estimation efficiency; however, including the supply side increases computational and programming complexity. Therefore, whether the supply side should be included should be determined according to the application and data.

Nevo (2001) discussed the market-specific valuation of the product in terms of instrument variables. The market-specific valuation of the product does not have a correlation among markets but a correlation within markets that changes over time. This property makes the market-specific valuation of the product valid as an instrument variable (Hausman, 1996). However, the independence assumptions across markets are frequently not acceptable in the real market. Thus, Nevo (2001) adopts a set of replacements representing the marginal costs of the markets, such as the cost of space and the average earnings of markets, instead of the market-specific valuation of the product variables.

Romeo (2010) considers instrument variables for the random coefficient logit demand model. He suggests using the moments of included demographics as instruments and conducts a set of endogenous price simulations to identify how much the new instruments improve numerical performance in various contexts. He finds that the moments of included demographics are useful and help to explain the heterogeneous specifications.

By introducing the instruments, the estimation algorithm becomes somewhat complicated. Berry (1994) proposes a method using the generalized method of moments (GMM hereafter)[18]. The parameters are estimated by minimizing the difference between the actual market share and the calculated market share, as shown in Eq. (35).

$$\underset{\alpha,\boldsymbol{\beta},\boldsymbol{\Pi},\boldsymbol{\Sigma}}{Min}\left\| s_{jt}(p_{jt},\mathbf{x}_{jt},\delta_{jt}(p_{jt},\mathbf{x}_{jt},\xi_{jt};\alpha,\boldsymbol{\beta});\boldsymbol{\Pi},\boldsymbol{\Sigma})-S_{jt}\right\|, \quad \cdots\cdots\cdots\cdots \text{Eq. (35)}$$

where S_{jt} is the actual market share of product j in market t. However, Eq. (35), which has a nonlinear form, is not adopted because too many parameters generated by brand or product dummy variables are difficulty to estimate properly. Berry (1994) tries to separate the nonlinear function from Eq. (35) by including the unobserved product characteristic in the mean utility only. This sufficiently addressed the endogeneity between the price and the unobserved characteristic. That is, the estimation is performed for $(\delta,\boldsymbol{\Pi},\boldsymbol{\Sigma})$ first instead of for $(\alpha,\boldsymbol{\beta},\boldsymbol{\Pi},\boldsymbol{\Sigma})$ because δ is treated as a fixed effect for

[18] For the details of the GMM, see Hansen (1982).

each product. The contraction mapping developed by BLP (1995) makes it possible to find the unique fixed point of δ (For the proof, see BLP (1995)).

$$\delta_{jt}^{h+1} = \delta_{jt}^{h} + \ln S_{jt} - \ln s_{jt}(p_{jt}, \mathbf{x}_{jt}, \delta_{jt}^{h}; \mathbf{\Pi}, \mathbf{\Sigma}), \quad h = 0, ..., H, \quad \cdots\cdots \text{Eq. (36)}$$

where H is the iteration number to guarantee that $\left\| \delta_{\cdot t}^{H+1} - \delta_{\cdot t}^{H} \right\|$ is smaller than a given tolerance level. After numerically computing the value of δ, α and $\boldsymbol{\beta}$ can be estimated easily through the linear function, which is the mean utility part. However, in this estimation, the GMM method is applied to overcome the endogeneity problem. Based on the value of δ, the unobserved product characteristic term is newly defined as

$$\omega_{jt} = \delta_{jt}(S_{jt}; \mathbf{\Pi}, \mathbf{\Sigma}) - (\mathbf{x}_{jt}\boldsymbol{\beta} + \alpha p_{jt}) \equiv \xi_{jt}. \quad \cdots\cdots \text{Eq. (37)}$$

A set of instrument variables Z that is independent of the error term ω_{jt} is given as

$$E\left[\omega_{jt} \mid \mathrm{Z}\right] = 0. \quad \cdots\cdots \text{Eq. (38)}$$

With Eq. (38), the GMM estimator that is satisfied with the moment condition is given as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \omega(\boldsymbol{\theta})' Z \Phi^{-1} Z' \omega(\boldsymbol{\theta}), \quad \cdots\cdots \text{Eq. (39)}$$

where Φ^{-1} is a weighting matrix $E\left[Z'\omega\omega' Z\right]^{-1}$ and $\boldsymbol{\theta}$ denotes all the parameters ($\alpha, \boldsymbol{\beta}, \boldsymbol{\Pi}, \boldsymbol{\Sigma}$). Finally, the parameters that minimize the objective function can be estimated by using search methods such as the non-derivative simplex search method or the quasi-Newton method with an analytic gradient[19].

Several studies have addressed searching methods. In his appendix, Nevo (2000) briefly compares the simplex method and the quasi-Newton method. The simplex method is slow to converge but robust, while the quasi-Newton method is faster but very sensitive to the starting values. In an advanced study, Dubé et al. (2012) propose a mathematical program with equilibrium constraints (MPEC) to avoid the sensitivity problem with starting points. This method is much faster than the classical methods and increases efficiency[20]. Following Dubé et al. (2012), Reynaerts et al. (2012) enhance the convergence property of contraction mapping.

4.2.3 **Model Specifications**

Most of consumer preference studies use individual-level data, but BLP (1995)

---

[19] For details, see Nelder and Mead (1965) and Press (2007) for the simplex method and the quasi-Newton method, respectively.

[20] Note that the MPEC code for MATLAB is not free.

demonstrate that consumer behavior is clarified by market-level data such as market shares. Nevo (2001) uses socio-demographic distribution information and explores the heterogeneous consumer preference in the cereal market. Similarly, in this dissertation, market shares and consumers' socio-demographic information from the Current Population Survey (CPS) will be used to examine consumer preference in the U.S. automobile market. However, consumer preference might be ambiguous if the entire U.S. market is considered because U.S. automobile market is extremely large for companies. For example, when a company utilizes consumer characteristics such as income, since income distributions vary among U.S. regions, it is difficult to find appropriate markets that include the preferred consumer groups. This weakness makes it difficult for a company to target the ideal consumer groups among potential consumers. In addition, the marketing expenditure for the entire U.S. market is likely to be wasted, and its effectiveness is in doubt. To overcome these limitations, an analysis of sub-markets is required.

Introducing information for multiple sub-markets improves estimation efficiency and accuracy. In the national-level approach, the applicable consumer characteristic distribution is only one; however, multiple sub-markets provides different consumer characteristic distributions across markets, which improves the identification of heterogeneity. Moreover, forecasting market shares for a new market is also possible if the consumer characteristic distribution of the new market is given. This approach helps explain the consumer heterogeneity, not only with socio-demographics but also with

environmental variables such as precipitation, snowfall, the gas price, and infrastructure. Therefore, this study analyzes consumer preference with a focus on sub-markets—four U.S. states—and compares the estimated market share and the actual market share in the new market.

Following *Chapter 4.2.1.*, the utility structure is defined as

$$\begin{aligned} u_{ijt} &= -(\alpha p_{jt} + \mathbf{x}_{jt}\boldsymbol{\beta} + \xi_{jt}) - (\boldsymbol{\Pi}_{\alpha}\mathbf{D}_i + \boldsymbol{\Sigma}_{\alpha}\mathbf{v}_{i\alpha})p_{jt} + \mathbf{x}_{jt}(\boldsymbol{\Pi}_{\boldsymbol{\beta}}\mathbf{D}_i + \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{v}_{i\boldsymbol{\beta}}) + \varepsilon_{ijt} \\ &= \delta_{jt} + \mu_{ijt} + \varepsilon_{ijt}. \end{aligned} \quad \cdots\cdots\cdots \text{Eq. (40)}$$

The heterogeneous preference for automobiles varies by consumer characteristics (income, age, and education level) and environmental characteristics (precipitation, snowfall, the gas price, and infrastructure), so the parameters are split into two parts to reflect the heterogeneity with a hierarchical structure.

$$\begin{aligned} \begin{pmatrix} \alpha_i \\ \boldsymbol{\beta}_i \end{pmatrix} &= \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{\Pi}\mathbf{D}_i + \boldsymbol{\Sigma}\mathbf{v}_i \\ &= \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Pi}_{\alpha} \\ \boldsymbol{\Pi}_{\beta} \end{pmatrix}(\mathbf{C}_i, \mathbf{R}_i) + \begin{pmatrix} \boldsymbol{\Sigma}_{\alpha} \\ \boldsymbol{\Sigma}_{\beta} \end{pmatrix} + (\mathbf{v}_{i\alpha}, \mathbf{v}_{i\beta}), \end{aligned} \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{Eq. (41)}$$

where $\mathbf{C}_i$ and $\mathbf{R}_i$ represent the consumer characteristics and environmental characteristics, respectively. As the values of $\mathbf{C}_i$ and $\mathbf{R}_i$ are not directly observed by researchers, the distributions are introduced into the model instead. Let the distributions

of $\mathbf{C}_i$ and $\mathbf{R}_i$ be $\mathbf{P}_{\mathbf{C}}^*(\mathbf{C})$ and $\mathbf{P}_{\mathbf{R}}^*(\mathbf{R})$, respectively. The choice probability s_{jt} for product j in market t is derived by aggregating the individual choice probability; in other words, the choice probability s_{jt} is the integration for the characteristic distributions as

$$\begin{aligned} s_{jt} &= \int_{\mathbf{v}}\int_{\mathbf{D}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_K e^{\delta_{kt}+\mu_{ikt}}}\right]\mathrm{d}\mathbf{P}_{\mathbf{D}}^*(\mathbf{D})\,\mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}) \\ &= \int_{\mathbf{v}}\int_{\mathbf{C}}\int_{\mathbf{R}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_K e^{\delta_{kt}+\mu_{ikt}}}\right]\mathrm{d}\mathbf{P}_{\mathbf{R}}^*(\mathbf{R})\mathrm{d}\mathbf{P}_{\mathbf{C}}^*(\mathbf{C})\,\mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}). \end{aligned} \qquad \cdots\cdots\text{Eq. (42)}$$

Eq. (42) can be calculated numerically via Monte Carlo simulation. The next step is to find estimators that minimize the distance between the aggregated market share and the actual market share. The estimation algorithm is as given in *Chapter 4.2.2*.[21]

In addition, market share forecasting is discussed. When all the parameters are properly estimated, the market shares in a new state can be calculated through random drawing from the distributions. First, the distributions of state-level characteristics, $\hat{\mathrm{P}}_{\mathbf{D}}(\mathbf{D})$, are defined. Most of the socio-demographic data are available in CPS, and data for other environmental characteristics can be collected from various sources. Second, after the distributions of state-level characteristics are defined, the random drawing from

[21] Nevo (2000) provides the estimation procedure step by step in the appendix of "A Practitioner's Guide To Estimation Of Random Coefficients Logit Models Of Demand: The Nitty-Gritty." (retrieved May 20, 2014, from http://faculty.wcas.northwestern.edu/~ane686/supplements/Ras_guide_appendix.pdf)

the distributions is repeated n_s times. Likewise, the unobserved parameter, $\mathbf{v}$, is drawn n_s times from a multivariate normal distribution, $\mathbf{P_v}(\mathbf{v}) \sim \mathrm{N}(0,1)$. Third, the predicted market share, $\hat{s}^1_{jt}$, is calculated using the random draws $(\mathbf{D}_i, \mathbf{v}_i)$ as

$$\hat{s}^q_{jt} = \frac{1}{n_s}\sum_{i=1}^{n_s} s_{ijt} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left(\frac{\exp(\delta_{jt} - (\hat{\mathbf{\Pi}}_\alpha \mathbf{D}_i + \hat{\mathbf{\Sigma}}_\alpha \mathbf{v}_{i\alpha})p_{jt} + \mathbf{x}_{jt}(\hat{\mathbf{\Pi}}_\beta \mathbf{D}_i + \hat{\mathbf{\Sigma}}_\beta \mathbf{v}_{i\beta}))}{1 + \sum_{m=1}^{J}\exp(\delta_{mt} - (\hat{\mathbf{\Pi}}_\alpha \mathbf{D}_i + \hat{\mathbf{\Sigma}}_\alpha \mathbf{v}_{i\alpha})p_{jt} + \mathbf{x}_{jt}(\hat{\mathbf{\Pi}}_\beta \mathbf{D}_i + \hat{\mathbf{\Sigma}}_\beta \mathbf{v}_{i\beta}))}\right), \cdots \text{Eq. (43)}$$

where $\delta_{jt} = -\hat{\alpha}\, p_{jt} + \mathbf{x}_{jt}\hat{\boldsymbol{\beta}}$.

Fourth, the mean ($\overline{s}^q_{jt}$) and standard deviations (σ^q_{jt}) for $\hat{s}^q_{jt}$ are calculated by repeating the second and third steps q times. Lastly, based on the calculated means and standard deviations, the forecasting accuracy can be tested.

$$\begin{aligned} -\sigma^q_{jt} &\le s_{jt} - \overline{s}^q_{jt} \le \sigma^q_{jt}, \\ -2\sigma^q_{jt} &\le s_{jt} - \overline{s}^q_{jt} \le 2\sigma^q_{jt}. \end{aligned} \quad \cdots\cdots \text{Eq. (44)}$$

By counting how many actual market share observations are included in the calculated deviation, the model's identification of heterogeneity using can easily tested.

Lastly, the method for market segmentation is described. The purpose of market segmentation is maximizing companies' profit by identifying the consumer group with the highest utility according to socio-demographic or environmental characteristics. If

consumers are classified into several groups according to the characteristics, the group with the largest market share can be identified through a simulation that provides the predicted market shares for the different consumer characteristic distributions. Let the consumer group g have a characteristic following the distribution $\hat{\mathrm{P}}_{\mathbf{D}}^{g}(\mathbf{D})$. Then, the predicted market share can be calculated as previously described.

$$
\begin{aligned}
s_{jt}^{*} &= \max_{g} s_{jt}(g) \\
&= \max_{g} \frac{1}{n_s} \sum_{i=1}^{n_s} \left(\frac{\exp(\delta_{jt} - (\hat{\mathbf{\Pi}}_{\alpha} \mathbf{D}_i^g + \hat{\mathbf{\Sigma}}_{\alpha} \mathbf{v}_{i\alpha}) p_{jt} + \mathbf{x}_{jt} (\hat{\mathbf{\Pi}}_{\boldsymbol{\beta}} \mathbf{D}_i^g + \hat{\mathbf{\Sigma}}_{\boldsymbol{\beta}} \mathbf{v}_{i\boldsymbol{\beta}}))}{1 + \sum_{m=1}^{J} \exp(\delta_{mt} - (\hat{\mathbf{\Pi}}_{\alpha} \mathbf{D}_i^g + \hat{\mathbf{\Sigma}}_{\alpha} \mathbf{v}_{i\alpha}) p_{jt} + \mathbf{x}_{jt} (\hat{\mathbf{\Pi}}_{\boldsymbol{\beta}} \mathbf{D}_i^g + \hat{\mathbf{\Sigma}}_{\boldsymbol{\beta}} \mathbf{v}_{i\boldsymbol{\beta}}))} \right),
\end{aligned}
\qquad \cdots\cdots \text{Eq. (45)}
$$

where $\delta_{jt} = -\hat{\alpha} p_{jt} + \mathbf{x}_{jt} \hat{\boldsymbol{\beta}}$.

Under the assumption that the profit is proportional to sales and that the price does not change despite the demand increase, the company can identify the consumer group that is most willing to choose its product. The above calculation can be applied to the environmental characteristics as well.

In summary, this chapter describes the methodology used to examine consumer preference by considering sub-markets. The consumer and environmental characteristics provide more information to identify the heterogeneity, and the estimates can be used to forecast a market share in a new market and identify the consumer group with the highest purchasing probability.

4.3 Analysis of Consumer Preference

In this chapter, consumer preference in the U.S. automobile market is analyzed. To easily identify the heterogeneous preference, four different states, California, Colorado, New York, and Texas, are chosen, and the aggregate market shares and socio-demographic information collected from each state are used.

4.3.1 Model and Data

Automobile attributes are selected based on the previous literature (Berry, 1994; BLP, 1995; Brownstone and Train, 1999). The chosen characteristics are price (in 10,000 USD), fuel efficiency (in 100 miles per gallon), weight per horsepower (in 100 lb/hp), and interior space (in 10,000 inch^2), which is calculated by multiplying the wheelbase and the width of the car. The actual purchasing price varies across consumers because each dealer offers a different discount or rate. It is assumed that the actual price is the same as manufacturer's suggested retail price (MSRP). The market share information is gathered from the Automotive News Data Center, and the automobile characteristic information is taken from MSN Autos. To identify the heterogeneous preference, both individual and environmental characteristics are considered, and income (in millions of USD) and rainfall (in inches) distributions are selected, and the cases are denoted as *Model 1* and *Model 2*, respectively. These variables are not included in the model simultaneously to

increase estimation efficiency and accuracy[22]. The income and rainfall distribution information is collected from the *Current Population Survey* (CPS) and the *National Environmental Satellite, Data, and Information Service* (NESDIS), respectively. Four separate states, California, Colorado, New York, and Texas, are selected as markets; these show different income and rainfall distribution.

The data sets of the models include the periods from January to December in 2005 and 2008: 24 months in total[23]. A problem is that new models were released and other models were discontinued during these periods, which makes the data sets unbalanced. Such models are excluded from the data sets. In addition, very expensive luxury cars and exotic models with very few sales are also excluded. Another problem is that too many cars exist in the market. BLP (1995) count two different models as the same model if the two different model year cars have the same name and similar characteristics to reduce the number of products. This empirical study does the same. Sports cars and vans are excluded as outside goods. Normally, consumers are concerned with car purchasing within a particular car segment (e.g., sports car, van, midsize sedan, compact sedan). To reduce the number of alternatives, midsize and compact sedans are grouped together, and the 32 models with the highest market share are chosen. The excluded models are treated as outside goods like very expensive luxury cars and some exotic cars since their market shares are small.

---

[22] A model including both income and rainfall variables was tested; however, the measured estimates are not statistically significant.
[23] Therefore, all the price and income values are converted into 2005 values.

Instead of the product dummy, a brand dummy is introduced. It is too difficult to estimate 32 product dummy variables; instead, 15 brand dummy variables are set up by manufacturing company (e.g., Chrysler, Toyota, Hyundai). The reason to introduce brand or product dummy variables is that it is not clear how well the observed characteristics explain the utility. Moreover, the brand dummy variables do not vary across time and markets, so they improve the fit of the model (Nevo, 2000). According to Berry (1994), this property decreases dependence on instrument variables for the estimation since the correlations between the price and brand effects such as quality are entirely explained by the brand dummy variables. However, data for more than one market are required to use brand dummy identification (BLP, 1995). Both models use the instrument variables suggested by BLP (1995): the attribute level of the competitors' products, the attribute level sum of all products excluding the own product level, and the attribute level sum of all the competitors' products.

Model 1, which considers income as an individual characteristic, has the following utility function:

$$\begin{aligned} u_{ijt} = {} & \beta^0 + \alpha p_{jt} + \beta^{cfe} x_{jt}^{cfe} + \beta^{wthp} x_{jt}^{wthp} + \beta^{space} x_{jt}^{space} + \sum_{h=1}^{15} \beta^{brand\_h} x_{jt}^{h} \\ & + \pi^{p\_income} p_{jt} d_i^{income} + \pi^{space\_income} x_{jt}^{space} d_i^{income} \\ & + \sigma^{p} \nu_i^{p} p_{jt} + \sigma^{space} \nu_i^{space} x_{jt}^{space} + \xi_{jt} + \varepsilon_{ijt}, \end{aligned} \qquad \cdots\cdots\cdots \text{Eq. (46)}$$

where α, β, π, and σ are the parameters and x_{jt}^{h} is a brand dummy variable[24].

Then, the calculated market share is shown as below.

$$
\begin{aligned}
s_{jt} &= \int_{\mathbf{v}}\int_{\mathbf{D}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_{K} e^{\delta_{kt}+\mu_{ikt}}}\right] \mathrm{d}\mathbf{P}_{\mathbf{D}}^{*}(\mathbf{D})\ \mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}) \\
&= \int_{v^{p}}\int_{v^{space}}\int_{d^{income}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_{K} e^{\delta_{kt}+\mu_{ikt}}}\right] \mathrm{d}\hat{\mathrm{P}}_{d^{income}}(d^{income})\mathrm{d}\mathrm{P}_{v^{space}}(v^{space})\ \mathrm{d}\mathrm{P}_{v^{p}}(v^{p}),
\end{aligned}
\qquad \cdots\cdots \text{Eq. (47)}
$$

where $\delta_{jt} = \beta^{0} + \alpha p_{jt} + \beta^{cfe} x_{jt}^{cfe} + \beta^{wthp} x_{jt}^{wthp} + \beta^{space} x_{jt}^{space} + \sum_{h=1}^{15} \beta^{brand\_h} x_{jt}^{h} + \xi_{jt}$,

$$
\mu_{ijt} = \pi^{p\_income} p_{jt} d_{i}^{income} + \pi^{space\_income} x_{jt}^{space} d_{i}^{income} + \sigma^{p} v_{i}^{p} p_{jt} + \sigma^{space} v_{i}^{space} x_{jt}^{space}.
$$

The value of individual income is obtained through stratified sampling based on the given income interval from CPS data since assuming a log-normal distribution does not accurately reflect the real income distribution.

Likewise, the utility of *Model 2*, which includes the rainfall variable as an environmental characteristic, is defined as

$$
\begin{aligned}
u_{ijt} = {} & \beta^{0} + \alpha p_{jt} + \beta^{cfe} x_{jt}^{cfe} + \beta^{wthp} x_{jt}^{wthp} + \beta^{space} x_{jt}^{space} + \sum_{h=1}^{15} \beta^{brand\_h} x_{jt}^{h} \\
& + \pi^{p\_rain} p_{jt} d_{i}^{rain} + \pi^{wthp\_rain} x_{jt}^{wthp} d_{i}^{rain} + \pi^{space\_rain} x_{jt}^{space} d_{i}^{rain} \\
& + \sigma^{p} v_{i}^{p} p_{jt} + \sigma^{wthp} v_{i}^{wthp} x_{jt}^{wthp} + \sigma^{space} v_{i}^{space} x_{jt}^{space} + \xi_{jt} + \varepsilon_{ijt},
\end{aligned}
\qquad \cdots\cdots\cdots \text{Eq. (48)}
$$

---

[24] The fifteen brands are BMW, Cadillac, Chevrolet, Chrysler, Ford, Honda, Hyundai (and Kia), Infiniti, Lexus, Mazda, Mercedes-Benz, Nissan, Subaru, Toyota, and Volkswagen.

where α, β, π, and σ are the parameters and x_{jt}^{h} is the same brand dummy as in *Model 1*. The calculated market share is presented as

$$
\begin{aligned}
s_{jt} &= \int_{\mathbf{v}}\int_{\mathbf{D}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_{K} e^{\delta_{kt}+\mu_{ikt}}}\right] \mathrm{d}\mathbf{P}_{\mathbf{D}}^{*}(\mathbf{D})\ \mathrm{d}\mathbf{P}_{\mathbf{v}}(\mathbf{v}) \\
&= \int_{v^{p}}\int_{v^{space}}\int_{v^{wthp}}\int_{d^{rain}}\left[\frac{e^{\delta_{jt}+\mu_{ijt}}}{1+\sum_{K} e^{\delta_{kt}+\mu_{ikt}}}\right] \mathrm{d}\hat{\mathrm{P}}_{d^{rain}}(d^{rain})\mathrm{d}\mathrm{P}_{v^{wthp}}(v^{wthp})\mathrm{d}\mathrm{P}_{v^{space}}(v^{space})\ \mathrm{d}\mathrm{P}_{v^{p}}(v^{p}),
\end{aligned}
\qquad \cdots\cdots \text{Eq. (49)}
$$

where $\delta_{jt} = \beta^{0} + \alpha p_{jt} + \beta^{cfe} x_{jt}^{cfe} + \beta^{wthp} x_{jt}^{wthp} + \beta^{space} x_{jt}^{space} + \sum_{h=1}^{15} \beta^{brand\_h} x_{jt}^{h} + \xi_{jt}$,

$$
\begin{aligned}
\mu_{ijt} &= \pi^{p\_rain} p_{jt} d_{i}^{rain} + \pi^{wthp\_rain} x_{jt}^{wthp} d_{i}^{rain} + \pi^{space\_rain} x_{jt}^{space} d_{i}^{rain} \\
&\quad + \sigma^{p} v_{i}^{p} p_{jt} + \sigma^{wthp} v_{i}^{wthp} x_{jt}^{wthp} + \sigma^{space} v_{i}^{space} x_{jt}^{space}.
\end{aligned}
$$

Since rainfall cannot be negative, the rainfall distribution is assumed by the censored normal distribution, which adjusts the value to zero if a drawn value is less than zero.

4.3.2 **Results**

In total, 3,072 observations are provided for four states. To calculate the market share, it is assumed that 5,000 individuals exist in each market for both *Model 1* and *Model 2*, and the tolerance level for contraction mapping is set up as 1e-14. Estimation is conducted using the STATA 11.0 software. The estimation results for *Model 1* are provided in Table 15.

Table 15. Estimation results for *Model 1*

Parameters	Estimate	Std. Error	z	P>|z|
constant	-6.5653**	0.9394	-6.99	0.000
cfe	2.9503**	0.9273	3.18	0.001
wthp	-4.3123**	1.6045	-2.69	0.007
space	13.2674**	2.5826	5.14	0.000
space * income	-206.2540**	16.2206	-12.72	0.000
sigma of space	2.43E-06	123.0666	0.00	1.000
price	-1.3815**	0.3225	-4.28	0.000
price * income	31.0589**	1.3242	23.45	0.000
sigma of price	0.0656	8.0129	0.01	0.993
BMW	0.3345	0.2933	1.14	0.254
Cadillac	-0.6169*	0.3309	-1.86	0.062
Chevrolet	0.4017**	0.0566	7.1	0.000
Chrysler (base)	-	-	-	-
Ford	-0.1943**	0.0927	-2.09	0.036
Honda	1.4004**	0.0893	15.69	0.000
Hyundai	-0.0883	0.0629	-1.41	0.160
Infiniti	-0.4699*	0.2810	-1.67	0.094
Lexus	-1.0397**	0.4079	-2.55	0.011
Mazda	-0.0194	0.0722	-0.27	0.788
Mercedes-Benz	-0.4321	0.4977	-0.87	0.385
Nissan	0.6935**	0.0744	9.33	0.000
Subaru	-0.2186**	0.0929	-2.35	0.019
Toyota	1.5696**	0.0817	19.22	0.000
Volkswagen	0.7294**	0.1081	6.75	0.000

Note: ** and * indicates statistical significance at the 5% and 10% level, respectively.

All the observed product variables except for some of the brand dummy variables are significant. Fuel efficiency and interior space have positive preferences, as expected, while price and weight per horsepower have negative preferences. The standard deviations of the price and interior space coefficients are not significant. This means that the heterogeneity is fully explained by the introduced income variable (Nevo, 2000). Regarding the coefficient for the interaction of product characteristics and income, consumers whose income is high prefer expensive cars. However, high-income consumers prefer less interior space. To confirm the heterogeneity, the frequency distributions of the price and interior space coefficients can be described using the given income (Nevo, 2001). Preferences for the 15 brand dummy variables are analyzed by setting Chrysler as the baseline. Chevrolet, Honda, Nissan, Toyota, and Volkswagen are preferred to Chrysler; however, Cadillac, Ford, Infiniti, Lexus, and Subaru are not. The rest of the brands, BMW, Hyundai, Mazda, and Mercedes-Benz, do not have meaningfully different preferences.

The estimation results for *Model 2* are shown in Table 16.

Table 16. Estimation results for *Model 2*

Parameters	Estimate	Std. Error	z	P>|z|
constant	-6.6990**	0.9677	-6.92	0.000
cfe	3.6213**	0.8666	4.18	0.000
space	-6.4477**	2.8417	-2.27	0.023
space * rain	58.9186**	14.1930	4.15	0.000
sigma of space	4.49E-06	11.3146	0.00	1.000
wthp	8.6317	9.0256	0.96	0.339
wthp * rain	-69.7637**	4.0930	-17.04	0.000
sigma of wthp	0.9599	1.6223	0.59	0.554
price	1.0701**	0.3965	2.70	0.007
price * rain	-6.8771*	3.8313	-1.79	0.073
sigma of price	1.80E-06	30.1788	0.00	1.000
BMW	0.6742**	0.2301	2.93	0.003
Cadillac	-0.1361	0.2146	-0.63	0.526
Chevrolet	0.3856**	0.0557	6.93	0.000
Chrysler (base)	-	-	-	-
Ford	-0.3387**	0.1027	-3.30	0.001
Honda	1.3271**	0.0871	15.23	0.000
Hyundai	-0.1155*	0.0623	-1.85	0.064
Infiniti	-0.0677	0.2125	-0.32	0.750
Lexus	-0.4317*	0.2499	-1.73	0.084
Mazda	-0.0771	0.0726	-1.06	0.289
Mercedes-Benz	0.1433	0.3346	0.43	0.669
Nissan	0.6357**	0.0723	8.79	0.000
Subaru	-0.2573**	0.0938	-2.74	0.006
Toyota	1.5250**	0.0793	19.22	0.000
Volkswagen	0.6877**	0.1079	6.37	0.000

Note: ** and * indicates statistical significance at the 5% and 10% level, respectively.

All the observed product variables except weight per horsepower and some of the brand dummy variables are significant. Fuel efficiency has a positive preference; however, price and interior space have negative preferences. This is because the overall preference is explained by the heterogeneity, and the interaction terms make the overall effects reasonable even though mean utility preference has the opposite result from what is expected. For example, the coefficient of price is 1.070 but the coefficient of the interaction term with rainfall is -6.877. The rainfall distribution is assumed to be a censored normal distribution whose values are all greater than zero. Therefore, the results show that price has a negative preference for most consumers. The heterogeneity is appropriately explained by the introduced rainfall variable. In the region with the highest precipitation, consumers tend to prefer cars whose interior space is larger and whose weight per horsepower is low. Consumers in this region are also more price-sensitive. Brand preferences differ from those in *Model 1*. Again using Chrysler as the baseline, BMW, Chevrolet, Honda, Nissan, Toyota, and Volkswagen are preferred, while Ford, Hyundai, Lexus, and Subaru are viewed less favorably.

Table 17 shows a sample of the average of the own price and the cross-price elasticity. In a given market, the elasticity (that is, the percentage of demand changes) for a product in a row is given with respect to a 1% price increase of a product in a column.

Table 17. A sample of the average of the own and cross-price elasticity for midsize sedans in 2005, California (*Model 1*)

	CHEVROLET MALIBU	FORD TAURUS	HONDA ACCORD	HYUNDAI SONATA	INFINITI G35	KIA SPECTRA	NISSAN ALTIMA	TOYOTA CAMRY
CHEVROLET MALIBU	-1.10638	0.003873	0.037087	0.004381	0.009541	0.002898	0.026623	0.038382
FORD TAURUS	0.004879	-1.20119	0.036564	0.004321	0.009393	0.002859	0.026253	0.037827
HONDA ACCORD	0.013504	0.010567	-2.25171	0.011292	0.023356	0.007973	0.079555	0.095057
HYUNDAI SONATA	0.010485	0.008208	0.07422	-1.62386	0.017721	0.005489	0.05246	0.070437
INFINITI G35	0.007526	0.005881	0.0506	0.005841	-2.22202	0.003881	0.036131	0.051579
KIA SPECTRA	0.011884	0.009304	0.08978	0.009404	0.020171	-1.63808	0.063078	0.080845
NISSAN ALTIMA	0.013102	0.010255	0.107526	0.010787	0.022541	0.007571	-2.52599	0.091316
TOYOTA CAMRY	0.008969	0.007016	0.061006	0.006878	0.015279	0.004607	0.04336	-1.67226

Consumers are most price-sensitive to the Nissan Altima. With a 1% price increase, the Nissan Altima loses the biggest market share among midsize sedans. If there is a 1% price decrease, Altima gains the highest demand increase. In contrast, the Chevrolet Malibu is the least price-sensitive of the midsize sedans. The results seem to indicate that if all the listed midsize sedans increase their price by 1%, the Nissan Altima loses the most and the Chevrolet Malibu loses the least.

Table 18. A sample of the average of own and cross price elasticity for compact sedan in August, Texas (*Model 2*)

	CHEVROLET AVEO	CHRYSLER SEBRING	FORD FOCUS	HONDA CIVIC	HYUNDAI ACCENT	MAZDA 3	NISSAN SENTRA	TOYOTA COROLLA	VOLKSWAGEN JETTA
CHEVROLET AVEO	-0.53389	0.004022	0.017861	0.042807	0.006601	0.007721	0.016834	0.039457	0.010079
CHRYSLER SEBRING	0.004459	-0.69167	0.016567	0.039799	0.006125	0.007185	0.01567	0.036647	0.00927
FORD FOCUS	0.004287	0.003587	-0.52623	0.038256	0.005882	0.006907	0.015049	0.035202	0.00885
HONDA CIVIC	0.004803	0.004028	0.017884	-0.51972	0.006609	0.00773	0.016856	0.039505	0.010093
HYUNDAI ACCENT	0.004462	0.003735	0.016567	0.039816	-0.48754	0.007185	0.015656	0.036661	0.009274
MAZDA 3	0.00493	0.004139	0.018377	0.043997	0.006787	-0.6239	0.01731	0.040569	0.010397
NISSAN SENTRA	0.004789	0.004021	0.017837	0.042738	0.006588	0.007711	-0.69282	0.039394	0.010062
TOYOTA COROLLA	0.004579	0.003837	0.017022	0.040863	0.006294	0.007373	0.016071	-0.50536	0.009557
VOLKSWAGEN JETTA	0.00342	0.002837	0.01251	0.030519	0.004655	0.005524	0.012	0.02794	-0.30585

The same interpretation can be applied to the compact sedan case. Consumers are most price-sensitive to the Nissan Sentra. Thus, Nissan should be careful about increasing the price. On the other hand, the Volkswagen Jetta is the least price-sensitive of the compact sedans. The Mazda 3 can withstand the largest price increase in Texas.

4.4 Forecasting

In this chapter, the product market share for a new market will be forecasted based on the estimators and compared to the actual market share. The estimation for consumer heterogeneity was performed for four states in the United States: California, Colorado,

New York, and Texas. It should be noted that the heterogeneity is analyzed for only income and rainfall, and this model structure assumes that other individual characteristics are treated as stochastic terms. Even if the standard deviations of the coefficients are not statistically significant and the introduced individual and environmental variables sufficiently explain the heterogeneity, the other individual and environmental characteristics certainly affect the consumer heterogeneity. To assume that the effects of other individual and environmental characteristics (excluding income and rainfall) can be ignored, the target market should have a similar background. Therefore, the province of Ontario in Canada is selected for forecasting since Canada and the United States have similar characteristics. The price and other attributes of automobiles in Canada are newly collected for forecasting[25].

For *Model 1*, the calculation of the market share of Ontario is as below.

$$\hat{s}_{jt}^{q} = \frac{1}{n_s}\sum_{i=1}^{n_s} s_{ijt} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left(\frac{\exp(\hat{\delta}_{jt} + \hat{\mu}_{ijt})}{1+\sum_{m=1}^{J}\exp(\hat{\delta}_{mt} + \hat{\mu}_{imt})}\right), \quad \text{Eq. (50)}$$

where $\hat{\delta}_{jt} = \hat{\beta}^{0} + \hat{\alpha} p_{jt} + \hat{\beta}^{cfe} x_{jt}^{cfe} + \hat{\beta}^{wthp} x_{jt}^{wthp} + \hat{\beta}^{space} x_{jt}^{space} + \sum_{h=1}^{15} \hat{\beta}^{brand\_h} x_{jt}^{h}$,

$$\hat{\mu}_{ijt} = \hat{\pi}^{p\_income} p_{jt} d_i^{income} + \hat{\pi}^{space\_income} x_{jt}^{space} d_i^{income}.$$

[25] Other characteristics, such as fuel efficiency, weight per horsepower, and interior space are the same in Canada; however, the prices are slightly different. The data are retrieved from http://autos.ca.msn.com/.

The number of products (J) is 32, and it is assumed that there are 3,000 consumers in the market. The market indicator (t) remains because 2005 and 2008 are still included in the forecasting. The income distribution varies across years, so it is expected that the market shares will differ according to each year's income distribution. The income data for Ontario are obtained from Statistics Canada. The number of market share calculations (q) is 1,000. The deviation obtained from the 1,000 calculated market shares provides the range of variation for the forecasts, and the mean of 1,000 market shares will be compared to the actual market shares.

The rainfall distribution for Ontario is utilized for forecasting in *Model 2*. The calculation for the market share of Ontario based on *Model 2* is shown as

$$\hat{s}_{jt}^{q} = \frac{1}{n_s}\sum_{i=1}^{n_s} s_{ijt} = \frac{1}{n_s}\sum_{i=1}^{n_s}\left(\frac{\exp(\hat{\delta}_{jt} + \hat{\mu}_{ijt})}{1+\sum_{m=1}^{J}\exp(\hat{\delta}_{mt} + \hat{\mu}_{imt})}\right), \quad \text{Eq. (51)}$$

where $\hat{\delta}_{jt} = \hat{\beta}^{0} + \hat{\alpha} p_{jt} + \hat{\beta}^{cfe} x_{jt}^{cfe} + \hat{\beta}^{wthp} x_{jt}^{wthp} + \hat{\beta}^{space} x_{jt}^{space} + \sum_{h=1}^{15} \hat{\beta}^{brand\_h} x_{jt}^{h}$,

$$\hat{\mu}_{ijt} = \hat{\pi}^{p\_rain} p_{jt} d_i^{rain} + \hat{\pi}^{wthp\_rain} x_{jt}^{wthp} d_i^{rain} + \hat{\pi}^{space\_rain} x_{jt}^{space} d_i^{rain}.$$

The number of products (J) and the number of consumers are the same as in *Model 1*: 32 and 3,000, respectively. The market indicator (t) indicates the monthly rainfall distributions. The rainfall distributions for Ontario are available from Environment

Canada. A total of 1,000 market shares are calculated for each product, and the mean and deviation for each product are obtained from these 1,000 market share forecasts.

For the exact forecasting, a calibration procedure should be performed. ASCs are introduced to explain the effect of unobserved factors in the utility structure (Train, 2003). When forecasting is conducted for different markets, the average effects of the unobserved factors might be different from those in the analyzed market. Therefore, the ASC should be adjusted for the forecasting. Train (2003) provides an iterative process to calibrate the constant as below.

$$\alpha_j^1 = \alpha_j^0 + \ln(s_j / \hat{s}_j^0), \quad \text{Eq. (52)}$$

where α_j^0 denotes the estimated alternative-specific constant for product j.

By repeating the above iteration, the forecasted market share will be close to the actual market share. In this study, calibration is conducted for the constant term, and the calibrated constant terms for four markets are obtained. Then, the most similar market's calibrated constant will be used to forecast Ontario. For *Model 1*, the median incomes of Colorado and California are the closest to Ontario's median income in 2005 and 2008, respectively. For *Model 2*, the average precipitation of New York is the most similar to Ontario.

The prediction interval is defined using the deviation of the forecasts. The interval is

defined by measuring whether the difference between the actual market share and the forecasted market share is in the range of the absolute deviation. The method briefly described shows that the actual market share is within the prediction interval. The forecasting results are shown in Table 19.

Table 19. The percentage of actual market shares in the prediction interval

	$-\sigma_{\hat{s}_{jt}} \leq s_{jt} - \bar{\hat{s}}_{jt} \leq \sigma_{\hat{s}_{jt}}$	$-2\sigma_{\hat{s}_{jt}} \leq s_{jt} - \bar{\hat{s}}_{jt} \leq 2\sigma_{\hat{s}_{jt}}$
Based on *Model 1*	59.4%	84.4%
Based on *Model 2*	50.5%	68.5%

Table 19 shows that the heterogeneity from income has higher explanatory ability to forecast the market share than precipitation. When the prediction interval allows two time deviations, 84.4% of the actual market shares are included in the prediction interval. Forecasts using *Model 2*, which has slightly lower inclusion with respect to the deviation, show that only half of the actual market shares are included in the interval. Even if the interval is expanded with two times deviation, only 68.5% of the actual market share points are in the prediction interval. This is evidence that the income information better represents the heterogeneity than precipitation information in terms of market share forecasting. Figure 15 displays a sample of the forecasted and actual market shares in Ontario, Canada.

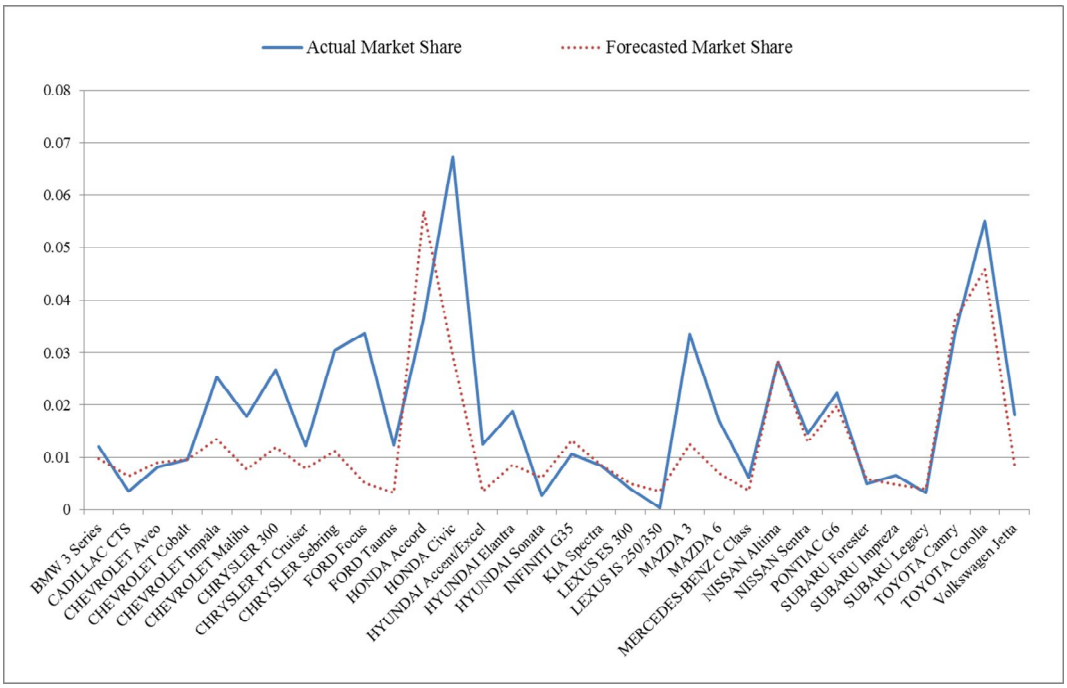

Figure 15. The forecasted market shares in Ontario based on *Model 2* (January 2005)

4.5 Market Segmentation

This chapter discusses market segmentation from the perspective of a company that wants to maximize sales. Heterogeneity causes different utilities depending on consumers' backgrounds. As discussed in the previous chapter, consumers' income brings diverse choice and market shares. From the perspective of a company, it is very important to identify the potential consumers with the highest utility. Each product differs in terms of price, fuel efficiency, and interior space, so the ideal consumer groups might differ

according to the estimated preferences. In this chapter, consumers are classified into multiple groups and regions by level of income and rainfall, respectively, and the group with the highest purchasing power group is identified through simulation.

The simulation procedure is the same as forecasting, but it has predetermined distributions. For forecasting, the distribution should reflect the target market environment. For the simulation, however, the distribution can be defined arbitrarily depending on the marketer's intuition to identify the target consumer group. For example, in the case of income, the group can be divided into high-income, medium-income, and low-income sub-groups. If the company has consumer information from previous sales, it is useful to set up these groups.

As previously, simulations are conducted separately for income and rainfall. Rather than separating consumer groups exclusively, it is assumed that the income of the target consumers is naturally concentrated at a certain level. Therefore, each income distribution is set as a normal distribution with a different mean but the same deviation. The groups have means of i. 20,000 USD, ii. 40,000 USD, iii. 60,000 USD, iv. 80,000 USD, and v. 100,000 USD, and all the groups have the same deviation, 5,000. With this distribution assumption, the optimal consumer group k is selected through Eq. (53).

$$s_{jt}^{*} = \max_{k} s_{jt}(d_i^{income}) = \max_{k} \frac{1}{n_s} \sum_{i=1}^{n_s} \left(\frac{\exp(\hat{\delta}_{jt} + \hat{\mu}_{ijt})}{1 + \sum_{m=1}^{J} \exp(\hat{\delta}_{mt} + \hat{\mu}_{imt})} \right), \quad \cdots\cdots \text{Eq. (53)}$$

where $\hat{\delta}_{jt} = \hat{\beta}^0 + \hat{\alpha} p_{jt} + \hat{\beta}^{cfe} x_{jt}^{cfe} + \hat{\beta}^{wthp} x_{jt}^{wthp} + \hat{\beta}^{space} x_{jt}^{space} + \sum_{h=1}^{15} \hat{\beta}^{brand\_h} x_{jt}^{h}$,

$$\hat{\mu}_{ijt} = \hat{\pi}^{p\_income} p_{jt} d_i^{income} + \hat{\pi}^{space\_income} x_{jt}^{space} d_i^{income},$$

$$d_i^{income} \sim \begin{cases} N(0.02, 0.005) & \text{for } k = 1 \\ N(0.04, 0.005) & \text{for } k = 2 \\ N(0.06, 0.005) & \text{for } k = 3 \\ N(0.08, 0.005) & \text{for } k = 4 \\ N(0.1, 0.005) & \text{for } k = 5. \end{cases}$$

Likewise, the simulation for the rainfall variable is performed. Each region is virtually defined with the different levels of rainfall. Each region has a different interval in the uniform distribution. The basic assumption for grouping is to define a low-rain region, a medium-rain region, a high-rain region and a very high-rain region. Each group has i. greater than 0 and less than 1, ii. greater than 1 and less than 3, iii. greater than 3 and less than 5, and iv. greater than 5 and less than 10 inch intervals for the uniform distribution. Eq. (54) shows the calculation to find the optimal consumer group k.

$$s_{jt}^* = \max_k s_{jt}(d_i^{rain}) = \max_k \frac{1}{n_s} \sum_{i=1}^{n_s} \left(\frac{\exp(\hat{\delta}_{jt} + \hat{\mu}_{ijt})}{1 + \sum_{m=1}^{J} \exp(\hat{\delta}_{mt} + \hat{\mu}_{imt})} \right), \quad \cdots\cdots \text{Eq. (54)}$$

where $\hat{\delta}_{jt} = \hat{\beta}^0 + \hat{\alpha} p_{jt} + \hat{\beta}^{cfe} x_{jt}^{cfe} + \hat{\beta}^{wthp} x_{jt}^{wthp} + \hat{\beta}^{space} x_{jt}^{space} + \sum_{h=1}^{15} \hat{\beta}^{brand\_h} x_{jt}^{h}$,

$$\hat{\mu}_{ijt} = \hat{\pi}^{p\_rain} p_{jt} d_i^{rain} + \hat{\pi}^{wthp\_rain} x_{jt}^{wthp} d_i^{rain} + \hat{\pi}^{space\_rain} x_{jt}^{space} d_i^{rain},$$

$$d_i^{rain} \sim \begin{cases} U(0,1) & \text{for } k=1 \\ U(1,3) & \text{for } k=2 \\ U(3,5) & \text{for } k=3 \\ U(5,10) & \text{for } k=4. \end{cases}$$

Both simulations assume that there are 10,000 consumers in each market, and the simulations are performed using MATLAB R2101a.

Table 20. The optimal consumer group for market share maximization (*Model 1*)

Group 1	Chevrolet Aveo, Chevrolet Cobalt, Chevrolet Impala, Chevrolet Malibu, Chrysler 300, Chrysler PT Cruiser, Chrysler Sebring, Ford Focus, Ford Taurus, Honda Accord, Honda Civic, Hyundai Accent, Hyundai Elantra, Hyundai Sonata, Kia Spectra, Mazda 3, Mazda 6, Nissan Altima, Nissan Sentra, Pontiac G6, Subaru Forester, Subaru Impreza, Subaru Legacy, Toyota Camry, Toyota Corolla, Volkswagen Jetta
Group 2	-
Group 3	Infiniti G35
Group 4	BMW 3 Series, Cadillac CTS
Group 5	Lexus ES, Lexus IS, Mercedes-Benz C-Class

The results show the optimal consumer groups for each product. In *Model 1*, *Group 1* represents the most price-sensitive people, who have low incomes. Among the 32 cars, 26 cars maximize their sales via *Group 1*. This simulation cannot extrapolate the competition

relationships among cars; however, the available evidence seems to suggest that the products in *Group 1* are competitors with each other since the target consumers are alike. On the other hand, premium brands such as Lexus and Mercedes-Benz can maximize their profit by targeting *Group 5*, whose income is highest among the groups. The target consumers of Infiniti, BMW, and Cadillac are medium and medium-high income groups, which are *Groups 3* and *4*, respectively. The simulation of *Model 1* shows that each consumer group divided by income level has a different preference for automobiles.

Table 21. The optimal region for market share maximization (*Model 2*)

Region 1	BMW 3 Series, Cadillac CTS, Chevrolet Aveo, Chevrolet Cobalt, Chevrolet Impala, Chevrolet Malibu, Ford Focus, Ford Taurus, Honda Civic, Hyundai Accent, Infiniti G35, Lexus ES, Lexus IS, Mazda 6, Mercedes-Benz C-Class, Nissan Sentra, Subaru Forester, Subaru Impreza, Subaru Legacy, Toyota Camry, Volkswagen Jetta
Region 2	-
Region 3	-
Region 4	Chrysler 300, Chrysler PT Cruiser, Chrysler Sebring, Honda Accord, Hyundai Elantra, Hyundai Sonata, Kia Spectra, Mazda 3, Nissan Altima, Pontiac G6, Toyota Corolla

In *Model 2*, the results are simpler. The target consumers are divided into only two groups: people living low-rain and heavy-rain regions. All the market shares of the cars preferred by the medium and medium-high income groups (*Groups 3*, *4*, and *5*) can be

maximized in *Region 1*. In contrast, the cars preferred by *Group 1* should focus on Region 4. A possible explanation for this result may be that high-income consumers do not want to live in the heavy-rain region. The people who live in the heavy-rain region tend to prefer cars that are inexpensive and have large amounts of interior space such as the Chrysler 300, Hyundai Elantra, Hyundai Sonata, and Kia Spectra. This is consistent with the estimation results of *Model 2*.

4.6 Summary and Discussion

This chapter has given an account of consumer preference for the U.S. automobile market. Data for four states are introduced to identify consumer heterogeneity. Income and precipitation distributions are included as individual and environmental characteristics, respectively. While previous studies mainly focus on the demand and cost parameters, this study focuses on marketing strategy development.

The results provide confirmatory evidence that consumer preference varies according to individual and environmental characteristics. For example, the interaction of product characteristics and income shows that rich consumers are more willing to purchase an expensive car than others. The finding suggests that other individual and environmental characteristics can be used to explain the consumer preference if significant estimation results are obtained using the chosen characteristics. Brand preference is properly estimated and helps to improve mean utility identification. Moreover, the price elasticity

describes the substitution patterns among products and helps manufactures predict demand changes according to price changes.

The Canadian province of Ontario was selected for forecasting since the unobserved factors are similar to those of the U.S. market. The forecasting results confirm that the income variable provides more accurate forecasts than the rainfall variable for the Ontario automobile market.

Following the forecasting, market segmentation was conducted to find the group with the highest purchasing power. Consumer groups are defined according to income and rainfall levels, and the group with the highest market share is identified for each product. More than half the products have the same target segment, and this segmentation indicates that these products are in very competitive markets. On the other hand, some premium brands, which target rich consumers, are in less competitive markets. Segmentation by rainfall classifies target consumers into two groups. Surprisingly, the cars preferred in the low-rain region are also preferred by the rich consumer group. This finding suggests that both income and rainfall variables should be used to consider the correlation between the two variables.

Finally, two improvements need to be considered. First, the income variable is selected as an individual characteristic in this essay. However, income information is not observable. Using other observable characteristics such as age or gender will help select the ideal consumers for marketing strategy execution. Second, multiple characteristics have not been used. The introduction of multiple individual or environmental

characteristics requires more market share data and estimation efforts. It is recommended that further studies consider methodological improvements and algorithm optimization.

Chapter 5. Conclusion

5.1 Summary

Various demand forecasting methods have been developed for industrialization. Qualitative and quantitative analyses can be used complementarily for better forecasting, and other theories have been combined with these analyses in many ways. Among the advanced methodologies, this dissertation focuses on the diffusion and choice models.

Regarding the diffusion model, the patterns of diffusion for new technology and innovation are relatively well forecasted; however, when the technology or innovation is not self-sustainable, the pattern is underestimated since the model only considers the communication factors. Previous studies point out that political intervention significantly affects diffusion if the technology or innovation is not self-sustainable.

To incorporate these findings, a new diffusion model was suggested based on the logistic model. The suggested diffusion model considers the politically planned goal in the RET market and the learning-by-doing effect to explain the shortfall between the planned and actual diffusion patterns. The newly introduced factors produced more accurate forecasts in the case of the South Korean RET market. The estimated diffusion pattern shows that the market potential for RET plant capacity is 32,680,000 kW, which is

a marginally larger volume than the politically planned 2027 capacity of 30,148,000 kW. In contrast, the classical models estimate that the market potential will be around 2,246,349 kW in 2027, so there is a large gap between the estimated and planned market sizes. The fixed-origin evaluation is used to compare forecasting accuracy. With the fixed time of 2003 to 2009 and three reserve years after 2009, the RMSE and MAPE show that the adjusted logistic model has better forecasting accuracy.

To determine further implications, the forecasted diffusion pattern was re-interpreted using the binary logit model. The effects of the network externality and subsidy were measured, and the results provide a basis for proper policy establishment. The total budget of the *Third Renewable Energy Technology Development, Use and Supply Master Plan* is enough to achieve the target level when executed as planned; however, there can still be improvement in terms of minimizing the budget expenditure. The DP solved the cost minimization problem from the perspective of policy makers. Three cases were assumed to evaluate the cost-minimized plan, and the cost-minimized plan was found to save 3,019.89 billion KRW over the politically planned diffusion plan. One finding from cost minimization is that an increase in the network externality is very important in the early stage of diffusion since stimulating potential adopters leads to faster diffusion and the subsidy can be changed to tax when self-sustainability is secured. Therefore, intensive expenditure at the initial stage is highly recommended for policy success.

Next, the choice model and random utility theory were reviewed briefly, and various DCE studies were introduced. DCE has exhibited outstanding performance in the

marketing and healthcare industries based on the strengths of conjoint analysis, but DCE also inherits conjoint analysis's disadvantages. Conjoint analysis requires a survey to overcome the data shortage, and it has intrinsic problems, such as the hypothetical bias problem during the survey. Some supplemental studies were proposed; however, stated preference data are still required for an accurate explanation of heterogeneous consumer preference.

BLP (1995) proposed a random coefficient logit demand model to estimate consumer preference with only aggregate market data. The random coefficient model has been used with aggregate market data before BLP (1995), but the endogeneity problem had not been explicitly addressed. BLP (1995) introduced instrument variables to solve the endogeneity problem between price and the unobserved factors, and this motivates consumer preference research with aggregate-level market data.

In *Chapter 4*, the consumer preference structure of the U.S. automobile market was considered based on the BLP model. Data from four different markets were collected, and the individual and environmental characteristics were used to measure the heterogeneity. Specifically, the hierarchical preference structure reflected the homogeneity and heterogeneity separately. Income and rainfall distributions were selected to determine consumer heterogeneity. For ease of estimation, brand dummy variables that were assumed to be the same across consumers were introduced. The results showed that the consumer preference varied according to the individual and environmental characteristics and that many brand coefficients were significant and helped identify the mean utility.

The interaction of product characteristics and consumer income has a statistically significant coefficient. Rich consumers tend to prefer expensive cars, and do not prefer cars with large amounts of interior space. Consumers who live in the high-rain region prefer cars with large amounts of interior space. Additionally, the own and cross price elasticities for each market were obtained. The price sensitivity of each product and the substitution pattern among products were determined based on the consumer heterogeneity.

Forecasting in a new market was conducted following the analysis of consumer preference. The province of Ontario in Canada was chosen as a new market, and the estimated coefficients for the U.S. automobile market were used. The forecasted market shares were compared with the actual market shares, and the income variable that represents the individual characteristic produced better forecasts than the rainfall variables for Ontario's automobile market. In addition, market segmentation methodology was revised to identify the consumer group that is most willing to purchase the product. Consumers were classified into several groups according to income and rainfall, and the average forecasted market shares for each group were calculated based on the estimated parameters. The highest market share indicates the target consumer group that maximizes the sales. Through the effect of market segmentation on consumer income, it was confirmed that the market is divided into a high premium brand market and a middle-low premium market. The middle-low premium market is very competitive for manufacturers. Likewise, the segmentation of rainfall showed two separate markets: low-rain and high-

rain regions.

5.2 Contributions and Limitations

This study mainly discusses the quantitative demand forecasting method when the available data are not sufficient. Specifically, two essays were written on the diffusion model and the choice model, respectively.

The first essay suggested a diffusion forecasting model when policy plans induce diffusion artificially. A case study of South Korean RET plant capacity and RPS policy was performed, and the results of the proposed diffusion model were used to consider the cost-minimization problem for a policy maker. In this essay, two contributions are made.

First, the role of change agents and facilitators is directly introduced into the diffusion. Most of the previous diffusion studies depend on the communication channel to trace the diffusion pattern, while the promotion efforts of the change agents and the facilitators were not explicitly addressed. The previous diffusion studies emphasize the effects of mass media and word-of-mouth with the S-curve assumption, but these approaches limit the effects of other factors that affect diffusion, such as change agents' promotion efforts (Bass, 1969; Mansfield, 1961; Rao and Kishore, 2000). Spontaneous diffusion under the free competitive market condition is well forecasted via communication channels; however, it does not provide appropriate forecasts for RETs because RETs do not have self-sustainability (Dulal et al., 2013). Although RETs have technological superiority,

they cannot be diffused by themselves because of high prices or incumbents' market power (Negro et al., 2012; Kemp et al., 1998). For this reason, change agents or facilitators should intervene to ensure RET diffusion. Policy plays a role as a change agent or a facilitator to induce RET diffusion at the national level; thus, this study tied to explain RET diffusion with consideration for the role of change agents and facilitators. Therefore, it is theoretically meaningful to address the change agents' promotion efforts explicitly, which was not done in previous studies.

Second, the diffusion of South Korean RET plant capacity is empirically analyzed in the essay. The proposed model provided more accurate estimates with both the actual capacity data and the planned capacity data, and it was shown that the estimates from the proposed diffusion model can be utilized to suggest a cost-minimizing policy direction for policy makers. The estimated market potential of the South Korean RET market with planned data is more realistic, while the traditional models estimate market potential improperly based on the actual data alone. In addition, the updating of the policy plan suitably calibrated the market potential. Therefore, this empirical analysis shows how diffusion forecasts can be used for policy establishment when only a few observations exist.

However, there are several limitations to this approach. First, the proposed model is developed under the non-competitive market condition because induced RET diffusion is driven not by free competition but by government policy. Under this scenario, considering competitors and the economic climate is not necessary. However, these

factors are very important after the RETs gain self-sustainability. There is no political inducement or support when RETs attain self-sustainability. In this situation, using classical diffusion models that assume the free competition market situation, such as Mahajan and Peterson (1978) and Bayus et al. (2000), are more appropriate for forecasting diffusion. Second, the proposed model is not able to identify the individual policy effect because of data shortages. If long-term observations that include policy execution and RET plant capacity are available, by following Mansfield (1961) and Chen et al. (2012), it is possible to identify the individual policy effects with the estimated diffusion rate. However, South Korean RETs entered the market recently, so additional observations are required to measure the individual policy effects.

The second essay proposed a demand forecasting method to provide consumers' heterogeneous preference information with only aggregate-level data in the situation where individual-level data are not obtainable. Based on the choice model, the essay analyzed consumer preference in the U.S. automobile market as a case study to identify the effects of individual socio-demographic variables and environmental variables. Two contributions are made.

First, it was shown that it is possible to predict a specific market share with only the consumer's socio-demographic distribution information and the new market's environmental information when prior market information does not exist. If a company considers releasing a new product or entering a new market segment, this platform is very useful. For new products or a new market, previous studies that include DCE require

individual-level survey data to predict the market situation. Forecasts using the surveyed information provide detailed consumer preference information. However, forecasting using survey data places a large burden on companies in terms of the time and costs required for the survey, and the survey includes intrinsic problems such as a hypothetical bias and response reliability (Cattin and Wittink, 1982). On the other hand, the proposed forecasting platform using the BLP model provides the same level of consumer preference information as research using individual-level survey data even though the proposed model uses only aggregate market data. Companies can utilize this empirical implication in various ways.

Second, this study shows that the BLP model can be used for demand forecasting, and the properties of the BLP model are still valid when used for demand forecasting even though the BLP model was developed for industrial organization research. BLP (1995) provide an explanation of the market power and strategic behaviors of firms by modeling the unobserved factors and their endogeneity (Kadiyali et al., 2001). This essay focused on a more realistic substitution pattern from that of the BLP model and applied it to demand forecasting. Several managerial contributions were provided through the application based on the estimated realistic substitution pattern of the BLP model. For example, marketers can establish proper marketing strategies by identifying which product is most preferred by a given consumer income segment. Moreover, finding a competitive market for their own products is also possible by considering the environmental conditions of the market. Therefore, this essay shows that the BLP model

can be utilized for demand forecasting and that this approach inherits the properties of the BLP model even for demand forecasting.

Lastly, there are avenues for further research. First, other individual characteristics could be introduced. The core individual or environmental characteristics differ across products and technologies. Various efforts to identify the core characteristics will bring additional implications for marketers. In addition, tangible characteristics have a higher utilization than intangible characteristics since they are easier to use for sales and market segmentation. Second, although many studies use the random coefficient logit demand model, most focus on methodological improvement. The random coefficient logit demand model with aggregate data has a great deal of inherent versatility, and from economics to the science of public administration, tremendous contributions can be made. As a part of this effort, this dissertation is expected to contribute to further studies.

Bibliography

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

Anderson, J. C., Jain, D. C., & Chintagunta, P. K. (1992). Customer value assessment in business markets: A state-of-practice study. *Journal of Business-to-Business Marketing, 1*(1), 3-29.

Armstrong, J. S., & Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method. *Journal of Business Research, 64*(7), 699-706.

Armstrong, J. S., Hamrin, J. (2000). *The renewable energy policy manual*. U.S. Export Council for Renewable Energy.

Ashton, A. H., Ashton, R. H., & Davis, M. N. (1994). White-Collar robotics: Levering managerial decision making. *California Management Review, 37*(1). 83-109.

Automotive News Data Center, Retrieved May 22, 2014, from http://www.autonews.com/section/datacenter

Bajic, V. (1988). Market shares and price-quality relationships: An econometric investigation of the U.S. automobile market. *Southern Economic Journal, 54*(4), 888-900.

Bass, F. M. (1969). A new product growth for model consumer durables, *Management*

Science, 15(1), 215-227.

Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass model fits without decision variables. *Marketing Science, 13*(3), 203-223.

Bayus, B. L. (1987). Forecasting sales of new contingent products: An application to the compact disc market. *Journal of Product Innovation Management, 4*(4), 243-255.

Bayus, B. L., Hong, S., & Labe, R. P. (1989). Developing and using forecasting models of consumer durables: The case of color television. *Journal of Product Innovation Management, 6*(1), 5-19.

Bayus, B. L., Kim, N., & Shocker, A. D. (2000). Growth models for multiproduct interactions: Current status and new directions. *New-product diffusion models*, 141-163.

Berry, S., (1994). Estimating Discrete-Choice models of product differentiation. *RAND Journal of Economics, 25*(2), 242-262.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica, 63*(4), 841-890.

Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science, 44*(11), 1533-1547.

Bhat, C. R., & Castelar, S. (2002). A unified Mixed logit framework for modeling revealed and stated preferences: Formulation and application to congestion pricing analysis in the San Francisco bay area. *Transportation Research: Part B:*

Methodological, 36(7), 593-616.

Bhat, C. R., & Sen, S. (2006). Household vehicle type holdings and usage: An application of the Multiple Discrete-Continuous Extreme Value (MDCEV) model. *Transportation Research: Part B: Methodological, 40*(1), 35-53.

Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. *Energy, 34*(9), 1413-1421.

Bosun, P., & Modrak, V. (2014). Using the Delphi method in forecasting tourism activity. *International Letters of Social and Humanistic Sciences,* (14), 66-72.

Boyd, J. H., & Mellman, R. E. (1980). Effect of fuel economy standards on the U.S. automotive market. *An Hedonic Demand Analysis, 14*(5), 367-378.

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

Brownstone, D., & Train, K. (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics, 89*(1-2), 109-129.

Cattin, P., & Wittink, D. R. (1982). Commercial use of conjoint analysis: A survey. *The Journal of Marketing, 46*(3), 44-53.

Chatterjee, R. A., & Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Science, 36*(9), 1057-1079.

Chen, Y. T., Chang, D. S., Chen, C. Y., & Chen, C. C. (2012). The policy impact on clean

technology diffusion. *Clean Technologies and Environmental Policy, 14*(4), 699-708.

Chow, G. C. (1967). Technological change and the demand for computers. *American Economic Review, 57*(5), 1117-1130.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559-583.

Conrad, K. (1996). Optimal environmental policy for oligopolistic industries under intra-industry trade. In C. Carraro, Y. Katsoulacos & A. Xepapadeas (Eds.), *Environmental policy and market structure* (pp. 65-83): Fondazione Eni Enrico Mattei Series on Economics, Energy and Environment, vol. 4. Dordrecht; Boston and London:Kluwer Academic.

Current Population Survey (CPS), Retrieved May 22, 2014, from http://www.census.gov/

Cuzán, A. G., & Bundrick, C. M. (2009). Predicting presidential elections with equally weighted regressors in fair's equation and the fiscal model. *Political Analysis, 17*(3), 333-340.

Dalla Valle, A., & Furlan, C. (2011). Forecasting accuracy of wind power technology diffusion models across countries. *International Journal of Forecasting, 27*(2), 592-601.

Davies, S. W., & Diaz-Rainey, I. (2011). The patterns of induced diffusion: Evidence from the international diffusion of wind energy. *Technological forecasting and social change, 78*(7), 1227-1241.

Dellaert, B. G., Borgers, A. W., & Timmermans, H. J. (1997). Conjoint models of tourist portfolio choice: Theory and illustration. *Leisure Sciences, 19*(1), 31-58.

Dodson Jr, J. A., & Muller, E. (1978). Models of new product diffusion through advertising and word-of-mouth. *Management Science, 24*(15), 1568-1578.

Dow, J. K., & Endersby, J. W. (2004). Multinomial probit and multinomial logit: A comparison of choice models for voting research. *Electoral studies, 23*(1), 107-122.

Dubé, J.-P., Fox, J. T., & Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica, 80*(5), 2231-2267.

Dulal, H. B., Shah, K. U., Sapkota, C., Uma, G., & Kandel, B. R. (2013). Renewable energy diffusion in Asia: Can it happen without government support? *Energy Policy, 59*, 301-311.

Durbin, J., & Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika, 38*(1-2), 159-178.

Easingwood, C., Mahajan, V., & Muller, E. (1981). A nonsymmetric responding logistic model for forecasting technological substitution. *Technological forecasting and Social change, 20*(3), 199-213.

Easingwood, C. J., Mahajan, V., & Muller, E. (1983). A nonuniform influence innovation diffusion model of new product acceptance. *Marketing Science, 2*(3), 273.

Ediger, V. S., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in

Turkey. *Energy Policy, 35*(3), 1701-1708.

Eggers, F., & Eggers, F. (2011). Where have all the flowers gone? forecasting green trends in the automobile industry with a choice-based conjoint adoption model. *Technological Forecasting and Social Change, 78*(1), 51-62.

Electric Power Statistics Information System (EPSIS), Retrieved May 22, 2014, from https://epsis.kpx.or.kr/

Environment Canada, Retrieved May 22, 2014, from http://climate.weather.gc.ca/data_index_e.html

Ewing, G., & Sarigöllü, E. (2000). Assessing consumer preferences for clean-fuel vehicles: A discrete choice experiment. *Journal of Public Policy & Marketing, 19*(1), 106-118.

Feess, E., & Muehlheusser, G. (2002). Strategic environmental policy, clean technologies and the learning curve. *Environmental and Resource Economics, 23*(2), 149-166.

Floyd, A. (1962). *Trend forecasting: A methodology for figure of merit. technological forecasting for industry and government: Methods and applications*. NJ: Prentice-Hall, Englewood Cliffs. 95-109.

Floyd, R. W. (1962). Algorithm 97: Shortest path. *Communications of the ACM, 5*(6), 345.

Frank, L. D. (2004). An analysis of the effect of the economic situation on modeling and forecasting the diffusion of wireless communications in Finland. *Technological forecasting and social change, 71*(4), 391-403.

Gerrard, P., & Cunningham, J. B. (2003). The diffusion of internet banking among

Singapore consumers. *International Journal of Bank Marketing, 21*(1), 16-28.

Giaccaria, S., & Dalmazzone, S. (2012). *Patterns of induced diffusion of renewable energy capacity: The role of regulatory design and decentralization* (No. 282). Collegio Carlo Alberto.

Givon, M., Mahajan, V., & Muller, E. (1995). Software piracy: Estimation of lost sales and the impact on software diffusion. *Journal of Marketing, 59*(1), 29-37.

Givon, M., Mahajan, V., & Muller, E. (1997). Assessing the relationship between the user-based market share and unit sales-based market share for pirated software brands in competitive markets. *Technological Forecasting and Social Change, 55*(2), 131-144.

Gray, W. B., & Shadbegian, R. J. (1998). Environmental regulation, investment timing, and technology choice. *Journal of Industrial Economics, 46*(2), 235-256.

Green, P. E., & Rao, Vithala R. (1969). *Nonmetric approaches to multivariate analysis in marketing*. Wharton School, University of Pennsylvania.

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

Grubel, H. G. & Lloyd, P. J. (1975). *Intra-industry trade: The theory and measurement of international trade in differentiated products*. London: Macmillan Press.

Guidolin, M., & Mortarino, C. (2010). Cross-country diffusion of photovoltaic systems: modeling choices and forecasts for national adoption patterns. *Technological forecasting and social change, 77*(2), 279-296.

Haaijer, R., Kamakura, W., & Wedel, M. (2001). The 'no-choice'alternative in conjoint choice experiments. *International Journal of Market Research, 43*(1), 93-106.

Haaijer, R., & Wedel, M. (2001). Conjoint choice experiments: General characteristics and alternative model specifications. In *Conjoint measurement* (pp. 345-386). Springer Berlin Heidelberg.

Hall, B. H., & Khan, B. (2003). *Adoption of new technology*. National Bureau of Economic Research, Inc, NBER Working Papers: 9730.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica, 50*(4), 1029-1054.

Hauser, J. R., & Rao, V. R. (2004). Conjoint analysis, related modeling, and applications. *Marketing Research and Modeling: Progress and Prospects,* Springer US. 141-168.

Hausman, J. A. (1997). Valuation of new goods under perfect and imperfect competition. In T. F. Bresnahan & R. J. Gordon (Eds.), *The economics of new goods* (pp. 209-237): National Bureau of Economic Research Studies in Income and Wealth, vol. 58. Chicago and London: University of Chicago Press.

Henderson, J., & Quandt, R. (1987). Financing structure. *American Economic Review, 39*(1), 123-145.

Horsky, D. (1990). A diffusion model incorporating product benefits, price, income and information. *Marketing Science, 9*(4), 342-365.

Islam, T., & Meade, N. (1997). The diffusion of successive generations of a technology: A

more general model. *Technological Forecasting and Social Change, 56*(1), 49-60.

Jaffe, A. B., Newell, R. G., & Stavins, R. N. (2005). A tale of two market failures: Technology and environmental policy. *Ecological Economics, 54*(2-3), 164-174.

Jaffe, A. B., & Stavins, R. N. (1995). Dynamic incentives of environmental regulations: The effects of alternative policy instruments. *Journal of Environmental Economics & Management, 29*(3), S43-S63.

Jeong, G., & Lee, J. (2010). Estimating consumer preferences for online music services. *Applied Economics, 42*(30), 3885-3893.

Jeuland, A. P. (1981). *Parsimonious models of diffusion of innovation. Part A: Derivations and comparisons.* Working Paper, Graduate School of Business, University of Chicago.

Jun, D. B., & Park, Y. S. (1999). A choice-based diffusion model for multiple generations of products. *Technological Forecasting and Social Change, 61*(1), 45-58.

Jun, D. B., Kim, S. K., Park, Y. S., Park, M. H., & Wilson, A. R. (2002). Forecasting telecommunication service subscribers in substitutive and competitive environments. *International Journal of Forecasting, 18*(4), 561.

Kamakura, W. A., & Balasubramanian, S. K. (1988). Long-term view of the diffusion of durables. *International Journal of Research in Marketing, 5*(1), 1-13.

Kauko, K., & Palmroos, P. (2014). The Delphi method in forecasting financial markets--an experimental study. *International Journal of Forecasting, 30*(2), 313-327.

Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through

processes of niche formation: The approach of strategic niche management. *Technology Analysis and Strategic Management, 10*(2), 175-195.

Kemp, R., Soete, L. (1990) *New explorations in the economics of technical change*. London: Printer.

Kemp, R., & Volpi, M. (2008). The diffusion of clean technologies: A review with suggestions for future diffusion analysis. *Journal of Cleaner Production, 16*(1), 14-21.

Kim, Y., Jeong, G., Ahn, J., & Lee, J.-D. (2007). Consumer preferences for alternative fuel vehicles in south korea. *International Journal of Automotive Technology & Management, 7*(4), 327-342.

Kim, W. J., Lee, J. D., & Kim, T. Y. (2005). Demand forecasting for multigenerational products combining discrete choice and dynamics of diffusion under technological trajectories. *Technological Forecasting and Social Change,72*(7), 825-849.

Krishnan, T. V., Bass, F. M., & Kumar, V. (2000). Impact of a late entrant on the diffusion of a new product/service. *Journal of Marketing Research (JMR), 37*(2), 269-278.

Kulendran, N., & King, M. L. (1997). Forecasting international quarterly tourist flows using error-correction and time-series models. *International Journal of Forecasting, 13*(3), 319-327.

Kwon, Y. S., Kim, S. W., Park, S., Lim, S. H., & Lee, J. B. (2009). The information diffusion model in the blog world. *In Proceedings of the 3rd Workshop on Social*

Network Mining and Analysis. ACM.

Lackman, C. L. (1978). Gompertz curve forecasting: A new product application. *Journal of the Market Research Society, 20*(1), 45-47.

Lawrence, M., O'Connor, M., & Edmundson, B. (2000). A field study of sales forecasting accuracy and processes. *European Journal of Operational Research, 122*(1), 151-160.

Lee, J., & Cho, Y. (2009). Demand forecasting of diesel passenger car considering consumer preference and government regulation in South Korea. *Transportation Research: Part A: Policy and Practice, 43*(4), 420-429.

Lee, J., Cho, Y., Lee, J. D., & Lee, C. Y. (2006). Forecasting future demand for large-screen television sets using conjoint analysis with diffusion model. *Technological Forecasting and Social Change, 73*(4), 362-376.

Lekvall, P., & Wahlbin, C. (1973). A study of some assumptions underlying innovation diffusion functions. *Swedish Journal of Economics, 75*(4), 362-377.

Levinsohn, J. (1988). Empirics of taxes on differentiated products: The case of tariffs in the U.S. automobile industry. In R. E. Baldwin (Ed.), *Trade Policy Issues and Empirical Analysis* (pp. 11-40): National Bureau of Economic Research Conference Report series Chicago and London: University of Chicago Press.

Lilien, G. L., Rao, A. G., & Kalish, S. (1981). Bayesian estimation and control of detailing effort in a repeat purchase diffusion environment. *Management Science, 27*(5), 493-506.

Lim, C. M. (2002). Time series forecasts of international travel demand for Australia. *Tourism Management, 23*(4), 389.

Lobel, R., & Perakis, G. (2011). *Consumer choice model for forecasting demand and designing incentives for solar technology*. Social Science Research Network, MIT, Cambridge.

Louviere, J. J., & Hensher, D. A. (1983). Using discrete choice models with experimental design data to forecast consumer demand for a unique cultural event. *Journal of Consumer Research, 10*(3), 348-361.

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research (JMR), 20*(4), 350-367.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. NY: Wiley.

Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology,1*(1), 1-27.

Mahajan, V., & Muller, E. (1979). Innovation diffusion and new product growth models in marketing. *The Journal of Marketing, 43*(4), 55-68.

Mahajan, V., & Muller, E. (1996). Timing, diffusion, and substitution of successive generations of technological innovations: The IBM mainframe case. *Technological Forecasting and Social Change, 51*(2), 109-132.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing, 54*(1), 1-

26.

Mahajan, V., Muller, E., & Bass, F. M. (1995). Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Science, 14*(3), G79-G88.

Mahajan, V., Muller, E., & Kerin, R. A. (1984). Introduction strategy for new products with positive and negative word-of-mouth. *Management Science, 30*(12), 1389-1404.

Mahajan, V., & Peterson, R. A. (1978). Innovation diffusion in a dynamic potential adopter population. *Management Science, 24*(15), 1589-1597.

Mahajan, V., & Peterson, R. A. (1979). Integrating time and space in technological substitution models. *Technological Forecasting and Social Change, 14*(3), 231-241.

Mahajan, V., & Peterson, R. A. (1985). *Models for innovation diffusion*. CA: Sage.

Mahajan, V., Sharma, S., & Buzzell, R. D. (1993). Assessing the impact of competitive entry on market expansion and incumbent sales. *Journal of Marketing, 57*(3), 39.

Mansfield, E. (1961). Technical change and the rate of imitation. *Econometrica, 29*, 741-766.

Marschak, J. (1960). Binary-choice constraints and random utility indicators. *In Proceedings of a Symposium on Mathematical Methods in the Social Sciences, 7*, 19-38.

McCarthy, P. S. (1996). Market price and income elasticities of new vehicles demand. *Review of Economics and Statistics, 78*(3), 543-547.

McCarthy, P. S., Kannan, P. K., Chandrasekharan, R., & Wright, G. P. (1992). Estimating loyalty and switching with an application to the automobile market. *Management Science, 38*(10), 1371-1393.

McDonald, A., & Schrattenholzer, L. (2001). Learning rates for energy technologies. *Energy Policy, 29*(4), 255.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, Zarembka P., *Frontiers in Econometrics*, NY: Academic Press.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447-470.

Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation - a 25-year review. *International Journal of Forecasting, 22*(3), 519-545.

Mickwitz, P., Hyvättinen, H., & Kivimaa, P. (2008). the role of policy instruments in the innovation and diffusion of environmentally friendlier technologies: Popular claims versus case study experiences. *Journal of Cleaner Production, 16*(1), 162-170.

Milkovich, G. T., Annoni, A. J., & Mahoney, T. A. (1972). The use of the Delphi procedures in manpower forecasting. *Management Science, 19*(4), 381-388.

Ministry of Commerce, Industry and Energy. (2002). *1st Demand Supply Program.*

Ministry of Commerce, Industry and Energy. (2004). *2nd Demand Supply Program.*

Ministry of Commerce, Industry and Energy. (2006). *3rd Demand Supply Program.*

Ministry of Knowledge Economy. (2008a). *4th Demand Supply Program.*

Ministry of Knowledge Economy. (2008b). *Third Renewable Energy Technology Development, Use and Supply Master Plan.*

Ministry of Knowledge Economy. (2010). *5th Demand Supply Program.*

Ministry of Knowledge Economy. (2013). *6th Demand Supply Program.*

MSN Auto, Retrieved May 22, 2014, from http://autos.msn.com/

National Environmental Satellite, Data, and Information Service (NESDIS), Retrieved May 22, 2014, from http://cdo.ncdc.noaa.gov/

Nawaz Sharif, M., & Kabir, C. (1976). A generalized model for forecasting technological substitution. *Technological Forecasting and Social Change, 8*(4), 353-364.

Negro, S. O., Alkemade, F., & Hekkert, M. P. (2012). Why does renewable energy diffuse so slowly? A review of innovation system problems. *Renewable & Sustainable Energy Reviews, 16*(6), 3836-3846.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer journal, 7*(4), 308-313.

Nemet, G. F. (2006). Beyond the learning curve: Factors influencing cost reductions in photovoltaics. *Energy Policy, 34*(17), 3218-3232.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy, 9*(4), 513-548.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica, 69*(2), 307-342.

Norton, J. A., & Bass, F. M. (1987). A diffusion theory model of adoption and substitution

for successive generations of high-technology products. *Management Science, 33*(9), 1069-1086.

Orr, L. (1976). Incentive for innovation as the basis for effluent charge strategy. *American Economic Review, 66*(2), 441-447.

Oster, S. (1982). The diffusion of innovation among steel firms: The basic oxygen furnace. *Bell Journal of Economics, 13*(1), 45-56.

Peterson, R. A., & Mahajan, V. (1978). Multi-product growth models. *Research in Marketing, 1*, 201-231.

Potoglou, D., & Kanaroglou, P. S. (2007). Household demand and willingness to pay for clean vehicles. *Transportation Research Part D: Transport and Environment, 12*(4), 264-274.

Prasad, A., & Mahajan, V. (2003). How many pirates should a software firm tolerate?: An analysis of piracy protection on the diffusion of software. *International Journal of Research in Marketing, 20*(4), 337.

Press, W. H. (2007). *Numerical recipes 3rd edition: The art of scientific computing.* Cambridge university press.

Purohit, P., & Kandpal, T. C. (2005). Renewable energy technologies for irrigation water pumping in India: Projected levels of dissemination, energy delivery and investment requirements using available diffusion models. *Renewable & Sustainable Energy Reviews, 9*(6), 592-607.

Purohit, P., & Michaelowa, A. (2008). CDM potential of SPV pumps in India. *Renewable*

& Sustainable Energy Reviews, 12(1), 181-199.

Rao, K. U., & Kishore, V. V. N. (2010). A review of technology diffusion models with special reference to renewable energy technologies. *Renewable & Sustainable Energy Reviews, 14*(3), 1070-1078.

Rapoport, J. (1978). Diffusion of technological innovation among nonprofit firms: A case study of radioisotopes in U.S. hosptials. *Journal of Economics and Business, 30*(2), 108-118.

Reynaerts, J., Varadhan, R., & Nash, J. C. (2012). *Enhancing the convergence properties of the BLP (1995) contraction mapping* (No. 35). Katholieke Universiteit Leuven, Faculteit Economie en Bedrijfswetenschappen, Vives.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Simon and Schuster.

Rogers, E. M., & Kincaid, D. L. (1981). *Communication networks: Toward a new paradigm for research*. NY: The Free Press.

Romeo, C. J. (2010). *Filling out the instrument set in mixed logit demand systems for aggregate data*. Department of Justice, Antitrust Division, EAG Discussions Papers: 201003.

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15*(4), 353-375.

Ryan, M., & Gerard, K. (2003). Using discrete choice experiments to value health care programmes: Current practice and future research reflections. *Applied Health Economics and Health Policy, 2*(1), 55-64.

Saab, S., Badr, E., & Nasr, G. (2001). Univariate modeling and forecasting of energy consumption: The case of electricity in Lebanon. *Energy, 26*(1), 1-14.

Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica, 45*(8), 1989-1996.

Schmittlein, D. C., & Mahajan, V. (1982). Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing Science, 1*(1), 57.

Soderholm, P., & Klaassen, G. (2007). Wind power in Europe: A simultaneous innovation-diffusion model. *Environmental and Resource Economics, 36*(2), 163-190.

Song, I., & Chintagunta, P. K. (2003). A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category. *Quantitative Marketing and Economics, 1*(4), 371-407.

Srinivasan, V., & Mason, C. H. (1986). Nonlinear least squares estimation of new product diffusion models. *Marketing Science, 5*(2), 169-178.

Statistics Canada, Retrieved May 22, 2014, from http://www.statcan.gc.ca/start-debut-eng.html

Statistics Korea, Retrieved May 22, 2014, from https://kostat.go.kr/

Stavins, R., Jaffe, A., & Newell, R. (2005). *A tale of two market failures: Technology and environmental policy*. Resources For the Future, Discussion Papers.

Stoneman, P., & Diederen, P. (1994). Technology diffusion and public policy. *Economic*

Journal, 104(425), 918-930.

Sudhir, K. (2001). Competitive pricing behavior in the auto market: A structural analysis. *Marketing Science, 20*(1), 42-60.

Sultan, F., Farley, J. U., & Lehmann, D. R. (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research (JMR), 27*(1), 70-77.

Tan, S. J. (1994). Predicting innovation adoption: A choice-based approach. *Asia Pacific Advances in Consumer Research, 1*, 72-78.

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437-450.

Teece, D. J. (1980). The diffusion of an administrative innovation. *Management Science, 26*(5), 464-470.

Teotia, A. P. S., & Raju, P. S. (1986). Forecasting the market penetration of new technologies using a combination of economic cost and diffusion models. *Journal of Product Innovation Management, 3*(4), 225-237.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*(4), 273-286.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

Train, K. E., & Winston, C. (2007). Vehicle choice behavior and the declining market share of U.S. automakers. *International Economic Review, 48*(4), 1469-1496.

Trandel, G. A. (1991). The bias due to omitting quality when estimating automobile

demand. *Review of Economics and Statistics, 73*(3), 522-525.

Tsoutsos, T. D., & Stamboulis, Y. A. (2005). The sustainable diffusion of renewable energy technologies as an example of an innovation-focused policy. *Technovation, 25*(7), 753-761.

Usha Rao, K., & Kishore, V. V. N. (2009). Wind power technology diffusion analysis in selected states of India. *Renewable Energy: An International Journal, 34*(4), 983-988.

Van Benthem, A., Gillingham, K., & Sweeney, J. (2008). Learning-by-doing and the optimal solar policy in California. *Energy Journal, 29*(3), 131-151.

Vives, X. (1997). Learning from others: A welfare analysis. *Games and Economic Behavior, 20*(2), 177-200.

Von Bertalanffy, L. (1957). Quantitative laws in metabolism and growth. *The Quarterly Review Of Biology, 32*(3), 217-231.

World Bank Group. (1996). *Photovoltaic market transformation initiative project documents*. International Finance Corporation Staff.

Zwerina, K. (1997). *Discrete choice experiments in marketing: Use of priors in efficient choice designs and their application to individual preference measurement*. Heidelberg: Physica.

Abstract (Korean)

기존의 많은 수요예측방법론 연구들은 소비자의 선호를 분석하거나 수요를 예측하는데 있어 소비자 수준의 대안 자료를 사용해왔다. 이는 시장 수준의 자료를 이용하였을 경우, 적절한 추정이 이루어지지 않거나 모형 속에 내재된 문제점을 극복하기 어렵기 때문이었다. 본 연구에서는 이러한 문제점을 극복하여 시장 수준의 자료만을 바탕으로 수요를 적절히 예측할 수 있는 방법론을 새로이 제안하고, 이를 활용하여 정책입안자 입장에서의 비용최소화 문제와 기업 입장에서의 시장 세분화 전략을 실증적으로 분석해본다.

먼저, 수요예측에서 활발히 이용되고 있는 확산 모형의 경우, 신재생에너지 기술과 같이 정책적 영향력이 크게 개입하는 시장에 대해서는 적절하지 못한 예측을 보여줌을 확인한다. 이를 극복하기 위해 기존의 확산 모형의 틀을 벗어나 정책이 가지는 혁신주도자의 역할에 주목하여 새로운 확산모형을 제안한다. 제안된 확산 모형의 추정 결과는 통계적으로 유의미한 결과를 보여주었으며, 기존 모형과의 예측력 비교에서도 보다 높은 예측력을 보여주었다. 제안된 모형은 새로이 발표되는 신재생에너지 정책에 명시된 목표 설비규모를 변수로 반영함으로써 잠재시장 규모를 보다 현실적으로 반영할 수 있을 뿐만 아니라 정책 목표에 미달되는 설비규모를 학습효과를 통해 적절히 설명하였다. 또한, 의사결정자의 효용 구조를 반영하는 선택모형을 바탕으로 한 확산 패턴의 재해석을 통해 정부의 보조금이

가져오는 효과를 계량적으로 분석하여, 신재생에너지 설비에 대한 정책적 목표 설비규모를 달성하기 위한 효과적인 예산배분계획을 동적계획법을 통해 추가적으로 제시하였다.

다음으로, 시장 데이터만을 이용한 선택모형연구를 기반으로 소비자의 이질적 선호를 확인하고 새로운 수요를 예측해본다. 선택모형 가운데 확률계수-확률효용 모형은 시장지배력에 대한 분석에 주로 사용되어 왔으나 소비자의 선호구조를 시장 데이터만으로 식별 가능하다는 점에서 큰 활용성을 내재하고 있다. 예를 들어, 하나의 시장이 여러 개의 하부시장으로 구분되는 경우, 일부 시장에서의 제품별 시장점유율과 소비자 특성의 분포 정보만을 바탕으로 다른 시장에서의 시장점유율을 예측하는 것이 가능해진다. 나아가 인구통계학 특성을 바탕으로 소비자를 여러 개의 그룹으로 세분화하고 제품의 선택확률이 최대화되는 그룹을 식별함으로써 이윤극대화 전략을 수립하는데 활용될 수 있다. 실증분석으로써 미국 4개 주 자동차 시장을 바탕으로 소비자의 선호를 분석하고, 새로운 시장에 대한 예측된 시장점유율을 실제 시장점유율과 비교해보았다. 그 결과, 제안된 모형은 높은 예측력을 보여주었으며, 소비자 선호 분석에 사용된 인구통계학 정보를 바탕으로 기업의 이윤을 극대화시킬 수 있는 소비자 계층 역시 적절히 식별되었다.

정리하자면, 본 연구는 제한적인 시장데이터만이 가용한 경우에 적용 가능한 정량적 수요 예측방법을 제시하였다. 이를 통해 기존의 확산 모형에서 간과되었던 혁신주도자의 역할을 모형화하였고, 설문에 의존하지 않고 시장데이터만을 바탕으로 한 수요예측이 가능함을 실증적으로 보였다. 두

가지 제안모형은 다른 재화나 서비스에 대해서도 적용될 수 있으며, 상호보완적으로 결합되어 보다 정확한 수요예측 역시 가능할 것으로 기대된다.

주요어 : 수요예측, 확산 모형, 선택 모형, 시장데이터
학 번 : 2009-21082