Industrial Organization Effects of High-Speed Rail Service Introduction in Korea

Jisun Baek

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Industrial Organization Effects of High-Speed Rail Service Introduction in Korea

Jisun Baek

The goal of this thesis is to investigate changes in consumers' choices and their welfare due to the introduction of new products, taking firms' reactions into consideration. I perform empirical analyses using Korean transportation industry data to evaluate the impact of high-speed train introduction on passenger travel. This work adds to the existing literature by considering the changes in product characteristics or the set of products offered to consumers after new product introduction, and investigates how those changes affect consumer welfare. The analysis provides a rich insight into the transportation industry and the relationship between the modes of transportation which contributes to enhancing the quality of government's policies regarding related industries.

The first part of my thesis investigates the changes in utilization of different modes of transportation in Korea after the introduction of high-speed train using a fixed effect model and a difference in differences model. My results show the significant impact of the introduction of high-speed train on the entire transportation industry and provide evidence that modes of transportation not only compete but also complement each other. After high-speed trains were introduced in 2004, inter-city bus and airline industries lost their customers in routes where they directly competed with high-speed rails, while the numbers of rail passengers in-

creased. The losses in the airline industry were particularly severe. On the other hand, the passengers of other rail lines for some routes not connected by high-speed trains but branch routes of high-speed rail lines, increased. The increase was perhaps induced by the consumers who traveled on those routes in order to reach high-speed rail lines.

After the introduction of high-speed trains, other changes such as service schedule adjustment ensued. The results from the reduced form models show only the overall impact of high-speed train introduction, but they cannot disentangle the impact of high-speed train introduction itself from that of ensuing changes. In order to separately examine the impact of high-speed train introduction and that of ensuing changes in product characteristics, I estimate a structural model of the demand for travel that incorporates consumers' heterogeneous preferences over travel schedules into a standard discrete choice model. The model treats the rail company's choice of train schedules as endogenous in order to take the firm's choices of product line into account. My results show that consumers are affected differentially by both the introduction of high-speed trains and the ensuing changes in train schedules. The welfare implications for consumers depend on the availability of high-speed trains in their choice set. Consumers who travel between two cities that are connected by high-speed trains are the main beneficiaries of the new service. However, reductions in schedule frequencies of non-high-speed trains operating along high-speed rail lines, generate losses that offset 50% of gains even for these consumers. Travelers on these lines who are not served by high-speed trains only experience substantial losses due to reduced schedule frequencies. Consumers who travel between two cities that are not located along high-speed rail lines gain from increased train frequencies, and the gains make up for the losses in other markets without high-speed trains. These results highlight the importance of accounting for changes in existing products when analyzing the impact of new product entry on consumers.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to all the people who made this dissertation possible.

I would like to express my sincere appreciation to my advisor, Professor Michael Riordan, and my committee chair, Professor Katherine Ho. Without their guidance and advice this dissertation would not have existed. They generously provided guidance and continuously encouraged me and pushed me to my limits during the process of completing of my dissertation. Not only did they Professor Michael Riordan and Professor Katherine Ho teach me how to approach my research question, they provided important intuition and criticism which enriched my dissertation. I am deeply indebted to them for all the patience and support they gave me since I started writing this thesis.

I also want to thank my dissertation committee members, Professor Chris Conlon, Professor Salanié, and Professor Brett R. Gordon for all of their insight, criticisms and guidance. While I encountered difficulties in my research due to computational issues, Professor Conlon generously shared his knowledge and experience.

I would like to acknowledge Korail and Korea Airport Corporation for providing valuable data for nothing in return. If there are any remaining errors in this final draft, those errors are entirely my own. I would like to emphasize that this dissertation is not a final product, but an ongoing research in empirical economics.

Above all, I thank my husband, Wooram Park for his constant support. He has been my best friend, and my counselor who emotionally supported me when I encountered difficulties in my graduate program at Columbia. He also has discussed my research topic with me every step of my way. It would not have been possible to finish my degree without his support.

Last but not the least, I extend my special thanks to my parents for their everlasting love for me and unconditional support.

Jisun Baek

New York, New York

July, 2012

To my parents

Chapter 1

Introduction

Generally speaking, introducing an additional differentiated product to a market benefits consumers due to the increased number of alternatives if everything else such as price remains the same. However, the effect on consumer welfare is not so simple if producers also change other products' characteristics and the set of other products offered.(Chen & Riordan, 2008) This thesis focuses on the changes occurred due to the introduction of new products. Consumers' reactions regarding their choices and firms' reactions such as changing product characteristics or changing the set of products offered are particularly interesting. The goal of this thesis is to investigate changes in consumers' choices and their welfare due to the introduction of new products, taking firms' reactions into consideration. To be specific, I perform empirical analyses using Korean transportation industry data to evaluate the impact of high-speed train introduction on passenger travel. The analysis provides a rich insight on transportation industry and the relationship between modes of transportation which contributes to enhancing the quality of government's policies regarding related industries.

High-speed rail systems were introduced in South Korea in April 2004. These rail systems continue to significantly impact on the nation's entire transportation industry, thereby affecting its consumers, which has motivated this research. The transportation market in South Korea is particularly interesting for the following reasons : i) data set observes the rich variations of choice sets and final choices, ii) high-speed trains were introduced during the data periods, iii) strong regulations imposed in South Korean transportation industry solve the endogeneity problems.[1]

In the first part of thesis, I will discuss consumers' intermodal and intramodal choices. I address the changes in consumers' intermodal and intramodal choices after high-speed train introduction adopting a fixed effect model and the method of difference in differences. There is literature regarding rail demand in Europe and airline industry in the United States.(Rhoades, Williams, & Green, 2006; Acutt & Dodgson, 1996; Mandel, Gaudry, & Rothengatter, 1994; Jones & Nichols, 1983; Berry, Carnall, & Spiller, 2006; Berry, 1994; Borenstein & Netz, 1999) However, only a handful of research in the literature considers the competition and substitution between closely related modes of transportation.(Wardman, 1997; Ivaldi & Vibes, 2005) Although the introduction of new product only occurred in the rail industry, it affected the entire transportation industry because other modes of transportation served as substitutes. Since consumers' demand for transportation is not restricted to one specific mode, changes in one industry may lead to changes in other transportation industries. The results of my analysis will imply the intercorrelation between modes of transportation.

After high-speed trains were introduced, other changes such as service sched-

[1]While the endogeneity problem is a common issue for the analyses on the effect of new products, the special economic circumstance in South Korean transportation industry resolves the endogeneity problems. First of all, the decision to introduce a new product, a high-speed train in this case, was made by the government. Pricing is also a common source of endogeneity, but the price being under the strong regulation of the government rather than entirely being under the discretion of the rail company also relieves a endogeneity issue from price decision process.

ule adjustment ensued. The results from the reduced form models show only the overall impact of high-speed train introduction, but they cannot disentangle the impact of high-speed train introduction itself from that of ensuing changes. In Chapter 3 of this thesis, I separately examine the impact of high-speed train introduction and that of ensuing changes in product characteristics. The focus will be on rail industry and the related impact on consumer welfare.

The possible effects of new product introduction are explored extensively in the literature. Trajtenberg (1989) proposes how to measure product innovations, and provides an example examination based on the social benefits from innovation of CT scanners. Petrin (2002) quantifies the effects of the introduction of minivans. However, many of the empirical studies of the markets with differentiated products primarily address firms' pricing strategies given the characteristics of each product and treat the market structure as being exogenous. Moreover, the effects of ensuing changes in product characteristics and product-line after new product introduction have not been discussed substantially in the empirical literature, although the corresponding theory is well-documented.[2] Berry et al. (2006); Berry and Jia (2010) also emphasize that producers might have an incentive to manipulate product characteristics other than price.[3]

In particular, a rail company in Korea might have a strong incentive to control product characteristics such as train schedules particularly because it has only limited power over pricing due to regulations. Accordingly, I will treat rail company's choice of train schedule as endogenous in all subsequent discussion, and I will instrument for it in the estimation. To study the effects of both new product introduction and the ensuing changes in product characteristics on consumer wel-

[2]Eizenberg (2011) accounts for the product-line choices after innovation.

[3]Bresnanhan also comments on Hausman (1996).

fare, I perform counterfactual analyses to separately quantify the gains resulting from introducing high-speed trains and the welfare changes resulting from the rail company's schedule adjustments.

My work adds to the existing literature by considering the impact not only on the rail industry but also on the competing modes of transportation such as domestic airline or intercity bus industries. It also takes the changes in product characteristics or the set of products offered to consumers after the new product entry into consideration, and examine how those changes affect consumer welfare. In order to take account of consumer welfare changes resulting from such adjustments, I observe the set of products offered in the Korean transportation markets before high-speed trains were introduced and I utilize the changes in my subsequent welfare analysis although I do not estimate a model of supply. Through this thesis, I emphasized the importance of considering all the related industries which are under potential influence and accounting the subsequent changes in existing products when we evaluated new product introduction.

1.1 Literature Review

In this section, I will briefly summarize the previous literature on various demand models for passenger travel as well as economic impact of new product innovation, which are closely related to my work.

1.1.1 Demand for Passenger Travel

Direct Demand Model

The *direct demand* model estimates the demand model for trips between origin and destination by a mode in a time period, using a flexible functional form of characteristics of origin and destination, and those of a given transportation mode and its competing modes. Explanatory variables commonly include land use, socio-economic, demographic factors such as the population, car ownership and income levels, which could potentially generate traveler population. They also include fare and service characteristics of the mode and its competing modes of transportation, which directly affect travelers' choice regarding means of travel. For example, a demand function for trips between i and j in time period t can be specified as

$$V_{ijt} = G_{it}^{\alpha_1} A_{jt}^{\alpha_s} \left(\Pi_p X_{ijrt}^{\delta_p} \right) \left(\Pi_q X_{ijmt}^{\gamma_q} \right)$$

where X_{ijrt} represents travel characteristics of rail and X_{ijmt} represents travel characteristics of competing mode m. G_{it} and A_{jt} represent socio-economic factors of the origin i and the attractiveness of the destination j. This model has been widely used in the United Kingdom to analyze intercity travel demand incorporating intermodal interactions.

While a conventional direct demand model adopts a constant elasticity model, Wardman (1997) amended the models to allow the elasticities to vary with the level their variables(e.g. fare) take, and with the level of competition from coach and car.

Then he estimated them using cross-sectional ticket sales data on 160 non-London flows for the years 1985-1986 and 1990-1991 with non-linear least squares. Comparing the models with a constant elasticity and those with variable elasticities, he found improved goodness of fit when the model allows variable elasticities with respect to the level of their variables and competitive effect. He also pointed out that it is important to let the elasticities vary with both the level of the variables and the competitive position since his results supported dampened elasticities for two important factors, generalized time and fare.[4]

[4]He defined "generalized time" as a composite variable which captures the principal service qual-

The advantage of this approach is that it automatically combines trip generation, distribution and mode choice decisions and the relationship between alternatives is freely estimated rather than imposed.(Small & Winston, 1999) However, the number of elasticities to be estimated will increase rapidly as the number of alternatives increases.

Direct demand models are applied to time-series data too. Gaudry (1975) employed time-series data observed monthly in Montreal to explain the aggregate demand for public transit in urban area in terms of fares, prices of non-transportation goods, service characteristics of competing modes of transportation and socio-economic variables such as income.

Disaggregated Demand: Multinomial Logit

Although aggregate models have a solid theoretical grounding and it is easy to estimate, disaggregate models that use data on individual choices usually lead to more precise estimates and more explicit description on individual's behavior. It is also more clear about the source of random disturbance in demands.(Small & Winston, 1999)

Most disaggregate models are based on a random component in a utility function. Suppose a traveler i whose utility function is expressed as:

$$U_j^i = V(X_j, Z^i, \beta) + \epsilon_j^i$$

where X_j represents characteristics of product j and Z^i represents characteristics of traveler i who make a choice a mode among J available modes of transportation. Then i will choose a mode j if and only if the utility generated by j is the largest

ity and specified it as a linear combination of journey time, inconvenience of train in terms of schedule constraints and interchange penalties.

among all the J alternatives. Hence, the probability that i will choose a mode j is:

$$P_j^i = \text{Prob}(U_j^i > U_k^i, \quad \forall j \neq k)$$

The choice probability depends on the distributional assumption on ϵ. When the random component is assumed to follow the extreme value distribution, it is particularly easy to obtain the choice probability, and it can be algebraically expressed. McFadden derives the model and it is known as "multinomial logit".

He used the introduction of BART as a natural experiment and applied this model to forecast the travel demand generated by San Francisco Bay Area commuters. He estimated the model of mode choices using data collected in 1972, before BART began operation and used to predict the share for BART usage.(Mcfadden, 1978) First, he estimated a disaggregate demand model, considering four alternatives: auto drive alone, auto shared with someone else, and bus, subdivided by access mode. The model included on-vehicle travel time, excess or out-of-vehicle time and cost divided by wage as well as mode specific dummy variables for the explanatory variables. Based on the estimated demand, he predicted the modal split after the introduction of BART, summing over the estimated choice probabilities for each alternative, with BART added to the set of available modes. For the prediction, he assumed that coefficients of alternative specific variable and interactions for BART are identical to the corresponding coefficients for the existing bus variable.[5] The prediction of total BART share produced by this model was 6.4%, which was quite close to the actual share of 6.2%. However, the model under- and over-predicted the shares for some other alternatives due to the over-prediction for the modes requiring walk access.

Multinomial logit models are useful and require low computational burden. In

[5]He pointed out that this was a weak point for this method in Mcfadden (2001).

addition, it is easy to analyze the impact of new product introduction since each alternative is assumed to be independent. However, it has limitation. In multinomial logit models, the ratio of any two choice probabilities(P_j^i/P_k^i) does not depend on utilities for any alternatives other than themselves. This property is called "Independence from Irrelevant Alternatives"(IIA). Since it implies that adding a new alternative will not affect the ratio P_j^i/P_k^i, it could be inconsistent with consumer's true behavior, particularly when the added alternative is a close substitute of j(or k). This problem can be more serious in aggregate models than in disaggregate models.(Mcfadden, 1978)

In order to avoid IIA property, McFadden developed multinomial logit by allowing idiosyncratic preferences in the model. The model is called "nested logit". It is based on "nests", a partition of the choice set.[6] In nested logit models, the random component in the utility function is correlated among products belonging to the same nest, but is assumed to be independent across nests. The model allows for more flexible substitution patterns than logit though the pattern strongly relies on how to define nests.

Morrison and Winston (1985) estimated a disaggregate model of intercity passenger transportation demand for vacation and business travelers using nested logit. They assumed that vacation travelers choose not only a mode for traveling but also a destination of the vacation. The advantage of their approach is that it is possible to analyze the impact of a change in a modal attribute on destination choice. They specified the utility of traveler taking a vacation trip if the traveler chooses a destination d, mode k and δ for a rental car, where δ takes 1 if he rents a car 0 otherwise, as a function $V_{d,k,\delta}$ of income, prices, characteristics of traveler,

[6] A set of similar alternatives is called a nest and each alternative belongs to only one nest.

mode and destination, and a random component. Then the choice probability

$$P_{d,k,\delta} = \text{Prob}(V_{d,k,\delta} > V_{d',k',\delta'} \quad \forall d', k', \delta' \neq d, k, \delta)$$

and this joint probability of choosing (d, k, δ) can be written as the product of marginal and conditional probabilities:

$$P_{d,k,\delta} = P_d \cdot P_{k|d} \cdot P_{\delta|d,k}$$

Assuming the random component is distributed as generalized extreme value, the above structure forms the basis for a nested logit model. They estimated the demand model using data drawn from the 1977 Census of Transportation National Travel Survey which provide socioeconomic data on traveling households and the characteristics of their trips. From the estimated results, they discovered consumers' preference regarding vacation travels. For example, if air fares to one city fall and those to other cities do not, then the city with decreased price is more likely to be selected as a trip destination. They also found that the coefficient on the expected value of the maximum utility obtained from the nest(log-sums or inclusive values) is placed between 0 and 1, which implies that the nested logit structure explains travelers' behavior better than the pure logit does.

Aggregate Demand from a Discrete Choice Model

Many of the recent empirical studies on transportation demand used aggregate data but their theoretical foundation is on discrete choice model unlike direct demand models. As pointed out earlier, there are too many elasticities to be estimated if we model a simple aggregated demand curve for a large number of products. In order to avoid such problem when only aggregate data is available, it is desirable to put some *a priori* restrictions on the demand problem which could reduce the number of parameters to be estimated. Such methods are appropriate for the analysis on travel demand, and it is particularly affluent in studies on the airline industry.

Although it is not applied to airline industry, the methodologies proposed in Berry (1994) and Berry, Levinsohn, and Pakes (1995) can be widely adopted to the studies on travel demand. They considered the problem of "supply-and-demand" analysis on an oligopoly market with differentiated products. Assuming that demand can be described by a discrete choice model and that prices are endogenously determined by price-setting firms, they set up a variety of demand models ranging from a simple logit to a full random coefficient model that explicitly allow for the correlation between prices and unobserved product characteristics. In their models, consumers' utility depends on observed product characteristics, unobserved product characteristics, which are known to consumers but unobserved to econometricians, and a random component. Each consumer is assumed to purchase only one unit of product which generates the highest utility among all the available products including "outside good", then product level market demand are derived as aggregate outcome of consumers' choice.[7] Therefore, it is possible to estimate demand parameters even if individual choices are not observed in the data.

In particular, the demand model whose basis is on the nested logit, is simply estimated from a following linear instrumental variables regression.(Berry, 1994)

$$\ln s_j - \ln s_0 = x_j\beta - \alpha p_j + \sigma \ln \bar{s}_{j|g} + \xi_j$$

where s_j is the market share of product j, and $s_{j|g}$ the market share of product j within a nest g. 0 represents an outside alternative. x_j contains observed character-

[7]Since each individual is limited to a single unit purchase, this model cannot be applied to some industries where consumers are free to buy multiple units. Recent studies such as Fan (2009) attempt estimating a multiple discrete choice model.

istics except its own price and p_j is the price of product j. ξ_j represents product j's unobserved product characteristics, and it is allowed to be correlated with price, thereby causing the endogeneity.

This kind of model can allow for random coefficients on product attributes. In the model with full random coefficients, the market share equation is computationally difficult to calculate unlike the above nested logit case. Since the mean utility(or equivalently unobserved product characteristics) does not have an analytical solution, the methodology requires to numerically solve for the mean utility from the model. Berry et al. (1995) proposed an estimation procedure by "inverting" the market share equation with a contraction mapping to find the mean utility and exploiting the calculated mean utility to set up the moment conditions.

Obviously, the model with full random coefficients requires heavier computational burden though it is more flexible. The full random coefficient model may be preferred when researchers want to obtain richer pattern of substitution.

The mixture model is one of the latest advances in the studies of choice models, and it is based on a probability distribution that is a convex combination of some other probability distributions. The early motivation of mixing choice models is to apply more flexible distribution to the models in order to account for heterogeneous tastes or to allow more flexible substitution patterns.[8] Although the mixture model can increase the goodness of fit, it also has its downside. There is a tendency to overfit the data, and it is impractical to estimate the parameter for the mixing distribution if number of types exceeds two.(Berry & Jia, 2010)

Berry and Jia (2010) estimated a structural model of the airline industry to analyze the impact of the various factors on the profitability of the legacy carriers. They adopted a nested logit model with a discrete type version of random coeffi-

---

[8] They are well-summarized in Walker and Ben-Akiva (2011).

cients, and estimated the model via a methodology suggested in Berry et al. (1995), treating both prices and flight frequency as endogenous variables. As in Berry et al. (2006), they considered multiple types of consumers who have different tastes for characteristics in order to capture the correlation of tastes for different product attributes.[9]

As the instruments, rival product attributes or competitiveness of the market environment are common choices. However, a market with wider price dispersion has a larger number of products since they defined a product as a group of tickets whose fares fall in a fixed bin. Instead of common choices of instruments, they used the route level characteristics including the percentage of rival routes that offer direct flights, the average distance of rival routes, the number of rival routes, or the number of all carriers. They also employed some cost shifters as another set of instruments. Since they treated flight frequency as an endogenous variable, they also needed instruments for the variable. They used the fitted value from the regression of segment departures on characteristics of the end cities.

They found that the price elasticity of air travel increased and the preference for direct flights became stronger from the comparison between 1999 and 2006. They also found that the expansion of low cost carriers, the changes in marginal cost due to fuel costs and these factors explained about 80% of the observed reduction in legacy carriers' profits. The increased competition with LCCs and increased fuel costs decreased legacy carriers' profit. Combining these results, they concluded that the change in demand is an important factor that influenced legacy carriers' losses, they concluded.

[9] In the model, they expanded the number of consumer types to R, but they assumed that there are two types in the estimated because type-specific parameters appeared to be sensitive to small changes in the model specification or instruments. Despite this issue, the discrete type version of random coefficient model has its advantage: i) the number of parameters to be estimated is fewer in the discrete version than in continuous version. ii) it does not require numerical integration to obtain market share expression.

1.1.2 Preference of Time-of-Day

Although there is a substantial amount of work done regarding demand models for passenger travel, studies on travel demand that consider consumers' preference over travel schedule are not common due to data limitation. One of common approaches that incorporate time-of-day preference into the model employs a set of dummy variables for each hour of departure time. An alternative approach assumes that each hour of departure time generates different values in travelers' via a parametric specification with a combination of sine and cosine curves.(Koppelman, Coldren, & Parker, 2008) These approaches presume that the value generated from each hour of departure time is common across travelers.

Some studies such as Koppelman et al. (2008) incorporated preference over time-of-day into air-travel itinerary using penalty function on schedule delay and the weights for each time interval within a day. However, travel demand incorporating preferences over time-of-day has been explored by engineers rather than economists, the endogeneity problems from fares or travel schedule have been rarely considered. Another reason why it is difficult to address the endogeneity from travel schedule together with preferences over time-of-day is because both fares and itineraries are usually determined by service providers and thus it is difficult to separate the effects of those two attributes in the estimation with a given data set. For example, in the airline industry, which researchers in various fields have explored, both fares and travel schedules are endogenously determined and thus closely related.[10] However, only few studies attempted to treat both fares and travel schedules as endogenous variables.[11]

---

[10]Borenstein and Netz (1999) applied spatial theory to airline studies to argue that airlines compete on departure times where the departure times of flights on a route can be interpreted as locations on a 24-hour clock.

[11]Berry and Jia (2010) attempted to treat fares and flight frequencies as endogenous, adopting a

Koppelman et al. (2008) applied a multinomial logit model with aggregate data to airlines. The authors examined air-travel itinerary share. The primary goal of their work is to assess the effectiveness of representing time of day preference by a continuous function and to analyze the effect of "schedule delay"(defined as the difference between preferred and itinerary departure time) on their trip choices. They estimated the demand for air-travel itinerary without data describing individual travelers' choices and trip characteristics, assuming that the value of itinerary is represented by itinerary characteristics. In order to model passengers' departure time preferences, they added i) hourly dummy variables, or ii) a continuous function using sine and cosine curves for time of day preference to the model, and compared the results. They also estimated a model with a penalty function for schedule delay which is non-linearly increasing. The results suggested that the models with a representation of travel time preference are superior to the models without it. They concluded that the models using sine-cosine curves are superior to others because they allow variation even within an hour range and they have fewer parameters to be estimated. However, it is not without limitation: the variation is restricted to a combination of sine-cosine curves. The limitation of these models with dummy variables or a continuous function of sine-cosine curves is that the underlying assumption includes every traveler value travel schedule in the same way.

In order to develop the models, they defined a schedule delay variable which is a weighted function of time deviations between each itinerary departure time and 15-min time periods, and transformed the variable into a penalty value which gradually increase in disutility during the one hour period before or after a desired

separate set of instrument for flight frequency variable. They regress segment departures on characteristics of the end cities, and then include the fitted value as instruments. However, their work do not allow for consumer heterogeneity over time-of-day.

departure time, and a more rapid increase between one and three hour before or after a preferred departure time. Their results suggested that the model does slightly better than and significantly rejects the base sine-cosine model. More important significance of considering schedule delay is that it describes well individual's behaviors underlying actual choices. Adopting this structure, they differentiated the effect of schedule delay between outbound and inbound passengers. They found that inbound and outbound passengers have different time of day preference, and such differentiation dramatically improves the goodness-of-fit.

Miller is one of the economists who raised the questions on the endogeneity from travel schedule. Miller (1972) presented an integer programming model of aircraft routing that maximizes welfare taking both consumer side and producer side into account. He compared the results from the criterion of maximum welfare with those from supply-oriented model. He also took notice of consumers' preference over time of day and the indivisibility problem in the production.

For the supply side model, he used Aircraft Operating Cost and Performance Report and Air Carrier Financial Statistics from Civil Aeronautics Board(CAB) Form 41 to break airline cost into 'direct operating cost' and 'indirect operating cost'. Since the data was drawn from system-wide average, he modeled the cost function for a representative city pair. For the demand side, he let the number of passengers who actually traveled as a function of the timing as well as the number of flights scheduled at the time. Then, he solved for the efficient flight schedule for a representative city pair that maximizes the sum of consumer surplus and producer surplus subject to constraints: i) the number of passengers flown to be less than or equal to the number of potential travelers, ii) the number of flights per time period for each type of aircraft to be less than the number of total equipment. Compared the results based on the supply-oriented model that minimizes cost given demand, the endogenous model put an emphasis on the demand side, particularly regarding preference over travel schedule. He concluded that the supply oriented model failed to internalize trade-offs between consumer surplus and producer surplus and the efficiency requires as much concern for demand as for cost.

1.1.3 New Good Introduction

My work can be considered in line with studies on a new product evaluation. There have been many attempts to measure qualitative improvement of new goods, and empirical studies on new product evaluation have been particularly active in the recent years. Estimating a demand system is the basis of quantifying the value of new goods, and most recent work in methodology tends to use a product space approach or a characteristic space approach.

In a product space approach, we regard a single full integrated entity as a product, while we regard a set of various characteristics as a product in a characteristic space. The estimation in a product space is computationally simple and the models tend to be flexible in terms of specification, but the number of parameters to be estimated increases rapidly as the number of products in consideration increases.[12] It is hard to forecast the demand of new products prior to their introduction. On the other hand, the models in a characteristic space avoid the dimensionality and it is useful to deal with the introduction of new products. These advantages do not come without a cost. First, it is difficult to obtain data on the relevant characteristics. Sometimes it may not be clear which characteristics constitute the product, and this may lead to the endogeneity problem due to omitted characteristics. Another disadvantage of using characteristic space can be computational burden.

One of recent work adopting a product space approach is Hausman (1996). He

[12] To avoid the dimensionality issue, we can impose more structure such as "nesting", which reduces the flexibility of the models as a trade-off.

raised the question that the current estimation of Consumer Price Index(CPI) did not take the effect of the introduction of new goods into account despite its importance. Investigating the introduction of a new cereal brand by General Mills in 1989-Apple Cinnamon Cheerios, he found that neglecting new products can result in overstated CPI compared to the true Cost-of-Living(COL) index. He used weekly cash register data, and adopted three-level Almost Ideal Demand System(AIDS) with bottom-level for each brand, middle-level for market segments such as kids, adults or family, and top-level for the overall consumption on cereal. For the endogeneity from prices, he adopted the changes in the prices of the same product in different cities. The basic idea of this instrument is that prices in one city which is affected by the cost of production are correlated with prices but uncorrelated with demand shocks. However, there is a debate on whether those instrumental variables are appropriate because national advertising or campaigns could be a source of unobserved nationwide demand shocks.[13] He found that the gain in consumer surplus was $66.8 million per year and the CPI may be overstated for cereal by about 25% because of neglecting the introduction of new cereal brands. When he took the imperfect competition into consideration, the increase in consumer welfare was smaller than in the perfect competition case, and the overstate was reduced to 20%. Based on his analysis, he emphasized the importance of considering the introduction of new products or brands when calculating CPI.

Nevo (2003) addressed the similar question using a discrete choice model in a characteristic space approach. He constructed a price index that takes the introduction of new products and quality changes into consideration and compared the results under various assumptions. His results showed that a price index can vary with assumptions regarding i) the interpretation of the changes over time in the

[13]See Bresnahan's comments on Hausman (1996).

average demand for all products, and ii) the interpretation of a change in the unobserved components in the demand equation, and the price index can range from 35% increase to a 2.4% decrease in prices over five years. He pointed out that these assumptions are not limited to the discrete choice models. Although the interpretation of time effects would not cause problem when comparing two economic outcomes, the assumption should be carefully reviewed when the time effects are linked into a price index.

Trajtenberg (1989) conceptualized the notion of product innovations and proposed an empirical measurement to assess the value of innovations. He adopted a characteristic approach to demand theory and discrete choice models, and empirically estimated the value of CT scanners. His idea to measure the value of new product is that product innovation can be regarded as a change in the set of available products if each product can be described in terms of product attributes and prices. Hence, the magnitude of innovations can be measured by

$$\Delta W = W(S_t) - W(S_{t-1})$$

where S_t is a set of available products in a time period t and $W(S_t)$ represents "social surplus" if consumers make choices among S_t.

He applied the proposed methodology to compute the value of CT scanner. First he separately estimated the nested logit model for every year from 1976 and 1981, and then took differences between computed aggregate consumer surplus which is calculated based on the estimated demand coefficients. Since the error term on the demand side follows an extreme value distribution, the surplus $W(S_t)$ is calculated using

$$W(S_t) = \frac{\ln\left(\sum \exp V_{it}\right)}{\alpha}$$

where V_{it} is the fitted value of indirect utility function from product i in period t and α is a price coefficient. He obtained a positive price sensitivity and discussed the potential causes, the correlation between prices and unobserved product characteristics. He also computed the surplus for both ex-ante and ex-post,[14] He found that the ex ante and ex post measures provide different values though the qualitative results are consistent for both measures.

As for more recent work, Petrin (2002) presented a more sophisticated process, addressing price endogeneity via a methodology proposed in Berry et al. (1995). To be specific, he proposed a technique for obtaining more precise estimates of demand and supply curves when micro data is not available, and apply the technique to estimate the economic effects of the innovation of Minivan. In his paper, he adopted the random coefficients discrete choice approach taken in Berry et al. (1995), and assumed that multiproduct firms compete in prices in Bertrand-Nash fashion. Instead of using micro-level data on purchasers, he exploited information on purchasers of new vehicles from the Consumer Expenditure Survey and related the average demographics of consumers to the characteristics of the products they purchase. This extra information on purchasers plays the same role as consumer-level data. His results suggested that the innovation of minivan resulted in the large increases of consumer surplus and the innovator, and almost half of those benefits came from increased price competition which accrued to consumers purchasing vehicles other than minivans. He also found that the models without micro data tend to yield larger welfare estimates than those with micro data because the models without micro data heavily depend on the idiosyncratic logit errors.

[14] Ex-ante surplus is calculated using the estimates in period t rather than $t+1$, i.e. $\Delta W^a = W_t(S_{t+1}) - W_t(S_t)$. Ex-post surplus is calculated using the estimates in period $t+1$, i.e. $\Delta W^p = W_{t+1}(S_{t+1}) - W_{t+1}(S_t)$.

1.2 Industry Background

In this section, I will describe the transportation industry in South Korea to provide a better understanding of consumers' behavior regarding travel methods. Although this paper focuses on the effect of high-speed train introduction, it is also important to understand other mass transit infrastructures such as inter-city buses and domestic flights because of their role as potential competitors. Thus, this section provides information on conventional modes of mass transportation available in Korea, and on the regulations imposed on the respective service providers.

1.2.1 Transportation Industry in Korea

Rail service in South Korea is provided by only one company, Korail, which leases railroads from Korea Rail Network Authority.[15] Korail handles about 20% of passenger travel, connecting about 600 train stations over more than 3500Km throughout the country.

Korail currently operates four different types of trains, categorized in terms of speed: KTX, Sae-ma-eul, Mu-gung-hwa and Tong-il. It had been operating the latter three types prior to the introduction of high-speed trains in April of 2004. KTX, the high-speed train introduced in 2004, is the fastest train type available in Korea, which makes only a few stops during its trips. It takes less than 150 minutes for Seoul-Busan route, while it takes more than 240 minutes by Sae-ma-eul trains, the second fastest train type. Sae-ma-eul is the second fastest train type, It skips small stations, but it stops at a large city in each region. It was the fastest train among rail service prior to the introduction of high-speed train, but it started

[15]Korea Rail Network Authority and Korail were governments' organizations until they were dismembered from the government. Korail and Korea Rail Network Authority become public enterprises financed by government in 2005 and in 2004 respectively. Korail is responsible for the entire rail operation and service, and Korea Rail Network Authority is responsible for the maintenance and construction.

fading out after high-speed train appeared. Mu-gung-hwa can be regarded as a "local" train, which stops even at stations in small cities. It was most common service until high-speed rail service was launched, thus it served most of the rail lines in Korea, complementing Sae-ma-eul service. Tong-il is slightly different from Sae-ma-eul and Mu-gung-hwa in that it only covers relatively short distances, and it stops more frequently than aforementioned train types. Tong-il is usually used by commuters who live in suburbs that are not reached by subways.[16]

Korail was a governmental organization until 2004, at which time it became a public enterprise financed by the government. Although it became a corporation, its general behavior such as pricing strategy did not change. According to Korail itself, the company determines fares primarily to achieve zero profit unlike private companies, and it has extremely limited power regarding its pricing. In particular, it needs the government's approval before changing fares; therefore the fare does not change frequently. Between 2001 and 2007, Korail increased prices only four times.[17] In addition, by regulation, the fare must only depend on the train type as well as travel distance, and the firm cannot set price differently for a given destination within the same day. Specifically, Korail determines a "Minimum Fare" and a "Rate per km" for each type of train, subject to government's approval, and calculates fares based on a combination of train-type and distance using the "Distance Scale Rates".[18]

[16]For these reasons, I excluded Tong-il in the dataset I used for the demand estimation.

[17]March 2002, December 2003, November 2006, July 2007. Figure 1.2 presents the nominal price of each type of trains for Seoul-Busan as well as express bus and air fares. The distance between two cities is about 400Km. The air fare fall in mid 2006 is due to the appearance of Jeju Air in the route between Seoul and Busan. Jeju Air launched the route on June 2006 and closed it on January 2007. Another route the Jeju Air operated is for Seoul-Yangyang. The routed was active between August 2006 and July 2007.

[18]It means *Fare = Greater value among Minimum Fare and (Rate per Km)·(Trip Distance)* . However, other types of price discrimination can be still offered to travelers. For example, the fares for weekdays are about 5% lower than the ones for weekends or holidays. There are also discount offers for members, students, and seniors. These discount offers remained regardless of the introduction of

This thesis takes advantage of these strict regulations on pricing. In the empirical literature, one major econometric issue is potential endogeneity bias caused by prices. That problem does not arise in this thesis, since rail pricing is under strict regulations; therefore, prices are assumed to be uncorrelated with unobserved product characteristics. In addition, since the rail fare is the same for a given destination within a day regardless of the departure time, consumers observed choices of travel schedule such as morning or evening reflected a preference based on travel schedule without being influenced by price.

Although this thesis focuses on the event occurred in the rail industry, it is still important to understand other modes of transportation in order to analyze the interrelation between those modes. In particular, intra-modal and intermodal substitutions affect the overall effect of high-speed train introduction on consumers' transportation mode choices.

Buses are the most common mode of transportation for intercity travel. Bus connection refers to inter-city buses and express intercity buses. Express inter-city buses connect two cities farther than 100Km apart and run on highways for more than 60% of the trip. There are more than 300 bus terminals throughout the country, thus travel by these bus services accounts for 70% of passenger transit in Korea. There are multiple bus companies operating on each route, and their pricing regulations resemble those of the rail company. Bus fares are also calculated using "Distance Scale Rates", and fare changes are similarly subject to the government's approval.

Domestic air travel is not as common as the other alternative modes since the Korean territory is rather compact. Air travel occupies the smallest share of pas-

---

high-speed train. Unfortunately, my data neither identifies weekend travelers from weekday travelers nor contains information on individual travelers, and thus any price discount or weekend surcharges would not be addressed throughout this paper.

senger travel among bus, air and rail, comprising only four percent of the market.[19] Only 62 out of 2107 city pairs included in my analysis are covered by airlines.[20][21]

The Korean domestic airline markets were characterized by duopolies with two legacy carriers, Korean Air and Asiana Air until the first low cost carrier(LCC) appeared in 2005. Since the first LCC in Korea, Hansung Air served only between Seoul and Jeju island, it does not yield any influence on my research, thus I excluded it from consideration. Jeju Air, the second LCC in Korea, operated on the route between Seoul and Busan from June 2006, and it affected the domestic airline business within the mainland.[22] However, Jeju Air took the market share from the legacy carriers, but it failed to increase the total air passengers.[23]

Pricing of air fare is much less restrictive than that of rail fare. Fares can be set at the discretion of airline companies as long as they provide public notice in advance. Changes in air fares are rarely observed, however.

1.2.2 Introduction of High-speed Trains

Korean High-speed train, KTX was introduced in April 2004, which may significantly affect the demand of all modes of transportation. Korean government started planning the introduction of high-speed train since 1970s, but it took more than 10 years to confirm the final plan and initiate the construction. When the construction for the dedicated rail lines began on June 1992, the government planned to com-

[19]The data set used in this thesis contains the information on 14 cities with airports: Seoul(Gimpo, Incheon), Busan(Gimhae), Yangyang, Wonju, Cheongju, Daegu, Ulsan, Pohang, Gwangju, Mokpo, Gunsan, Yeosu, Jinju(Sacheon), Jeju.

[20]If Jeju island is excluded from the dataset, 38 out of 2083 city pairs are included.

[21]There are two major airline and three low-cost carriers.

[22]See Figure 1.2.

[23]Figure 1.4 presents the monthly aggregated number of air passengers for Seoul-Busan and there is no clear changes for the time periods when Jeju Air was active for Seoul-Busan route.

plete it in January 1999, but the completion had been postponed until 2002 due to economic crisis in 1997, and postponed again until 2005. The government decided to launch new service using only a part of dedicated rail line in 2004 since it seemed impossible to complete the construction for the entire rail line in 2005.

Figure 1.1 displays two high-speed rail lines which have been available since 2004.[24] For Gyeonbu line, new rail line designated to high-speed trains were constructed, but the new rail line between Dongdaegu(in Daegu) and Busan did not open until 2010. Thus high-speed trains between Dongdaegu and Busan used the existing rail lines until the construction of the new rail line was completed. Similarly, the construction of designated rail for Honam line is still ongoing, and all the high-speed trains running on Honam line use the designated line only between Seoul and Daejeon.

Although only a section of the new rail line opened in 2004, it imposed a considerable impact on the entire transportation industry in Korea. Table 1.1 describes the market shares of each mode for the three major routes, Seoul-Daegu, Seoul-Gwangju and Seoul-Ulsan.[25] The market shares change over time and the structures of the market are different across cities. Although Seoul-Daegu and Seoul-Gwangju have similar distances, the share between rail and buses are alike. While buses are popular for Seoul-Gwangju, trains are more common for Seoul-Daegu. They had similar share for domestic airlines before high-speed train introduction, but the magnitude of impact seemed different. While the market share of domestic flights for Seoul-Daegu plunged after 2004, that for Seoul-Ulsan have remained at a

[24] Among the stations along Gyeongbu Line shown in 1.1(a), Singyeongju and Ulsan station opened in November 2010. Although Korail currently offers high-speed service for two more lines, Jeonla line and Gyeongjeon line, Singyeongju, Ulsan stations and stations along these two lines will be treated as being non-high-speed rail stations in this research because this research focuses on the first wave from high-speed rail service.

[25] The market share in this paper is defined as the proportion of travelers using each mode for a route out of the total number of travelers using flights, buses or trains for the route).

certain level.[26] Seoul-Daegu and Seoul-Gwangju routes also have different market shares across modes of transportation despite their similar distances.

Figure 1.3 displays the trend in the number of travelers using domestic flights and trains, and the vertical line indicates the timing of high-speed rail introduction. Overall, Figure 1.3 shows a considerable impact of introducing high-speed trains. The noticeable reductions in the number of air travelers after the introduction confirm competition in passenger transportation market. In particular, a surge in train usage is evident after the introduction of high-speed train. During the first three years in Figure 1.3, the number of rail travelers decreased by 30%. However, it shows an upturn after April 2004. At the same time, the number of air travelers has dropped significantly. Due to the limited number of flight routes, domestic airlines operate only in large cities, and high-speed trains were introduced in most of those cities. This supports that high-speed trains and air travel are indeed substitutes.

On the other hand, the impact of high-speed train introduction on the number of bus passengers is weaker than that on air travelers. This is likely to be due to the fact that intercity bus may compete with all types of trains, but considering the fare and trip duration, it is more likely that its main competitor is Mu-gung-hwa train. Hence the effect of the introduction of high-speed train is smaller for bus travel than domestic air travel.

These changes could be explained by the introduction of high-speed trains. However, other changes such as train schedule adjustment ensued to the introduction, and those changes could significantly affect consumers' modal choices. For example, 38 Mu-gung-hwa trains and 25 Sae-ma-eul trains departed at Seoul station for Busan in a day in 2002, but only 17 Mu-gung-hwa trains and 14 Sae-ma-eul trains departed in 2008. At the same time, high-speed trains were sched-

[26] The distance between Seoul and Ulsan is about 400Km while that between Seoul and Daegu is 300Km.

uled to depart frequently. These dramatic changes were not restricted to a few routes, but the entire rail service was rescheduled after the introduction of high-speed trains.(See Table 3.6) Hence, such changes would have caused substantial changes in consumer surplus together with high-speed train introduction itself.

Some routes without high-speed trains experienced a severe reduction in schedule frequency. Consider a station on the high-speed rail line without high-speed train connection. Due to the congestion in high-speed train stations, this station only experienced the schedule reduction without any benefits from high-speed train service. Some other routes without high-speed trains underwent opposite changes. Since the rail company adjusted train schedules for more convenient connections from/to high-speed trains, conventional trains were scheduled for the routes connected to high-speed rail lines in a more efficient way. For some routes for which Sae-ma-eul trains used to be unavailable, Sae-ma-eul trains were newly scheduled and consumers who travel for such routes would have benefited from those changes.

Observed data reflect consumers' reaction to all those subsequent changes as well as new service itself. The goal of this thesis is to analyze the overall changes in transportation utilization and to break down the changes according to the causes.

1.3 Overview

Did consumers benefit from high-speed trains in Korea? High-speed rail systems were introduced in South Korea in April 2004. These rail systems continue to significantly impact on the nation's entire transportation industry, thereby affecting its mass-transportation consumers. I observed and analyzed the changes in utilization of transportation and the differences in train schedules after high-speed rail introduction. This thesis considers firms' reactions to the introduction of new products, particularly changing product characteristics or changing the set of products offered, and analyze the effects of new products on consumer surplus, taking those reactions into account. I perform empirical analyses using Korean transportation industry data to evaluate the impact of high-speed train introduction on passenger travel.

Other researchers have theoretically considered firms' choices of product characteristics and product-lines in response to the introduction of a new product. Spence (1976) demonstrates that introducing new products may result in social inefficiency due to product choices. In his work, he illustrates two forces in the product selection under monopolistic competition. On one hand, he demonstrates that products important to social welfare could be inadvertently eliminated because revenue may not cover their costs. On the other hand, he demonstrates that the number of products will exceed the socially optimal number when a firm introduces a substitute product, which negatively affects other firms' profits in the market. He also considers the specific case of a multi-product firm. He considers the possible negative effects of launching a new product on the profits generated by the firm's other products. As a result, the firm tends to limit the number of products it offers by not introducing close substitutes for its existing products, leading to ambiguous implications regarding the introduction of new products on consumer welfare. In the context of my own work, the aforementioned findings imply that firms might choose a set of products.[27]

Gabszewicz, Shaked, Sutton, and Thisse (1986) illustrate how a monopolist would choose product quality if it can only produce a bounded number of products. Such a firm can provide optimal product lines, given a range of possible product quality, and the quality of each product may change as the range of possi-

[27]In Korea, high-speed train introduction was determined by government in 1989.

ble product quality changes. The lesson to be learned from both of these analyses is that firms can react to new product introduction by manipulating product characteristics other than prices; therefore, it is important to take changes in product selection into account when analyzing the effects of new products on consumer surplus.

The possible effects of new product introduction can be explored by reviewing the considerable amount of literature available.(Trajtenberg, 1989; Hausman, 1996; Petrin, 2002) However, many of the empirical studies of the markets with differentiated products primarily address firms' pricing strategies given the characteristics of each product and treat the market structure as being exogenous. Moreover, the effects of ensuing changes in product characteristics and product-line after new product introduction have not been discussed substantially in the empirical literature, although the corresponding theory is well-documented.[28]

In order to examine the impact of high-speed train introduction and that of ensuing changes in product characteristics separately, I adopt a multinomial logit with aggregate data with a modification to allow consumers' heterogeneous preference over travel schedule.[29] For consumer's preference over travel schedule, I employ a notion of "schedule delay" from Miller (1972), which defines it as the absolute difference between the passenger's most preferred time to travel on 24-hour clock and that of his actual time of travel.

As discussed in Section 1.1, the endogeneity from prices may result in a substantial bias. However, the special circumstance in Korean transportation industry, which is the rail company having an extremely limited power regarding its pric-

[28] Eizenberg (2011) accounts for the product-line choices after innovation.

[29] The model presented in Chapter 3 is close to Berry and Jia (2010); Koppelman et al. (2008). However, I do not estimate the proportion of each type of consumers from the model unlike Berry and Jia (2010). I differentiate my model from Koppelman et al. (2008) by allowing the travel schedule to be endogenous.

ing, alleviates the concern. Although strict regulations on pricing mitigate the endogeneity problem from prices, the endogeneity from train schedules is of concern to this research. The model treats the rail company's choice of train schedules as endogenous in order to take the firm's choices of product line into account. The detail identification strategy is described in Section 3.3.

Quantifying consumer surpluses due to the introduction of high-speed trains, I adopt the estimated demand for travel as presented in Section 3.5.1. Since the train schedules changed as a result of the introduction of high-speed trains, I separately considered the changes in consumer surplus caused by train rescheduling and those caused by high-speed train introduction. The change in consumer welfare can be measured by the difference between the expected utilities in two different situations. I primarily compared the consumers' expected utilities from the set of products offered after the introduction of high-speed trains to those from the products offered before high-speed trains were introduced.

The rest of this thesis is organized as follows. In Chapter 2, I will analyze the impact on travelers' choices regarding means of transportation from high-speed train introduction by applying a fixed effect model and the method of difference-in-differences. In Chapter 3, I will employ a standard discrete choice model for travel demand in order to estimate consumers' demand, and the estimated demand will constitute the foundation of counterfactual analyses on consumer welfare. I will summarize and conclude the thesis in Chapter 4.

In Chapter 2, I will analyze the impact on travelers' choices regarding means of transportation from high-speed train introduction by analyzing the changes in ridership of each model of transportation after the introduction of high-speed trains. To be specific, I will use a panel data on rail, bus and domestic airline industries that observes the utilization of each mode by routes(directional pair of two stations) for 83 month. A primary idea of the identification is to partition routes into the control and the treatment groups, and compare the ridership before and after the introduction of high-speed train between groups. Exploiting a large number of observations, I will let the specifications as flexible as possible. I will also provide evidence showing that the changes in ridership occurred after the introduction, not in any other periods. Based on the similar approach, I will examine the changes in the utilization of rail service by train types and discuss consumers' intramoal choices.

Although the introduction of new product only occurred in the rail industry, it affected the entire transportation industry because other modes of transportation are regarded as substitutable goods and consumers' demand for transportation is not restricted to one specific mode. Therefore the changes in one of transportation modes may lead to changes in other transportation industries. The results of my analysis will reflect how each mode of transportation substitutes and complements each other, and how train types were differentially affected by high-speed train introduction.

I found high-speed train introduction causing significant changes not only in the ridership of rail industry but also in that of the entire Korean transportation industry. In addition, the impact from high-speed train introduction was not restricted to the routes where high-speed trains have been made available. The routes without high-speed train service also have been affected, and how each route was affected depends on whether it is connected to high-speed rail service. My results show that the ridership of rail service increased 32% in the routes with high-speed trains. In contrast, the train ridership for the routes which can be partially replaced with high-speed trains, decreased by 44.4%. On contrary, the routes that are located along branch lines of a high-speed rail line attracted more passengers, thus the ridership in these routes increased by 69.6%.

The utilization of domestic flights and intercity buses also significantly changed. The domestic airline routes in which domestic airlines were directly competing with high-speed rail service, lost more than 30% of their customers. The ridership for intercity bus routes where high-speed trains have been made available also decreased more than 15%. It is noticeable that intercity buses lost their customers even in the routes which can be partially replaced with high-speed trains while domestic airlines did not.

I also found that the impact from high-speed trains varies within rail services. Each type of rail service was differentially affected. Since high-speed trains are closely substitutable to Sae-ma-eul trains, the ridership of Sae-ma-eul trains has stronger negative impact of high-speed train introduction than Mu-gung-hwa trains in the routes where Sae-ma-eul trains directly compete with high-speed trains. Therefore, Sae-ma-eul trains' ridership in those routes decreased by 31%, while Mu-gung-hwa train's ridership for the routes with high-speed train connections remained the same. Mu-gung-hwa and Sae-ma-eul trains' ridership for the routes which can be partially replaced with high-speed trains, decreased by 18% and 14% respectively. Mu-gung-hwa and Sae-ma-eul trains ridership for the routes that are located along branch lines of a high-speed rail line increased by 32% and 43% respectively.

The lesson to be learned in Chapter 2 is that the impact of high-speed train introduction was not limited to the rail industry but was significant in all the related industries. It is noteworthy that the routes without direct high-speed rail connections were also under the influence of high-speed rail service, and how each route was affected depends on whether it is connected to high-speed rail service. Although the results from those reduced form models show the significance of the impact from high-speed train introduction, they cannot disentangle the impact from high-speed train introduction itself from hat from other changes which ensued after high-speed train introduction.

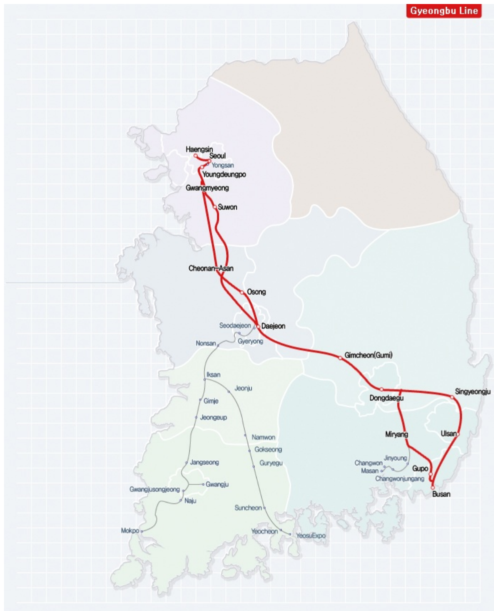
In order to examine the impact of high-speed train introduction and that of ensuing changes in product characteristics separately, in Chapter 3, I estimate a structural model of the demand for travel that incorporates consumers heterogeneous preferences over travel schedules into a standard discrete choice model. The model treats the rail company's choice of train schedules as endogenous in order to take the firm's choices of product line into account. The estimated demand will constitute the foundation for counterfactual analyses on consumer welfare. Due to the data limitation, the travel demand is estimated using data set observing only for the periods after the introduction of high-speed trains, and the counterfactual analyses project the demand for the period before the introduction using the estimated demand and information on the set of products offered to consumers before high-speed train introduction. One strong assumption for this extrapolation is that consumers' utility functions over travel remain the same after the introduction of high-speed trains.

In Chapter 3, I found that the introduction of high-speed trains caused sizable increases in consumer surplus in the Korean transportation markets where high-speed trains have been made available. However, due to the losses caused by the changes in the sets of products offered to consumers, the overall change in consumer surplus in the Korean transportation market as a whole after the introduction of high-speed train is smaller than the increases resulted from adding high-speed trains. I also found that there are significant differences in the magnitude of consumer welfare changes across heterogeneous consumers. The benefits from the new product introduction are somewhat confined to a small number of the markets, while the changes in choice set affects a broader range of consumers.

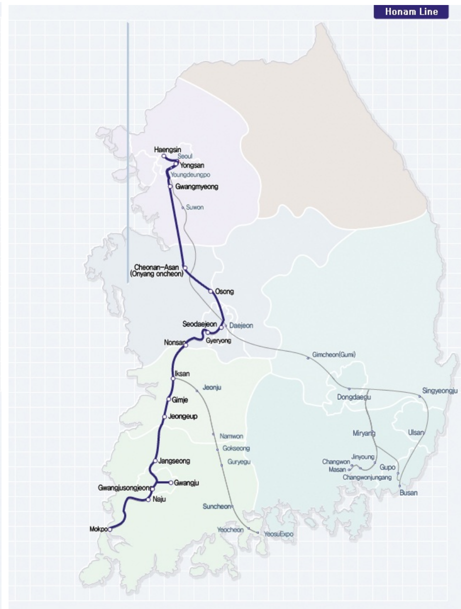
In order to examine how differentially heterogeneous consumers are affected by new product introduction, I divided the consumers into three groups based on high-speed train availability. The first group of consumers has high-speed trains in their choice set of transportation options. The second group of consumers travel between two cities, that are not connected by high-speed trains, but are located along a high-speed rail line. The third group of consumers travel between two cities at least one of which is not located along a high-speed rail line. Thus consumers in the second and the third group do not have high-speed trains in their choice set. The first two groups of consumers are expected to experience a stronger effect from introducing high-speed trains and schedule adjustment than the rest of consumers because of the mere existence of a high-speed rail line.

Each of the three groups of consumers experience different changes in the products in their choice set after the introduction of high-speed trains, which leads to variations in consumer surplus changes across those consumer groups. On the surface, consumers who had high-speed trains added to their choice set benefited as a result. However, this group endured about 50% reduction in non-high-speed trains after the introduction, which offset the gains from high-speed trains. Thus, the net gains for that consumer group are not as large as intuitively expected since the schedule changes caused substantial welfare losses and that offset 50% of the gains from having high-speed trains. Consumers who travel between two cities, that are not connected by high-speed trains but are located along a high-speed rail line, are also subjected to about 50% fewer trains. As a result, that consumer group only experienced the losses in consumer surplus. On the other hand, consumers who travel between two cities which are not located along a high-speed rail line, experienced an increased number of trains, thus a substantially increased consumer surplus. These changes in the train schedules are more noticeable than mere price changes after the high-speed train introduction, yielding more significant effects on consumer surplus than those of price changes.

Overall, the losses for consumers in the second consumer group(available high-speed rail line but no high-speed train available) outweigh the gains for the first consumer group(available high-speed train). However, the increased consumer surplus for the third consumer group(no high-speed rail line available) made up for the losses, which incidentally increased the overall consumer surplus after high-speed train introduction.



(a) Gyeongbu Line



(b) Honam Line

Figure 1.1: High-Speed Train Lines in Korea (Source: www.korail.com)

Table 1.1: Market Sahres(%); Total Passengers(10^6)

	Seoul-Daegu				Seoul-Gwangju				Seoul-Ulsan			
Year	AIR	BUS	RAIL	Total	AIR	BUS	RAIL	Total	AIR	BUS	RAIL	Total
2001	20.74	19.12	60.14	3.76	22.51	51.17	26.32	3.16	60.65	31.25	8.10	1.07
2002	20.92	20.50	58.59	3.63	21.44	52.57	25.99	3.02	59.35	31.78	8.87	1.08
2003	21.25	20.83	57.92	3.42	20.34	55.60	24.06	2.85	59.43	29.48	11.10	1.07
2004	8.18	13.55	78.27	3.98	16.35	55.23	28.42	2.84	62.59	29.70	7.70	0.98
2005	2.78	10.04	87.18	4.36	13.06	55.26	31.68	2.82	63.72	30.06	6.23	0.89
2006	1.27	8.22	90.50	5.89	10.98	49.51	39.51	3.17	63.77	30.84	5.39	0.91
2007	1.49	12.65	85.86	4.26	11.27	56.49	32.25	2.76	65.35	28.95	5.70	0.89

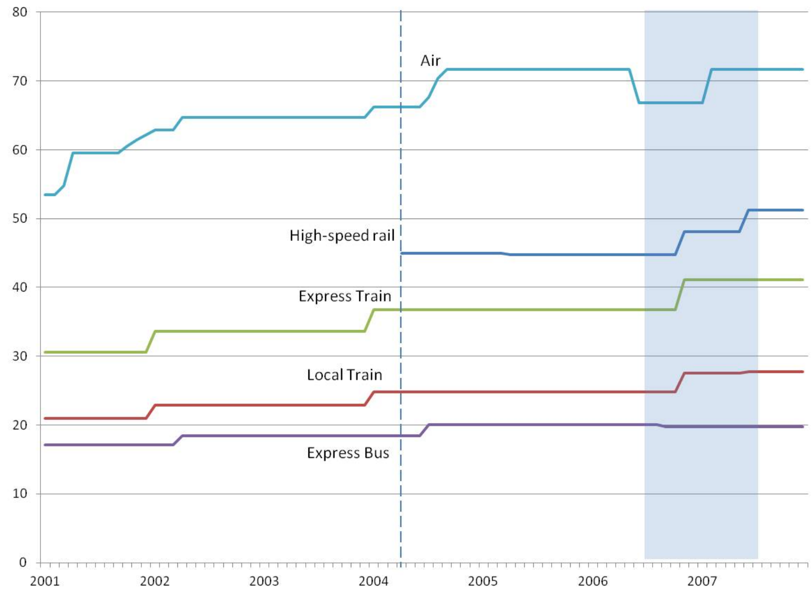
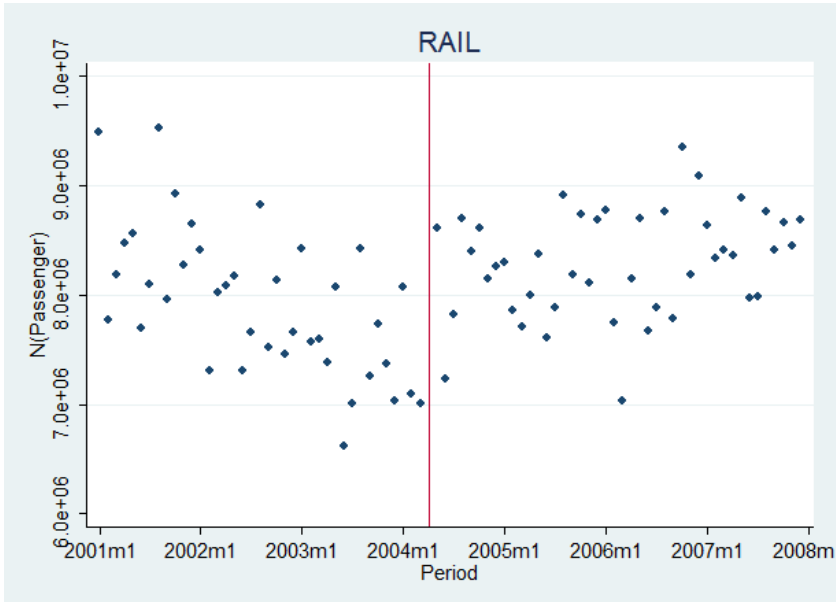


Figure 1.2: Nominal Price for Seoul-Busan(10^3 KRW): The vertical line indicates April 2004, the time of high-speed rail introduction; The shaded area indicates the periods used for the demand estimation in Chapter 3



(a) RAIL

(b) AIR

Figure 1.3: Monthly aggregated nationwide rail travelers(10^6) between 2001 and 2008: The vertical line indicates April 2004, the time of high-speed rail introduction

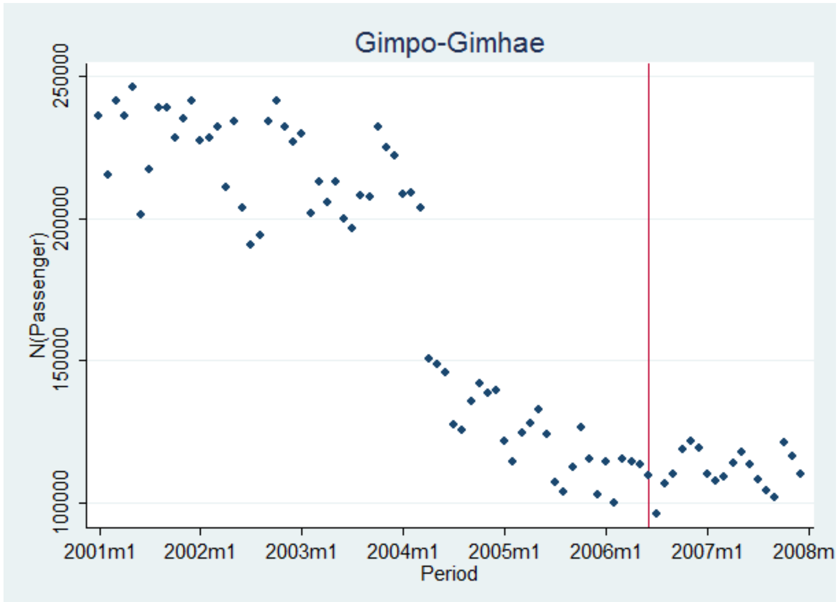


Figure 1.4: Monthly aggregated number of air passengers for Seoul(Gimpo)-Busan(Gimhae) between 2001 and 2008: The vertical line indicates June 2006, the time of Jeju Air's entry

Chapter 2

Rivalry in Transport Industry?

In this chapter, I evaluate the impact of high-speed train introduction on passenger transport. I discuss consumers' intermodal and intramodal choices, analyzing the changes in ridership of each mode of transportation after high-speed trains were introduced using a fixed effect model and difference in differences model. Although the introduction of new product only occurred in the rail industry, it affected the entire transportation industry because other modes of transportation are regarded as substitutable goods and consumers' demand for transportation is not restricted to one specific mode. Therefore the changes in one of transportation modes may lead to changes in other transportation industries. The results of my analysis will reflect how each mode of transportation substitutes and complements each other.

From the results of analyses, I found high-speed train introduction caused significant changes not only in the ridership of rail industry but also in that of the entire Korean transportation industry. In addition, the impact from high-speed train introduction was not restricted to the routes where high-speed trains have been made available. Even the routes without high-speed train service have been affected, and how each route were affected depends on whether it is connected to high-speed rail service. I also found that the impact from high-speed trains varies across different rail services. Each type of rail service was differentially affected. Since Sae-ma-eul trains are considered as a closer substitute to high-speed trains, it has stronger native impact from high-speed train introduction in the routes where Sae-ma-eul trains directly compete with high-speed trains.

The remainder of this chapter is organized as follows: Section 2.1 describes the data with summary statistics of key variables. Section 2.2 addresses the empirical strategy adopted for the analyses with required identifying assumptions, followed by the discussion on the results in Section 2.3. Summary and concluding remarks of Chapter 2 are offered in Section 2.4.

2.1 Data

This chapter provides analyses regarding intermodal choices and intramodal choices. The analysis on intermodal choices uses data from the entire transportation industries while the analysis on intermodal choices uses data on railway industry.

The main analysis for the intermodal choices employs three different sets of data. This dataset is self constructed by using raw data provided by Korail, Korea Airports Corporation(KAC), Korean Statistical Information Service(KOSIS) and Statistical Yearbook of Land, Transport & Maritime Affairs. It combines information on three different modes of transportation - domestic airlines, intercity buses and railroads. The key variable of this data is the monthly aggregated number of passengers for each unidirectional pair of stations(terminals or airports) by each mode.[1] The data set also contains the major characteristics of each route, including

[1]Unfortunately, data on intercity bus industry only recognizes two cities involved in a route. In other words, the riderships for intercity buses are observed only at a *non-directional* pair of terminals while the riderships for trains and flights are observed at a *unidirectional* pair of stations(or airports). In order to prevent from losing information regarding other modes, I assume that all the travelers

average fares, travel distances and the city where each station(terminal or airport) is located and its characteristics.[2]

The data set covers 2663 city pairs over 84 months between January 2001 and December 2007. During the month of the high-speed train introduction, all three modes of transportation underwent major changes in their service schedules. For this reason, I excluded the data observed for April 2004, the month of the high-speed train introduction. Therefore, the data set used for the analyses contains 83 months. I also excluded one of four train types, Tong-il, from the data set, because it is usually used for commuters who live in suburbs not reached by subways as discussed in Section 1.2, thus it services a different demand than this paper is concerned with.

Using this combined data set, this paper will provide travelers' intermodal choices and thus the changes in their choices after the introduction of high-speed trains. In addition to number of passengers, this paper also performs the analyses on passenger-Km and the revenue generated from each route.[3] The analyses using revenue and passenger-Km will facilitate understanding the composition of the changes after high-speed train introduction.

Tables 2.1, 2.2 and 2.3 summarize the key variables for the periods before and after high-speed train introduction by each mode. *Passenger* measures the passengers of each mode in a month t who travel for a route i. *Revenue* measures the revenue of each mode generated from a route i within a month t using real price. Tables 2.1, 2.2 and 2.3 also summarize travel distance measured in Km and travel

---

purchase round trip travels.

[2]Since there are multiple types of train service offered to consumers and multiple carriers in domestic airlines, it is not straightforward to define their fares. I use the average fares to define a fare for a pair of stations(terminals or airports) when there are more than one type of trains or more than one carriers.

[3]Passenger-Km is a unit of measurement commonly used in transportation. It is determined by multiplying the number of passengers by the average distance of their trips.

time measured in minute. The average price of airline routes is much higher than that of other two modes, and the average distance is also longer in airline industry than in other two industries. The average distance of routes is longer for intercity buses than rail services although trains and buses connect similar distance routes. It is because the data set used in this research does not fully observe all the bus connections while it does for rail and airline connections.

The data set used for the analysis of intramodal choices pertains to the South Korean railroad industry for years between 2001 and 2007 and combines three different types of information from Korea Railroad(Korail) - i) the number of train passengers for a route-train type combination aggregated monthly, ii) the major characteristics of a route-train type combination including fares and travel distances, iii) two cities involved in a route with their characteristics.[4]

As in the data set used for the intermodal choices, the data set used for intramodal choices also contains 2663 routes across 1359 city pairs over 83 months between January 2001 and December 2007.[5] Table 2.12 summarizes key variables for the periods before high-speed train introduction and those after high-speed train introduction by train type. Q_{it} measures the passengers of each mode in a month t who travel for a route i. *Revenue* and *RealRevenue* measure the revenue of each mode generated from a route i within a month t respectively using nominal price and real price. Table 2.12 also summarizes travel distance measured in Km and travel time measured in minute.

This data set is used for the analysis of travelers' intra-modal choices within rail way industry, and thus will reflect the substitute patterns between train types.

[4] A route is defined as a unidirectional pair of two stations. Available train types are described in Section 1.2.

[5] The observations for April 2004 and a train type, Tong-il are excluded for the aforementioned reason.

In addition to the ridership, passenger-Km and the revenue generated from each route reflect the pattern of consumers' intramodal choices.[6] As *Passengers* for conventional trains fell after the introduction of high-speed train, *PassengerXDistance* and *Revenue* also fell. The drop in *PassengerXDistance* and *Revenue* for Sae-ma-eul trains particularly distinctive since travelers who used to take Sae-ma-eul trains have been more strongly affected.

2.2 Empirical Strategy

Using the exogenous introduction of high-speed trains in 2004, this paper applies a differences-in-differences and a fixed effect models in order to evaluate the impact of high-speed train introduction on passenger transport, particularly mass-transit. The selection of high-speed train routes and high-speed train stations might have been endogenous because they could have been deliberately chosen by the government. However, as described in Section 1.2, the timing of the completion was exogenously determined. In other words, the impact of high-speed train introduction will be captured by comparing the mean changes of passenger travel after high-speed train introduction in the treatment group with that in the control group.

Since the definition of treatment group is unclear in this paper unlike ordinary natural experiments, it is important to classify routes into treatment group and control groups. First, all the routes connected by high-speed trains are indisputably affected by high-speed train introduction. There are other routes that could have been affected. For example, travelers who used to take a conventional train connecting from a station A to C directly, can transfer at a high-speed train station located between A and C if high-speed train service is introduced for a route be-

[6]Passenger-Km is a unit of measurement commonly used in transportation. It is determined by multiplying the number of passengers by the average distance of their trips.

tween A and B. The data used in this paper treat one complete trip connecting A and C with a transfer at B as two separate trips, a trip from A to B and another trip from B to C. Therefore, if there are many of such travelers who transfer at B in order to arrive at C, the data will present increases in number of passengers for a route from A to B, and a route from B to C while that for a direct connection from A to C decreases. In order to take those routes under the potential influence into consideration, I define a separate treatment group from the group with a direct high-speed train connection.

In order to take both direct and indirect effect from high-speed train introduction into consideration, I partition all the routes into four groups based on high-speed train availability. Group 1 consists of the routes with high-speed connections. Group 2 and Group 3 contain the routes that are not connected by high-speed trains but are under the potential impact. To be specific, the routes in Group 2 can be partially replaced with high-speed trains, thus the ridership for those routes are expected to be decreased. In contrast, the routes in Group 3 are along branch lines of a high-speed rail line, thus the ridership for these routes are expected to increase due to the travelers who want to transfer at a high-speed train station. All other routes are considered as in the control group.

Tables 2.4, 2.5 and 2.6 compare the ridership, passenger-Km for each route and revenue generated from each route during the periods before and after high-speed train introduction, presented by each group and by each mode. The first row of each panel shows the number of routes included in each group. While the number of train passengers for the routes in the control group did not change much, that for the routes in Group 1, 2 and 3 changed noticeably after high-speed train introduction. Domestic airline industry also underwent evident changes, particularly in Group 1 routes. Intercity bus industry, by contrast, did not experience as severe changes as domestic airline industry did. The number of bus passengers shown in Table 2.6 slightly decreased in Group 1 routes, but the magnitude of the decrease was not as large as in domestic airline industry.

2.2.1 Difference in Differences

Although Tables 2.4, 2.5 and 2.6 suggest preliminary findings, the comparison in those tables does not allow for other sources of variation in ridership of each mode such as regional differences. Incorporating those variations, the impact on passenger transport from high-speed train introduction can be summarized by the estimates from (2.1).

$$\ln y_{igt} = \sum_{k=1}^{3} \gamma_k TREAT_{k,igt} + \lambda_t + \delta_g + X'_{igt}\beta + \epsilon_{igt} \tag{2.1}$$

where $\ln y_{igt}$ is natural log of the outcome variable such as number of passengers for route i of group g in period t. δ_g is a time-invariant group effect and λ_t is a time period effect which is common across groups.[7] In the main specification of this paper, a month-year combination is used for time period, thus the time period effects vary over month-year combinations.[8] Let $TREAT_{k,igt}$ be a dummy variable for treatment group g and periods after the introduction of high-speed train. In other words,

$$TREAT_{k,igt} = \begin{cases} 1 & \text{if } g = k \text{ and } t \text{ is a period after the introduction} \\ 0 & \text{if } g \neq k \text{ or } t \text{ is a period before the introduction} \end{cases} \tag{2.2}$$

[7] δ_g is a time-invariant group-specific effect and this will be replaced with individual route fixed effect in Section 2.2.2.

[8] λ_t may be assumed common within all the periods after the introduction, depending on specifications.

X_{igt} is a set of characteristics of route i of group g in period t such as cities' populations, which vary across times and geographical regions. For each mode, the coefficient of interest in this paper is γ_1, γ_2 and γ_3. Under the identifying assumptions, γ_k summarizes the percent change in an outcome variable y after high-speed train introduction.

Depending specifications, the model may include a set of dummy variables that equals 1 if i is connecting a directional city pair m or dummy variables for month to make an allowance for a variation across city pairs and months. When the estimating equation contains city pair dummy variables for the airline industry, they absorb all the group-specific effects δ_g and the model becomes a fixed effect model.[9] For the rail industry, the group-specific effects will remain unless individual route fixed effects are included in the model.

The key identifying assumption of difference in differences is that trends of outcome variable would be same in both the treatment group and the control group.(Angrist & Pischke, 2009) Figure 2.1 presents similar trend in the outcome variables before high-speed train introduction and it supports the internal validity of the analyses. To check the identification strategy of the difference in differences model suggested in this paper, I add group-specific linear time trends to the list of controls. In other words, I estimate

$$\ln y_{igt} = \sum_{k=1}^{3} \gamma_k TREAT_{k,igt} + \delta_{g0} + \delta_g t + \lambda_t + X'_{igt}\beta + \epsilon_{igt} \qquad (2.3)$$

where δ_{g0} is a group-specific intercept and δ_g is a group-specific trend. Adding a group-specific linear time trend allows the control group and the treatment groups to follow different time trends, thus the results from the model with group-specific trend is more likely to be persuasive although it is limited.

[9]Data set include one route for each city pair for the airline industry.

In order to more formally see if the consequences follow the causes not vice versa, I also estimate the following model of difference in differences including l_1 lags and l_2 leads for the treatment in the spirit of Granger.

$$\ln y_{igt} = \sum_{k=1}^{3} \left(\sum_{\tau=0}^{l_1} \gamma_{k,-\tau} TREAT_{k,ig,t-\tau} + \sum_{\tau=1}^{l_2} \gamma_{k,+\tau} TREAT_{k,ig,t+\tau} \right) + \lambda_t + \delta_g + X'_{igt}\beta + \epsilon_{igt} \tag{2.4}$$

Figure 2.2 presents the estimates lags and leads relative to the one for March 2004 by groups. It indicates no effects in the six month before the high-speed train introduction while it shows drastic changes after the introduction and the effects appear to remain.[10]

The results from the causal test also support the fact that high-speed rail service launched on April 2004 is the essential cause of the changes in the utilization of transportation modes rather than some other events occurred on other months of 2004. In particular, the National Assembly cleared a legislative bill to amend the Labor Standard Act(LSA) on August 2003, and it introduced a five-day work week by reducing the maximum legal working hours from 44 to 40 per week. The law was imposed first on the public sector, financial institutions and private companies with more than a thousand employees in July 2004. Some could suspect that the changed transportation utilization is resulted from LSA rather than high-speed train introduction. However, the causality test confirms that the change occurred on April 2004.

[10] As I pointed out, the data set excludes observations for April 2004, the month of adoption.

2.2.2 Fixed Effect Model

While the difference in differences model suggested in this paper is based on the presumption of group-invariant omitted variables, the fixed effect model allows time-invariant omitted variables which may vary across individual routes within a group. In other words, the effect of high-speed train introduction is assumed to be fixed across routes within groups in the difference in differences model, but it is assumed to be fixed only across time periods for each individual route. Assuming that the individual route fixed effect appears to be linear without a time subscript and the effect of high-speed train introduction is additive and constant, the fixed effect model to be estimated is :

$$\ln y_{igt} = \sum_{k=1}^{3} \gamma_k TREAT_{k,igt} + \lambda_t + \delta_i + X'_{igt}\beta + \epsilon_{igt} \tag{2.5}$$

As I pointed out in an earlier paragraph, the data set includes one route for each city pair for the airline industry, thus the coefficients on city pair dummy variables reveal individual route fixed effects. In the intercity bus industry, only a few city pairs have more than one terminal pairs, thus the coefficients estimated from (2.1) with city pair dummy variables and without city pair dummy variables are almost same except standard errors. However, in the rail industry, a set of dummy variables for routes(a directional pair of two stations) will be added to estimate individual fixed effects since city pairs and routes(a directional pair of two stations) do not coincide.

It is also possible to incorporate group-specific time trends to the fixed effect model. Adding $\delta_g t$ to (2.5) allows each group to have different time trends with an individual specific intercept. Although it does not individual specific time trends, it could relieve the problem when the common trend assumption is violated.

2.2.3 Intramodal Choices

In order to investigate the changes in travelers' intra-modal choices after high-speed trains were introduced, I apply the same models under the same definition of treatment groups as suggested above. While the analysis on the intermodal choices focuses on the overall changes in the rail industry, I investigate how those changes estimated can be split into the changes in Sae-ma-eul and Mu-gung-hwa trains. The results will suggest the substituting patterns within the rail industry, and compare the magnitude of the impact of high-speed train introduction between train types.

Allowing other sources of variation such as fares and regional differences, (2.6) estimates the impact on passenger rail with difference in differences.

$$\begin{aligned} \ln y_{icgt} \quad = \quad & \sum_{k=1}^{3} \gamma_k^E (TREAT_{k,icgt} \cdot E_{ic}) + \sum_{k=1}^{3} \gamma_k^L TREAT_{k,icgt} \\ & + \delta_{gL} + \delta_{gE} E_{ic} + \lambda_{Lt} + \lambda_{Et} E_{ic} + \alpha E_{ic} + X'_{icgt}\beta + \epsilon_{icgt} \end{aligned} \tag{2.6}$$

where $\ln y_{icgt}$ is natural log of outcome variable such as the number of passengers for route-train type c combination i of group g in period t. E_i is a dummy variable that equals 1 if i's train type is Sae-ma-eul. δ_{gL} is an time-invariant group effect for Mu-gung-hwa trains and δ_{gE} is an time-invariant group effect for Sae-ma-eul trains relative to that for Mu-gung-hwa trains. λ_{Lt} is a time period fixed effect for Mu-gung-hwa trains which is common across groups, and λ_{Et} is a time period fixed effect for Sae-ma-eul trains relative to that for Mu-gung-hwa trains. λ_{Lt} and λ_{Et} allow Sae-ma-eul and Mu-gung-hwa trains to have different time trends. In the main specification of this paper, a month-year combination is used for time period, thus the time period effects vary over month-year combinations.[11] Let $TREAT_{k,icgt}$

[11] λ_{ct} may be assumed common within all the periods after the introduction, depending on specifications.

be a dummy for treatment group g and periods after the introduction of high-speed train. In other words,

$$TREAT_{k,icgt} = \begin{cases} 1 & \text{if } g = k \text{ and } t \text{ is a period after the introduction} \\ 0 & \text{if } g \neq k \text{ or } t \text{ is a period before the introduction} \end{cases} \tag{2.7}$$

The coefficient of interest in this paper is γ^E and γ^L where $\gamma^h = (\gamma_1^h, \gamma_2^h, \gamma_3^h)$. Under the identifying assumptions such as common trends over groups, γ_k^L summarizes the percent change in an outcome variable y for Mu-gung-hwa trains and $\gamma_k^E + \gamma_k^L$ summarizes the percent change in an outcome variable y for Sae-ma-eul trains after high-speed train introduction.[12] X_{icgt} is a set of characteristics of route-train type c combination i of group g in period t such as fare. Depending specifications, the model may include the involved cities' populations or a set of dummy variables that equals 1 if i is connecting a directional city pair m or dummy variables for month to make an allowance for a variation across city pairs and months.

To reinforce the identification strategy of the difference in differences model suggested in this paper by allowing different time trend across groups, I add group-specific linear time trends to the list of controls. In other words, I estimate

$$\begin{aligned} \ln y_{icgt} &= \sum_{k=1}^{3} \gamma_k^E (TREAT_{k,icgt} \cdot E_{ic}) + \sum_{k=1}^{3} \gamma_k^L TREAT_{k,icgt} \\ &\quad + \delta_{gL} + \delta_{gE} E_{ic} + \delta_g t + \lambda_{Lt} + \lambda_{Et} E_{ic} + \alpha E_{ic} + X'_{icgt} \beta + \epsilon_{icgt} \end{aligned} \tag{2.8}$$

where δ_g is a group-specific trend. Adding a group-specific linear time trend allows the control group and the treatment groups to follow different time trends,

[12]Therefore, γ_k^E shows the effect on Sae-ma-eul trains relative to that on Mu-gung-hwa trains.

thus the results from the model with group-specific trend is more likely to be persuasive although it is limited to the linear trend.

In addition to difference in differences model, I consider a fixed effect model which allow individual specific intercepts. In the fixed effect model, the effect of high-speed train introduction is assumed to be time-invariant but variant across individual routes. I allow for route-train type specific intercepts to make the model more flexible. As in the difference in differences model, train type specific time trends are also allowed. Assuming that the individual route fixed effect appears to be linear without a time subscript and the effect of high-speed train introduction is additive and constant, the fixed effect model to be estimated is :

$$\begin{aligned} \ln y_{icgt} &= \sum_{k=1}^{3} \gamma_k^E (TREAT_{k,icgt} \cdot E_{ic}) + \sum_{k=1}^{3} \gamma_k^L TREAT_{k,icgt} \\ &\quad + \delta_{iL} + \delta_{iE} E_{ic} + \lambda_{Lt} + \lambda_{Et} E_{ic} + \alpha E_{ic} + X'_{icgt}\beta + \epsilon_{icgt} \end{aligned} \tag{2.9}$$

It is also possible to incorporate group-specific time trends to the fixed effect model. Adding $\delta_{gt} t$ to (2.9) allows each group to have different time trends with an individual specific intercept. Although it does not individual specific time trends, it could relieve the problem when the common trend assumption is violated.

2.2.4 Limitation

A critical limitation of these results is that it shows only the overall impact of high-speed train introduction. After high-speed trains were introduced, other changes such as service schedule adjustment ensued, and the mixture of these changes resulted in the overall changes in utilization of the entire transportation industry. However, the fixed effect model or the difference in difference in this paper cannot disentangle the impact of high-speed train introduction itself from that of ensuing changes, which raises the necessity of structural model for travel demand. The discussion of the structural model will be continued in Chapter 3.

Another limitation of this model is that it is not true that the control group has not been affected by the introduction of high-speed train. In an ordinary difference in differences model, the control group is not affected by the treatment by definition, and the difference in differences can be interpreted as the effect of the treatment. However, in this paper, high-speed train introduction has affected the entire transportation industry. According to the evidence shown in Chapter 3, some routes without any high-speed train connections or potential direct and indirect influence, underwent schedule changes after high-speed train introduction. In this sense, the control group defined in this paper is atypical although I classify some routes as in “control group”. Therefore, the results from the model adopted in this paper imply the impact of high-speed train introduction on the ridership of the routes in the treatment groups relative to that in the control group.

Apart from the model, someone might suspect the endogenous selection of high-speed train stations. Since those selections were made at the government’s discretion, the selected cities might have chosen because they were expected to grow faster than others. However, Figure 2.2 presents that there was no effect in the six month before the high-speed train introduction and the dramatic changes after the introduction. It suggested the reaction in the market under the effect of high-speed train was immediate and such immediate changes in the transportation utilization were not due to gradual growth of the cities at least.

2.3 Results

2.3.1 Intermodal Choices

Table 2.7, 2.8 and 2.9 present the main findings of this paper summarized by each mode of transportation . The entries in the tables are regression coefficients from models of 2.1. *AFTER* is a dummy variable which equals to 1 if the observation is for the periods after the introduction of high-speed train. This variable appears only when time-specific effect does not enter into the model. Columns (1)-(3) of Table 2.7 present the results estimating 2.1, and Column (4) is based on 2.5. Column (5) shows the estimates from 2.3 with individual route fixed effect instead of δ_{g0}. Columns (1),(2) of Tables 2.8-2.9 are based on 2.1, and Column (3) of Tables 2.8 and (4) of Tables 2.9 estimate 2.5, and Column (4) of Tables 2.8 and (5) of Tables 2.9 show the estimates from 2.3 with individual route fixed effect instead of δ_{g0}.[13]

To be more specific, the estimates in Column (1) of Tables 2.7-2.9 is directly comparable to the simple difference-in-differences of the ridership changes without any other sources of variation. Column (2) in Table 2.7-2.9 presents the estimates by each mode from the models which allow regional differences and the variation across months, which captures seasonal variation, using a set of dummy variables for city pairs as well as months[14]. Column (3) of Tables 2.7-2.9 are based on the model controls for regional variation as well as time period specific effects. Since the dataset contain only one airline route(a directional pair of two airports) for one city pair, the model controlling for the variation across city pairs, is equivalent

[13]While each city pair has only one airline route, there are some city pairs with more than one bus route. Columns (3) of Table 2.9 present the results estimating 2.1.

[14]Since the dependent variable is aggregated at a route level and each route may include more than one type of trains or airlines, the price cannot be described by one variable. I employ the involved cities' population in log for other controls in Column (3)-(5) of Tables 2.7-2.9. I tried including the average fare in a route and the results are robust.

to fixed effect model. Unlike the airline industry, there are more than one train stations or one bus terminal in a city, and thus it is possible that more than one routes exist for one city pair. Column (4) of Table 2.7 and Table 2.9 shows the results from the fixed effect model in the rail industry that contains individual route fixed effect.[15] Column (5) of Table 2.7 and Table 2.9, and Columns (4) of Table 2.8 present the results from the fixed effect model with group-specific linear time trends.

The estimates of coefficient on $TREAT_k$ in Table 2.7 summarize the impact of high-speed train introductions on train ridership for the routes in Group k. From the results, I conclude that estimates of the coefficient on $TREAT_k$ are generally robust to different specifications. In particular, the estimated coefficients of interest are stable unless the group-specific linear trends are included in the model. The comparison between Column (3) and (4) suggests that most of individual route fixed effects are from city pairs where the route is located. From the comparison between Column (4) and (5), it is likely that the outcome variable for Group 1 and Group 3 is increasing over time relative to that for other groups since the coefficients on $TREAT_1$ and $TREAT_3$ are smaller in Column (5) of 2.7 than in Column (4).

Column (5) of Table 2.7, which is based on the fixed effect model with group-specific time trends, provides the most robust estimates. The train ridership for Group 1, which consists of the routes with direct high-speed train connection, increased by 32%. Although the decrease in ridership of Sae-ma-eul or Mu-gung-hwa trains is expected, the increase in ridership of high-speed trains outweighed the decrease. Besides the passengers who switched from conventional trains to high-speed trains, there are other sources for the increase - (i) travelers who switched from other modes of transportation including flights, buses or auto mobiles, (ii)

[15]Since there are only 20 city pairs with multiple routes, the estimates shown in Columns (3) and (4) of Table 2.9 are almost same except the standard errors.

travelers who increased their trip frequencies after the introduction of high-speed trains.[16]

The train ridership for Group 2 routes, which can be partially replaced with high-speed trains, decreased by 44.4%. Although these routes are not under direct effect of high-speed trains, they might have been affected by transfers from or to high-speed trains. On the contrary, the routes in Group 3 attracted more passengers after high-speed train introduction, thus the ridership in these routes increased by 69.6%. This increase implies that travelers would take conventional trains to arrive at high-speed train stations for transfer, and that different train classes are not only substitutes of each other but also complements. I will discuss those travelers who switched from conventional trains to high-speed trains in Section 2.3.2.

The ridership of rail service increased by more than 30%, and this increase was solely due to passengers in high-speed trains. The number of train passengers excluding those for high-speed trains did not significantly changes after the introduction of high-speed trains. Table 2.10 presents the results of the model that estimates the same equation as 2.1 and 2.5 with data excluding observations for high-speed trains. The changes of ridership for Group 2 and 3 are consistent with the findings shown in Table 2.7.

Table 2.11 summarizes the effect from high-speed train introduction on the revenue generated in each route using the same estimating equation as in Table 2.7. Since fares of intercity buses and flights in a given route do not vary significantly across service provides, the changes in the revenue are similar to the changes in the ridership. However, the revenue changes in the rail industry are different from the changes in ridership of the rail industry shown in Table 2.7 because there are multiple types of trains operating and the fares vary across train types in a route. The

[16]The effect of high-speed train introduction on domestic flights and intercity buses confirms (i).

increase in revenue generated from Group 1 routes tends to be greater than that of the ridership, since most of the increase of ridership comes from high-speed train passengers. In contrast, the decrease in revenue generated from Group 2 routes is smaller than that of the ridership in Group 2. It suggests that Sae-ma-eul trains might have lost more customers than Mu-gung-hwa trains or the decrease of ridership might have been concentrated on relatively short routes which cost less. Similarly, the increase in revenue generated from Group 3 routes is also smaller than that of the ridership in Group 3. It suggests that the increase in ridership might have been concentrated on either Mu-gung-hwa trains or relatively short routes which cost less. The discussion regarding the changes within rail industry will be continued in Section 2.3.2.

Korean intercity bus and domestic airline industries also experienced large impact after the introduction of high-speed train. The estimate of $TREAT_k$ shown in Tables 2.8-2.9 summarizes the impact of high-speed train introductions on domestic flight ridership and bus utilization for the routes in Group k.

Although the coefficient on $TREAT_k$ in those industries are also generally robust to different specifications, it is noticeable that the coefficient on $TREAT_1$ is greater in Column (4) of 2.8 than in Column (3). It might imply that the outcome variable for Group 1 is decreasing over time relative to that for other groups in the airline industry, thus the decrease due to the group-specific trends are captured as the effect of the policy until the group-specific trends are controlled. Similarly, the coefficient on $TREAT_2$ is greater in Column (4) of 2.8 than in Column (3), and it loses its significance after group-specific time trends are added to the model. This result suggests that the outcome variable for Group 2 in the airline industry is decreasing over time relative to that for other groups.

According to the estimated impact shown in Column (4) of Table 2.8, Group 1 routes, where domestic airlines directly compete with high-speed trains, lost 33.2% of their customers due the introduction of high-speed trains. Unlike trains, the domestic airline routes in Group 2 were not significantly influenced by high-speed train introduction.

In the bus industry, the coefficient on $TREAT_1$ is smaller in Column (5) of Table 2.9 after the group-specific trends are controlled. The coefficients on $TREAT_2$ is also smaller in Column (5), and it becomes significant. It is likely that there are increasing trends for Group 1 and 2 in the bus industry, thus the increase due to the trend might have made up for the fall due to high-speed train introduction when the group-specific trends are not controlled.

Although the estimated impact does not seem to be as severe as the airline industry underwent, it may be broader than it affected the airline industry. Based on the results shown in Column (5) of Table 2.9, the number of passengers for intercity buses decreased by 15.3% in Group 1 routes, where buses directly compete with high-speed trains. Bus companies did not lose as many customers as domestic airlines in Group 1 routes since buses are closer rivals to conventional trains rather than high-speed trains. However, unlike in the airline industry, the utilization of buses for the routes in Group 2, which can be partially replaced with high-speed trains, decreased by 12.7%. This result implies that consumers who take flights for the routes with direct air connections, are not willing to switch to high-speed trains if they have to transfer, but some consumers who take buses for the routes with direct bus connections, are willing to switch to high-speed trains, putting up with a transfer.

2.3.2 Intramodal Choices

Table 2.13 presents the main findings of this chapter. The entries in the table are regression coefficients from models of 2.6. Columns (1)-(3) of Table 2.13 present the results estimating 2.6. Columns (4) and (5) show the estimates from 2.9, the fixed effect model, but Column (5) differs from (4) because it allows group-specific linear trends. To be specific, the estimate in Column (1) of Table 2.13 is directly comparable to the simple difference-in-differences of the ridership changes without any other sources of variation. Column (2) presents the estimates from the models controlling for price, involved cities' population as well as seasonal and regional variations.[17] Column (3) presents the estimates from the models which additionally allow regional differences and the variation across years using dummy variables.

The estimate of coefficient on $TREAT_k(\gamma_k^L)$ in Table 2.13 summarizes the impact of high-speed train introductions on train ridership of Mu-gung-hwa trains for the routes in Group k. The estimate of coefficient on $TREAT_k \cdot E(\gamma_k^E)$ summarizes the impact of high-speed train introductions on Sae-ma-eul train's ridership relative to that on Mu-gung-hwa train's ridership for the routes in Group k. Thus, the estimates for $\gamma_k^E + \gamma_k^L$ reflects the impact of high-speed train introductions on train ridership of Sae-ma-eul trains for the routes in Group k. The rows labeled as Sae-ma-eul in Group k in Table 2.13 present the estimates with standard errors.

From the results, I conclude that estimates of γ_k^L and γ_k^E are generally robust to different specifications. However, there are noticeable changes in the coefficients on $TREAT_2 * E$ and $TREAT_3 * E$ when the fixed effect model is adopted instead of the difference in differences model. It suggests that the involved city pair is not enough to explain the variation specially for the routes in Group 2 and 3 unlike

[17] To be specific, a vector X_{icgt} contains i's price and the involved cities' population in log, the squared log price as well as their interactions in order to allow for flexible variations.

the results found in Section 2.2.4. Another change is that γ_k^L for $TREAT_k$ tends to become smaller when I control for the group-specific trends. From the comparison between Column (4) and (5), it is likely that the outcome variable for Mu-gung-hwa trains is decreasing over time relative to that for Sae-ma-eul, thus the part of the decreasing trend is captured as the effect of high-speed train introduction except in Column (5).

Column (5) of Table 2.13, which is based on the fixed effect model with group-specific time trends, provides the most robust estimates. While Mu-gung-hwa train's ridership for Group 1, which consists of the routes with direct high-speed train connection, remained the same, Sae-ma-eul train's ridership for Group 1 decreased by 31%. However, these changes are small relative to the increase in the number of passengers for high-speed trains, since the number of rail passengers for conventional trains decreased but the decrease is not statistically significant.(See Table 2.10) Thus, the number of passengers traveling for the routes in Group 1 increased by more than 30% as shown in 2.3.1 despite this decrease, and the increase was solely due to passengers in high-speed trains.

Why did the changes in Mu-gung-hwa trains' ridership decrease only insignificantly in Group 1 unlike those in Sae-ma-eul trains? The first reason would be that Sae-ma-eul trains are closer substitutes to high-speed trains. Another explanation would be found in the ensuing service frequency changes. Since the service frequency of Sae-ma-eul trains for Group 1 routes decreased noticeably after high-speed train introduction, travelers who used to take Sae-ma-eul trains had to switch to other type of trains.[18] While some of them switched to high-speed trains, some might have switched to Mu-gung-hwa trains, and thus this potential increase might have made up for the decrease due to the substitution.

[18]Table 3.6 presents the reduction in schedule frequency.

Mu-gung-hwa train's ridership and Sae-ma-eul train's ridership for Group 2 routes, which can be partially replaced with high-speed trains, decreased by 18% and 14% respectively. Although these routes are not under direct effect of high-speed trains, they might have been affected due to transfers from or to high-speed trains. A route from Seoul to Ulsan represents an example of this case. During my data periods, high-speed trains between Seoul and Ulsan were not introduced. However, I observed decrease in train ridership for this route and increase in train ridership for the route between Dongdaegu and Ulsan. This observation suggests that some of travelers for Seoul-Ulsan route transfer at Dongdaegu station where high-speed trains stop instead of taking a direct train between Seoul and Ulsan.

On the contrary, the routes in Group 3 attracted more passengers after high-speed train introduction, thus Mu-gung-hwa train's ridership and Sae-ma-eul train's ridership in these routes increased by 32% and 43% respectively. This increase implies that travelers would take conventional trains to arrive at high-speed train stations for transfer as observed in above example for the route between Dongdaegu and Ulsan.

As briefly discussed in Section 2.2.4, it might not be true that the control group has not been affected by the introduction of high-speed train. Unlike an ordinary difference in differences or fixed effect model, high-speed train introduction has affected the entire transportation and the routes even in the control group might not be free from the influence. Therefore, it would be more reasonable to interpret the estimated effects as relative to those for the routes in the control group.

2.4 Conclusion

In Chapter 2, I addressed the changes in consumers' choice regarding modes of transportation resulting from the introduction of high-speed trains. I examined the changes in ridership of each mode of transportation with a fixed effect model and a method of difference in differences using Korean transportation industry data. With this data, I estimated a reduced form model for change in ridership of each mode of transportation allowing variation across regions and years. From the analyses, I yielded the implications regarding rivarly in transportation industry. Based on the results, I discussed in detail a rich analysis of ridership changes in the entire transportation industry after the introduction of high-speed trains and the relationship between modes of transportation.

My results show that high-speed train introduction caused significant changes not only in the ridership of rail industry but also in that of the entire Korean transportation industry. In addition, the impact from high-speed train introduction was not restricted to the routes where high-speed train has been made available. Even the routes without high-speed train service have been affected, and how each route was affected depends on whether it is connected to high-speed rail service.

For the analyses, I partitioned all the routes into four groups based on high-speed train availability. Group 1 consists of the routes with high-speed connections. Group 2 and Group 3 contain the routes that are not connected by high-speed trains but are under potential influence. To be specific, the routes in Group 2 can be partially replaced with high-speed trains, thus the ridership for those routes are expected to decrease. In contrast, the routes in Group 3 are along the branch lines of a high-speed rail line, thus the ridership for these routes are expected to increase due to thed travelers who transfer at a high-speed train station. All other routes are considered as included in the control group. I found that the ridership of rail service increased 32% in the routes with high-speed trains. In contrast, the train ridership for the routes which can be partially replaced with high-speed trains, decreased by 44%. On the contrary, the routes that are located along the branch lines of a high-speed rail line attracted more passengers, thus the ridership in these routes increased by 70%.

At the same time, Korean intercity bus and domestic airline industries also experienced large impact after the introduction of high-speed trains. I found that the ridership of domestic airlines for the routes, where they directly compete with high-speed trains, lost more than 30% of their customers. These losses are concentrated on the Group 1 routes unlike in the rail industry. Intercity buses companies were also affected. The number of passengers for intercity buses for Group 1 routes decreased by 15%, which is less severe than the decreases in the airline industry. The utilization of buses for Group 2 route, which can be partially replaced with high-speed trains, decreased by 13%, while there is no evident that the airline industry lost in those routes. This result implies that consumers who take flights for the routes with direct air connections, are not willing to switch to high-speed trains if they have to transfer, but some consumers who take buses for the routes with direct bus connections, are willing to switch to high-speed trains, putting up with a transfer.

I also found that the impact from high-speed trains varies within rail services. Each type of rail service was differentially affected. Since high-speed trains are more closely substitutable to Sae-ma-eul trains, the ridership of Sae-ma-eul trains has stronger negative impact from high-speed train introduction in the routes where Sae-ma-eul trains directly compete with high-speed trains. Therefore, Sae-ma-eul trains' ridership in those routes decreased by 31%, while Mu-gung-hwa train's ridership for the routes with high-speed train connections remained the same. Mu-gung-hwa and Sae-ma-eul trains' ridership for the routes which can be partially replaced with high-speed trains, decreased by 18% and 14% respectively. Mu-gung-hwa and Sae-ma-eul trains ridership for the routes that are located along branch lines of a high-speed rail line increased by 32% and 43% respectively.

Although the results from the reduced form models show that high-speed train introduction significantly affected the entire passenger transport in Korea, it might not be true that high-speed train introduction itself brought the whole changes. Since other changes such as service schedule adjustment ensued after the introduction of high-speed trains, such ensuing changes could have affected consumers' choice too. However, the fixed effect model or the difference in difference in Chapter 2 show only the overall impact of high-speed train introduction, but they cannot disentangle the impact of high-speed train introduction itself from that of ensuing changes, which raises the necessity of structural model for travel demand. In Chapter 3, I will discuss the structural demand model and examine the impact of introducing high-speed trains on consumer welfare, taking the effect from the ensuing changes into account.

Table 2.1: Summary Statistics: Rail Industry

RAIL				
Periods Before High-Speed Train Introduction (N=103857)				
	Mean	SD	Min	Max
Passengers(Q)	2328.72	8621.39	1	191271
PassengerXDistance(10^6 Km)	0.44	2.30	1.0E-05	84.95
Revenue(10^6 KRW)	26.14	144.03	4.7E-03	5322.61
Real Revenue(10^6 KRW)	28.64	158.10	0.01	5839.40
Fare(10^3 KRW)	9.28	5.33	4.70	33.30
Real Fare(10^3 KRW)	10.16	5.82	5.26	36.64
Trip Distance(Km)	153.48	107.16	9.00	489.65
Duration(min)	130.09	89.46	7.65	489.26
Periods After High-Speed Train Introduction (N=117172)				
	Mean	SD	Min	Max
Passengers(Q)	2782.79	15459.41	1	661003
PassengerXDistance(10^6 Km)	0.50	4.82	2.1E-05	265.93
Revenue(10^6 KRW)	45.24	543.80	2.8E-03	28616.59
Real Revenue(10^6 KRW)	44.49	532.97	0.00	27783.10
Fare(10^3 KRW)	10.25	7.10	2.80	41.20
Real Fare(10^3 KRW)	10.10	6.98	2.71	40.27
Trip Distance(Km)	151.87	106.22	6.00	494.00
Duration(min)	131.17	87.29	6.00	490.35

Table 2.2: Summary Statistics: Airline Industry

AIR				
Periods Before High-Speed Train Introduction (N=936)				
	Mean	SD	Min	Max
Passengers(Q)	57520.70	76338.82	1817	365598
PassengerXDistance(10^6 Km)	21.97	32.86	5.7E-01	171.47
Revenue(10^6 KRW)	3788.40	5410.87	1.0E+02	28114.49
Real Revenue(10^6 KRW)	4144.11	5907.26	116.37	29953.64
Fare(10^3 KRW)	61.32	7.63	34.50	78.40
Real Fare(10^3 KRW)	67.07	8.08	39.55	86.96
Trip Distance(Km)	338.50	64.72	209.00	469.00
Duration(min)	49.58	6.60	35.00	60.00
Periods After High-Speed Train Introduction (N=1056)				
	Mean	SD	Min	Max
Passengers(Q)	48013.06	71011.42	164	368482
PassengerXDistance(10^6 Km)	19.01	33.07	5.0E-02	172.82
Revenue(10^6 KRW)	3610.57	5863.82	1.0E+01	30751.85
Real Revenue(10^6 KRW)	3560.66	5784.19	10.55	31264.59
Fare(10^3 KRW)	68.71	6.22	53.45	84.40
Real Fare(10^3 KRW)	67.72	6.32	54.47	86.33
Trip Distance(Km)	338.50	64.72	209.00	469.00
Duration(min)	49.58	6.60	35.00	60.00

Table 2.3: Summary Statistics: Intercity Bus Industry

BUS				
Periods Before High-Speed Train Introduction (N=8814)				
	Mean	SD	Min	Max
Passengers(Q)	13799.88	17906.90	538	125032
PassengerXDistance(10^6 Km)	3.00	4.21	1.5E-01	40.06
Revenue(10^6 KRW)	134.39	197.25	6.4E+00	6640.79
Real Revenue(10^6 KRW)	147.22	217.35	7.04	7545.49
Fare(10^3 KRW)	10.86	4.28	1.80	54.00
Real Fare(10^3 KRW)	11.88	4.69	1.96	62.20
Trip Distance(Km)	243.98	91.57	42.79	460.30
Duration(min)	178.01	62.69	40.00	359.00
Periods After High-Speed Train Introduction (N=9944)				
	Mean	SD	Min	Max
Passengers(Q)	12806.92	16362.71	532	121336
PassengerXDistance(10^6 Km)	2.73	3.76	1.5E-01	35.30
Revenue(10^6 KRW)	139.55	189.31	7.2E+00	1832.17
Real Revenue(10^6 KRW)	137.44	186.42	7.47	1782.26
Fare(10^3 KRW)	12.34	4.44	1.90	21.80
Real Fare(10^3 KRW)	12.15	4.36	1.96	21.43
Trip Distance(Km)	241.18	89.14	41.90	435.20
Duration(min)	178.01	62.69	40.00	359.00

Table 2.4: Means of Key Variables by Groups: Rail Industry

RAIL				
	Control	Group 1	Group 2	Group 3
N(Routes)	1700	148	652	163
Periods Before High-Speed Train Introduction				
Passengers(Q)	1154.17	11152.67	3200.47	3079.70
PassengerXDistance(10^6 Km)	0.18	2.47	0.69	0.31
Real Revenue(10^6 KRW)	11.04	170.91	43.84	22.29
Real Fare(10^3 KRW)	9.80	10.59	11.93	6.42
Periods After High-Speed Train Introduction				
Passengers(Q)	1279.29	21885.25	1918.56	4575.87
PassengerXDistance(10^6 Km)	0.15	5.36	0.34	0.35
Real Revenue(10^6 KRW)	9.32	566.88	22.66	24.28
Real Fare(10^3 KRW)	9.55	11.67	12.39	5.20

Table 2.5: Means of Key Variables by Groups: Airline Industry

AIR				
	Control	Group 1	Group 2	Group 3
N(Routes)	16	4	4	
Periods Before High-Speed Train Introduction				
Passengers(Q)	41492.13	138234.79	40920.92	
PassengerXDistance(10^6 Km)	17.46	48.57	13.42	
Real Revenue(10^6 KRW)	3206.71	9251.35	2786.47	
Real Fare(10^3 KRW)	67.70	64.28	67.33	
Periods After High-Speed Train Introduction				
Passengers(Q)	45346.40	73790.38	32902.38	
PassengerXDistance(10^6 Km)	19.34	25.89	10.81	
Real Revenue(10^6 KRW)	3506.81	5073.70	2263.03	
Real Fare(10^3 KRW)	68.18	65.66	67.93	

Table 2.6: Means of Key Variables by Groups: Intercity Bus Industry

BUS				
	Control	Group 1	Group 2	Group 3
N(Routes)	96	58	72	
Periods Before High-Speed Train Introduction				
Passengers(Q)	9756.27	21145.60	13273.98	
PassengerXDistance(10^6 Km)	1.78	4.81	3.17	
Real Revenue(10^6 KRW)	91.87	232.07	152.66	
Real Fare(10^3 KRW)	11.55	11.65	12.50	
Periods After High-Speed Train Introduction				
Passengers(Q)	9538.78	18824.60	12316.88	
PassengerXDistance(10^6 Km)	1.73	4.15	2.93	
Real Revenue(10^6 KRW)	88.61	206.91	146.58	
Real Fare(10^3 KRW)	11.82	11.88	12.82	

Table 2.7: Changes in Ridership: Rail Industry

	RAIL				
	(1)	(2)	(3)	(4)	(5)
$TREAT_1$	0.547***	0.566***	0.563***	0.561***	0.320**
	(0.0395)	(0.1584)	(0.1585)	(0.1543)	(0.1473)
$TREAT_2$	-0.426***	-0.424***	-0.424***	-0.425***	-0.444***
	(0.0189)	(0.0450)	(0.0450)	(0.0354)	(0.0327)
$TREAT_3$	0.895***	0.894***	0.894***	0.894***	0.696***
	(0.0388)	(0.1160)	(0.1160)	(0.1030)	(0.0952)
δ_1	2.143***	3.371***	3.372***		
	(0.0309)	(0.2792)	(0.2793)		
δ_2	0.934***	1.731***	1.732***		
	(0.0144)	(0.1086)	(0.1086)		
δ_3	0.245***	0.464*	0.464*		
	(0.0294)	(0.2493)	(0.2493)		
AFTER	0.237***	0.254***			
	(0.0094)	(0.0317)			
Constant	5.291***	-8.755**	-6.684*	-4.417	-4.522
	(0.0070)	(3.7846)	(3.8794)	(3.9418)	(3.9532)
OTHER CONTROLS	NO	YES	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES	YES
CITY PAIR DUMMY	NO	YES	YES	NO	NO
INDIVIDUAL FE	NO	NO	NO	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	NO	YES
adj. R-sq	0.106	0.672	0.673	0.910	0.910
N	221029	221029	221029	221029	221029

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

Table 2.8: Changes in Ridership: Domestic Airline Industry

	AIR			
	(1)	(2)	(3)	(4)
TREAT_1	-0.414***	-0.400***	-0.413***	-0.332***
	(0.1121)	(0.1208)	(0.1181)	(0.0615)
TREAT_2	-0.105	-0.111	-0.106	-0.008
	(0.0968)	(0.1390)	(0.1450)	(0.0509)
$delta_1$	1.917***			
	(0.0780)			
δ_2	0.869***			
	(0.0606)			
AFTER	-0.205**	-0.201*		
	(0.0806)	(0.0987)		
Constant	9.684***	-9.537	8.803	9.622
	(0.0529)	(75.3104)	(75.2047)	(76.9100)
OTHER CONTROLS	NO	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES
CITY PAIR DUMMY	NO	YES	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	YES
adj. R-sq	0.216	0.948	0.953	0.953
N	1992	1992	1992	1992

Standard errors in parentheses
* $p<0.10$ ** $p<0.05$ *** $P<0.01$

Table 2.9: Changes in Ridership: Intercity Bus Industry

	BUS				
	(1)	(2)	(3)	(4)	(5)
$TREAT_1$	-0.102***	-0.099**	-0.099**	-0.099**	-0.153***
	(0.0376)	(0.0388)	(0.0388)	(0.0398)	(0.0422)
$TREAT_2$	-0.015	-0.022	-0.022	-0.022	-0.127**
	(0.0341)	(0.0421)	(0.0422)	(0.0399)	(0.0591)
δ_1	0.692***				
	(0.0273)				
δ_2	0.238***				
	(0.0250)				
AFTER	-0.021	-0.020			
	(0.0224)	(0.0252)			
Constant	8.713***	-10.115*	-10.103*	-10.103*	-9.854
	(0.0164)	(6.0734)	(6.0920)	(6.0064)	(5.9920)
OTHER CONTROLS	NO	YES	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES	YES
CITY PAIR DUMMY	NO	YES	YES	NO	YES
INDIVIDUAL FE	NO	NO	NO	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	NO	YES
adj. R-sq	0.061	0.717	0.718	0.959	0.959
N	18758	18758	18758	18758	18758

Standard errors in parentheses
* $p<0.10$ ** $p<0.05$ *** $P<0.01$

Table 2.10: Changes in Ridership: Rail Industry Excluding High-Speed Trains

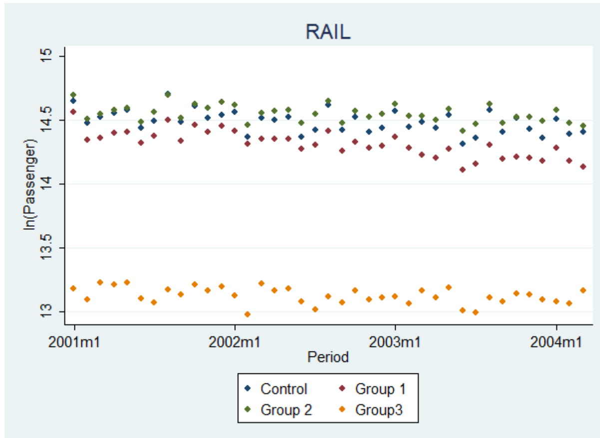
	RAIL : Excluding KTX				
	(1)	(2)	(3)	(4)	(4)
$TREAT_1$	-0.040	-0.023	-0.026	-0.029	-0.050
	(0.0369)	(0.1445)	(0.1445)	(0.1421)	(0.1361)
$TREAT_2$	-0.426***	-0.424***	-0.425***	-0.425***	-0.445***
	(0.0189)	(0.0450)	(0.0450)	(0.0354)	(0.0327)
$TREAT_3$	0.895***	0.894***	0.894***	0.894***	0.695***
	(0.0388)	(0.1160)	(0.1161)	(0.1031)	(0.0952)
δ_1	2.143***	3.308***	3.309***		
	(0.0309)	(0.2810)	(0.2810)		
δ_2	0.934***	1.720***	1.720***		
	(0.0144)	(0.1092)	(0.1092)		
δ_3	0.245***	0.445*	0.445*		
	(0.0294)	(0.2498)	(0.2498)		
AFTER	0.237***	0.252***			
	(0.0094)	(0.0317)			
Constant	5.291***	-7.439*	-5.226	-1.887	-1.794
	(0.0070)	(3.8184)	(3.9199)	(3.9054)	(3.9129)
OTHER CONTROLS	NO	YES	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES	YES
CITY PAIR DUMMY	NO	YES	YES	NO	NO
INDIVIDUAL FE	NO	NO	NO	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	NO	YES
adj. R-sq	0.089	0.663	0.664	0.910	0.910
N	221029	221029	221029	221029	221029

Standard errors in parentheses
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

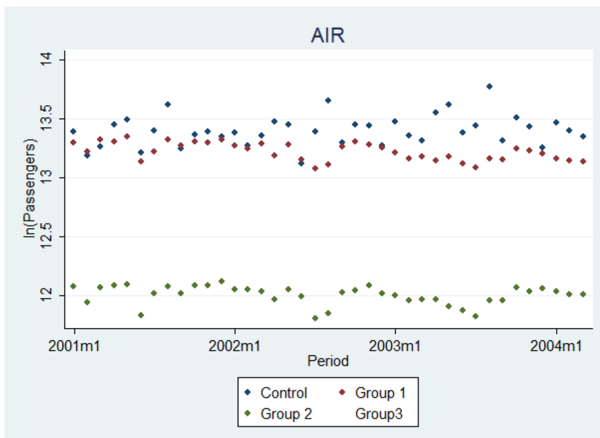
Table 2.11: Changes in Revenue: Rail Industry

	RAIL				
	(1)	(2)	(3)	(4)	(5)
$TREAT_1$	0.727***	0.752***	0.750***	0.751***	0.468***
	(0.0446)	(0.1698)	(0.1699)	(0.1649)	(0.1585)
$TREAT_2$	-0.295***	-0.293***	-0.293***	-0.293***	-0.313***
	(0.0206)	(0.0378)	(0.0378)	(0.0312)	(0.0292)
$TREAT_3$	0.699***	0.696***	0.697***	0.696***	0.488***
	(0.0396)	(0.1104)	(0.1105)	(0.1002)	(0.0935)
δ_1	2.144***	3.442***	3.442***		
	(0.0326)	(0.2729)	(0.2730)		
δ_2	1.119***	1.715***	1.715***		
	(0.0158)	(0.1064)	(0.1064)		
δ_3	-0.080***	0.616**	0.616**		
	(0.0300)	(0.2506)	(0.2507)		
AFTER	0.077***	0.098***			
	(0.0098)	(0.0247)			
Constant	0.507***	-18.438***	-16.917***	-16.980***	-17.117***
	(0.0075)	(3.1122)	(3.1740)	(3.5434)	(3.5548)
OTHER CONTROLS	NO	YES	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES	YES
CITY PAIR DUMMY	NO	YES	YES	NO	NO
INDIVIDUAL FE	NO	NO	NO	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	NO	YES
adj. R-sq	0.106	0.719	0.720	0.929	0.929
N	221029	221029	221029	221029	221029

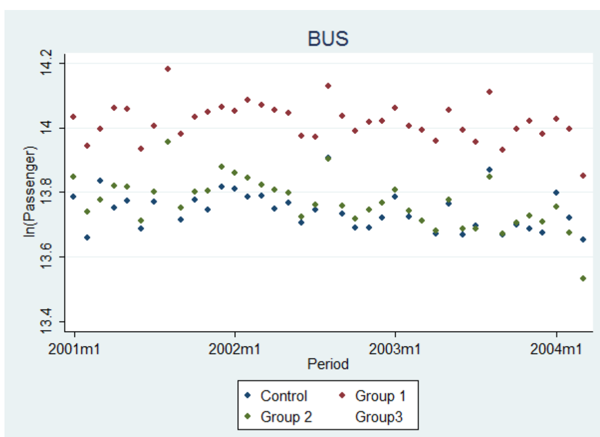
Standard errors in parentheses
* p¡0.10 ** p¡0.05 *** P¡0.01



(a) RAIL

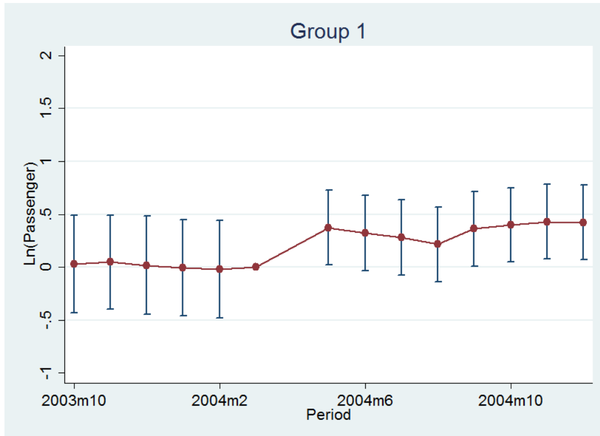


(b) AIR



(c) BUS

Figure 2.1: Trend of Ridership by Groups of Routes



(a) Group 1

(b) Group 2

(c) Group 3

Figure 2.2: Coefficients from 2.4 on lags and leads in the rail industry: Relative to March 2004(Vertical lines mark two standard errors)

Table 2.12: Summary Statistics by Train Types

	KTX		Sae-ma-eul		Mu-gung-hwa	
Periods Before High-Speed Train Intro			(N=29074)		(N=102746)	
	Mean	SD	Mean	SD	Mean	SD
Passengers(Q)			1495.82	6452.89	1930.63	6399.39
PassengerXDistance(10^6 Km)			0.44	2.37	0.32	1.19
Revenue(10^6 KRW)			32.30	174.01	17.29	61.14
Real Revenue(10^6 KRW)			35.36	190.84	18.94	67.16
Fare(10^3 KRW)			14.21	7.70	8.58	4.63
Real Fare(10^3 KRW)			15.54	8.40	9.39	5.05
Trip Distance(Km)			174.93	121.09	152.91	106.96
Duration(min)			123.48	83.21	132.73	91.92
Periods After High-Speed Train Intro	(N=5954)		(N=48694)		(N=116133)	
	Mean	SD	Mean	SD	Mean	SD
Passengers(Q)	17524.87	51494.42	716.43	1683.72	1608.82	4605.63
PassengerXDistance(10^6 Km)	4.89	17.21	0.14	0.41	0.19	0.50
Revenue(10^6 KRW)	566.21	2017.53	12.61	35.05	11.33	28.90
Real Revenue(10^6 KRW)	555.75	1978.38	12.46	34.89	11.17	28.60
Fare(10^3 KRW)	15.16	10.21	15.62	9.33	8.94	5.95
Real Fare(10^3 KRW)	14.92	10.04	15.38	9.16	8.80	5.85
Trip Distance(Km)	135.77	98.95	171.86	117.15	150.52	105.48
Duration(min)	68.93	43.07	126.95	82.44	134.25	89.82

Table 2.13: Changes in Ridership: Within Rail Industry

	(1)	(2)	(3)	(4)	(5)
$TREAT_1$*E	0.075	0.086	-0.075	-0.356**	-0.356**
	(0.0562)	(0.1788)	(0.1922)	(0.1712)	(0.1713)
$TREAT_2$*E	-0.202***	0.209**	0.050	0.042	0.041
	(0.0379)	(0.0944)	(0.1102)	(0.1056)	(0.1056)
$TREAT_3$*E	-0.309***	0.127	-0.027	0.116	0.116
	(0.0801)	(0.2723)	(0.2799)	(0.1834)	(0.1835)
$TREAT_1$	-0.044	-0.026	0.000	-0.019	0.043
	(0.0354)	(0.1371)	(0.1376)	(0.1367)	(0.1373)
$TREAT_2$	-0.416***	-0.206***	-0.173***	-0.226***	-0.181***
	(0.0189)	(0.0395)	(0.0393)	(0.0319)	(0.0301)
$TREAT_3$	0.932***	0.527***	0.546***	0.563***	0.317***
	(0.0385)	(0.1256)	(0.1263)	(0.1086)	(0.1089)
Constant	5.252***	51.297***	51.706***	-13.721*	-13.093
	(0.0070)	(8.4488)	(8.6864)	(8.1511)	(8.1466)
Sae-ma-eul in Group 1	0.031	0.060	-0.075	-0.375***	-0.313***
	(0.0440)	(0.1100)	(0.1250)	(0.1050)	(0.1000)
Sae-ma-eul in Group 2	-0.618***	0.003	-0.124	-0.183*	-0.140
	(0.0330)	(0.0980)	(0.1110)	(0.1010)	(0.1010)
Sae-ma-eul in Group 3	0.623***	0.654***	0.519**	0.680***	0.433***
	(0.0700)	(0.2270)	(0.2330)	(0.1480)	(0.1440)
OTHER CONTROLS	NO	YES	YES	YES	YES
TIME DUMMY	NO	NO	YES	YES	YES
CITY PAIR DUMMY	NO	YES	YES	NO	NO
INDIVIDUAL FE	NO	NO	NO	YES	YES
GRP-SPECIFIC TREND	NO	NO	NO	NO	YES
adj. R-sq	0.097	0.654	0.655	0.915	0.915
N	296647	296647	296647	296647	296647

Standard errors in parentheses
* $p<0.10$ ** $p<0.05$ *** $p<0.01$

Chapter 3

Did Consumers Benefit from High Speed Trains in Korea?

Generally speaking, introducing an additional differentiated product to a market benefits consumers due to the increased number of alternatives if everything else such as price remains same. However, the effect on consumer welfare is not so simple if producers also change other products characteristics and the set of other products offered. This paper considers firms' reactions to the introduction of new products, particularly changing product characteristics or changing the set of products offered, and analyze the effects of new products on consumer surplus, taking those reactions into account. The goal of my analysis is to investigate changes in consumer welfare due to the introduction of a new product, based on available Korean transportation industry data. Specifically, this paper decomposes the effects of high-speed train introduction into the gains or losses attributable to having high-speed trains and those attributable to firms' choices of products to offer across different types of consumers.

Did consumers benefit from high-speed trains in Korea? High-speed rail systems were introduced in South Korea in April 2004. These rail systems continue to significantly impact the nation's entire transportation industry, thereby affecting its mass-transportation consumers, which has motivated this paper. I observed and analyzed differences in train schedules and train availability after high-speed rail introduction, both of which affect the alternatives available to consumers. Heterogeneity across consumers is also an important factor in the analysis of consumer welfare because consumers might be differentially affected by the newly-introduced high-speed trains. Two dimensions of heterogeneity were relevant to my analysis: preferences regarding travel schedule and the choice sets available to consumers.

Other researchers have theoretically considered firms' choices of product characteristics and product-lines in response to the introduction of a new product. Spence (1976) demonstrates that introducing new products may result in social inefficiency due to product choices. In his work, he illustrates two forces in the product selection under monopolistic competition. On one hand, he demonstrates that products important to social welfare could be inadvertently eliminated because revenue may not cover their costs. On the other hand, he demonstrates that the number of products will exceed the socially optimal number when a firm introduces a substitute product, which negatively affects other firms' profits in the market. He also considers the specific case of a multi-product firm. He considers the possible negative effects of launching a new product on the profits generated by the firm's other products. As a result, the firm tends to limit the number of products it offers by not introducing close substitutes for its existing products, leading to ambiguous implications regarding the introduction of new products on consumer welfare. In the context of my own work, the aforementioned findings imply that firms might choose a set of products.[1]

Gabszewicz et al. (1986) illustrate how a monopolist would choose product quality if it can only produce a bounded number of products. Such a firm can provide optimal product lines, given a range of possible product quality, and the quality of each product may change as the range of possible product quality changes. The lesson to be learned from both of these analyses is that firms can react to new product introduction by manipulating product characteristics other than prices; therefore, it is important to take changes in product selection into account when analyzing the effects of new products on consumer surplus.

The possible effects of new product introduction can be explored by reviewing the considerable amount of literature available. Trajtenberg (1989) proposes how to measure product innovations, and he provides an example examination based on the social benefits from innovation of CT scanners. Petrin (2002) quantifies the effects of the introduction of the minivan. However, many of the empirical studies of the markets with differentiated products primarily address firms' pricing strategies given the characteristics of each product and treat the market structure as being exogenous. Moreover, the effects of ensuing changes in product characteristics and product-line after new product introduction have not been discussed substantially in the empirical literature, although the corresponding theory is well-documented.[2] Berry et al. (2006); Berry and Jia (2010) also emphasize that producers might have an incentive to manipulate product characteristics other than price.[3] In particular, a rail company in Korea might have a strong incentive to control product characteristics such as train schedules particularly since by regulation it has only limited power over pricing. Accordingly, I will treat rail company's

[1] In Korea, high-speed train introduction was determined by government in 1989.

[2] Eizenberg (2011) accounts for the product-line choices after innovation.

[3] Bresnanhan also comments on Hausman (1996).

choice of train schedule as endogenous in all subsequent discussion, and I will instrument for it in the estimation.

To study the effects of both new product introduction and the ensuing changes in product characteristics on consumer welfare, I performed counterfactual analyses to separately quantify the gains resulting from introducing high-speed trains and the welfare changes resulting from the rail company's schedule adjustments. My work adds to the existing literature by considering the changes in product characteristics or the set of products offered to consumers after new product introduction, and by investigating how those changes affect consumer welfare. In order to take into account consumer welfare changes resulting from such adjustments, I observe the set of products offered in the Korean transportation markets before high-speed trains were introduced and I utilize the changes in my subsequent welfare analysis although I do not estimate a model of supply.

Estimation of consumers' demand for travel is necessary for my examination of the impact of introducing high-speed trains on consumer welfare based on the counterfactual analyses. In order to consider travelers' heterogeneous preferences regarding travel schedules along side the rail company's schedule changes, I estimate consumers' demand for travel by explicitly incorporating preference heterogeneity into an otherwise standard discrete choice model.(Koppelman, 2006) Heterogeneity is captured in my model through a modification of the concept of "Schedule Delay" suggested in Miller (1972) and Douglas and Miller (1974). Although preference over travel schedule is an essential factor in travel demand, there has been limited modeling of it in the past due to data constraints. Some research that analyzes travel demand such as Koppelman et al. (2008) models departure time preferences, but in general they consider neither potential endogeneity from the schedules, nor heterogeneity of preferences over travel schedule across consumers.

As a result of the research I will present in the remainder this paper, I found that the introduction of high-speed trains caused sizable increases in consumer surplus in the Korean transportation markets where high-speed trains have been made available. However, due to the losses caused by the changes in the sets of products offered to consumers, the overall change in consumer surplus in the Korean transportation market as a whole after the introduction of high-speed train is smaller than the increases resulted from adding high-speed trains. I also found that there are significant differences in the magnitude of consumer welfare changes across heterogeneous consumers. The benefits from the new product introduction are somewhat confined to a small number of the markets, while the changes in choice set affects a broader range of consumers.

In order to examine how differentially heterogeneous consumers are affected by new product introduction, I divided the consumers into three groups based on high-speed train availability. The first group of consumers has high-speed trains in their choice set of transportation options. The second group of consumers travel between two cities, that are not connected by high-speed trains, but are located along a high-speed rail line. The third group of consumers travel between two cities at least one of which is not located along a high-speed rail line. Thus consumers in the second and the third group do not have high-speed trains in their choice set. The first two groups of consumers are expected to experience a stronger effect from introducing high-speed trains and schedule adjustment than the rest of consumers because of the mere existence of a high-speed rail line.

Each of the three groups of consumers experience different changes in the products in their choice set after the introduction of high-speed trains, which leads to variations in consumer surplus changes across those consumer groups. On the surface, consumers who had high-speed trains added to their choice set benefited as a result. However, this group endured about 50% fewer non-high-speed trains after the introduction, which offset the gains from high-speed trains. Thus, the net gains for that consumer group are not as large as intuitively expected since the schedule changes caused substantial welfare losses and that offset 50% of the gains from having high-speed trains. Consumers who travel between two cities, that are not connected by high-speed trains but are located along a high-speed rail line, are also subjected to about 50% fewer trains. As a result, that consumer group only experienced the losses in consumer surplus. On the other hand, consumers who travel between two cities which are not located along a high-speed rail line, experienced an increased number of trains, thus a substantially increased consumer surplus. These changes in the train schedules are more noticeable than mere price changes after the high-speed train introduction, yielding more significant effects on consumer surplus than those of price changes.

Overall, the losses for consumers in the second consumer group(available high-speed rail line but no high-speed train available) outweigh the gains for the first consumer group(available high-speed train). However, the increased consumer surplus for the third consumer group(no high-speed rail line available) made up for the losses, which incidentally increased the overall consumer surplus after high-speed train introduction.

The remainder of this chapter is organized as follows. In Chapter 3, I estimates travel demand treating train schedule as an endogenous variable. Section 3.1 describes the data used. Section 3.2 and Section 3.3 presents the model, the estimation procedure and the assumptions imposed. Section 3.4 addresses the procedure to calculate consumer welfare, followed by the discussion on the results in Section 3.5. Summary and concluding remarks of this chapter are offered in Section 3.6.

3.1 Data

The main analysis employs three different sets of data. This dataset is self constructed by using raw data provided by Korail, Korea Airports Corporation(KAC), Korean Statistical Information Service(KOSIS) and Statistical Yearbook of Land, Transport & Maritime Affairs. The first data set pertains to the South Korean railroad industry and consists of market shares and product characteristics for years 2006 and 2007. The second data set includes the market size and the market share of outside alternatives. These two data sets, used in the demand estimation, only contain observations during the period after the introduction of high-speed trains. The third data set contains characteristics of products offered to travelers in 2002, when high-speed trains were not available. This data set is used for the calculation of traveler's surplus as well as for performing counterfactual analysis.

The first dataset pertaining to the railroad industry combines three different types of information from Korea Railroad(Korail) - i) the number of train passengers for each route(defined as a directional pair of stations) by train type and departure time of day aggregated monthly, ii) the major characteristics of each route, including fares, travel distances, and distance from a station to a city-center, iii) the train schedules with train types, routes, departure times and arrival times. In all, the data set covers 6,456 routes throughout the country in existence during the time period of the data, and it contains the monthly aggregate numbers of train passengers for each route by train type and departure hour of day, observed for 12 months between July 2006 and June 2007. This data set also contains major characteristics of the route-train type combinations, such as fares, travel distance, distance from a station to a city-center, that are key variables for demand estimation.

Lastly, the schedule data provides for each train identified by a train identification number, the stations at which stops are made, train type, and departure and arrival times. The ideal data set for my research would include the numbers of train passengers aggregated for each train and for each route to facilitate more robust cross-referencing with the schedule of train services.[4] Unfortunately, the available data only summarizes counts by train type and the hour of the departure time; therefore, to infer a train-level data set I imposed an assumption on the distribution of train passengers over trains departing within an hour. Each train for each route within a given hour, is assumed to have the same number of passengers. Using this assumption, the unit of observation for the combined data is a single train, identified by its train ID number, running on a specific route over a month. Therefore, my analysis treats a train running on a route A and a train running on a route B as different observations even if train ID number is the same.

The second dataset contains the market size and market share of the outside alternatives. "The market" as used herein, is defined as a one-way travel choice from an origin to a destination city hence, I treat a directional city pair and month combination as a separate market. "Travel choice" refers to traveling by rail, bus, car or domestic flight or choosing to forego travel. Potential travelers were estimated rather than observed, however, by assuming that the number of potential travelers is proportional to the geometric average of the populations of the two respective cities involved in a route.(Berry et al., 2006) An alternative assumption would be that travelers cannot forego travel, but the increase in rail travel shown in Figure 1.3(a) suggests that this assumption is not plausible. I will discuss this alternative assumption in 3.2.7 and the results in 3.5.3.

Table 3.1 provides the definition of the variables, and Table 3.2 summarizes the data used in the demand estimation of this paper, which combines the first and the second data. It contains 392,459 products(station pair and trainID combina-

[4] It is possible that there are multiple trains departing and arriving within a given hour.

tions) over 1,114 directional city pairs and 12 months, thus the number of combinations of city-pair and month, which is recognized as a market, is 13,347. I excluded one of four train types, Tong-il, from the first dataset, because it is usually used for commuters who live in suburbs not reached by subways as discussed in Section 1.2, thus it services a different demand than this paper is concerned with. On average, 182 passengers travel on a train-route combination over a month period. N(Own Type Train/Day), N(Other Type Train/Day) and Station-City Centers are the variables used to capture the convenience of each route. N(Own Type Train/Day) counts a single type of train scheduled for a particular route within a day. N(Other Type Train/Day) similarly counts the other types of trains. Distance from city-center for a given route is defined as the sum of the distances between the departure and arrival stations from their respective city centers. This variable is meant to capture how conveniently located departure and arrival stations are in terms of in terms of intra-city transportation.

As pointed out in Section 1.2, there are four different train types that are differentiated in terms of speed, fare and facilities. Table 3.3 summarizes the data by train type. Out of 392,459 observations, KTX trains are 3.9%. Sae-ma-eul and Mu-gung-hwa trains are 12.3%, and 83.7% respectively.

The price variations within a market primarily come from price differences across train types and from routings, since fares for each route-train type combination does not vary within a day or between markets due to the distance scale rates system. Another source of price variation is nominal rail fare changes, which were observed twice in my data period.

A third dataset is employed to compare traveler's surplus before and after the introduction of high-speed trains. It contains information on the products offered to consumers before high-speed rail was inaugurated. Table 3.6 compares the number of products offered in 2006 with that in 2002 by train type. Each panel summarizes a specific type of train. The first row of each panel shows the number of city pairs for which the given train type is available, and the next three rows show the mean, median and standard deviation. Each column of the panels summarizes a separate group of markets. In order to compare the train frequencies in 2006 to those of 2002, I partitioned markets into three groups based on high-speed train availability and location. Group 1, containing the city pairs with high-speed connections, is summarized in Columns (1) and (4). Group 2, containing the city pairs which are located along a high-speed rail line but are not connected by a high-speed train is summarized in Columns (2) and (5). The city pairs that belong to Group 3, which are not on a high-speed rail line(thus, not connected by high-speed trains), are summarized in Columns (3) and (6).

Each group has been differentially affected by the introduction of high-speed trains. The numbers of Mu-gung-hwa trains offered to Group 1 and Group 2 markets in 2006, were significantly lower than in 2002, while the numbers of Mu-gung-hwa trains offered to markets of Group 3 did not substantially decrease. The panel for Sae-ma-eul, reveals two distinctive patterns. First, the numbers of Sae-ma-eul trains offered to markets of Group 1 and Group 2 in 2006, also decreased compared to those in 2002.[5] This change was caused by major reductions in the number of train scheduled for the routes along high-speed rail lines. Second, the panel also reveals that the number of city pairs where Sae-ma-eul trains are available increased from 127 city pairs to 260. This increase occurred because Sae-ma-eul trains stop more frequently and therefore became available in the cities where these additional stops are made. Group 3 experienced only relatively minor changes. In that group,

[5]The number of Sae-ma-eul trains offered to the city pairs of column (2) in 2006 is understated because the number of city pairs where Sae-ma-eul is larger than the one in 2002. However, it still significantly decreased, compared to the one in 2002. The average number of Sae-ma-eul trains offered in 2006 to the 127 city pairs where Sae-ma-eul trains have been available since 2002, is 11.

Mu-gung-hwa trains became available bewteen more city pairs despite the average number of Mu-gung-hwa trains slightly decreasing in the group of markets. The average number of Sae-ma-eul trains increased slightly for Group 3.

3.2 Model on Empirical Demand

In order to evaluate consumer surpluses resulting from the introduction of high-speed trains, one must analyze the demand that describes how travelers choose a means of transportation, taking into account their preferred travel schedules. I estimated the demand for travel using a discrete choice model, that has been used effectively in the past.(see Berry (1994); Berry et al. (2006); Koppelman et al. (2008); Berry and Jia (2010).) I also extended the standard multinomial logit model by allowing for heterogeneous travel schedules among consumers.

3.2.1 Notions of Markets and Products

This section describes in detail markets and products as I've conceptualized them for this research.

A "market", as used in this paper, is defined as a unidirectional travel from an origin city to destination city. Each unique market is identified by a unidirectional city pair and a month. Each market has own set of products offered.

A "product" is defined as a specific train operating for a specific route(a unidirectional pair of two stations) within a specific market. Each train, which is identified by a unique ID number, runs from a start-node station to an end-node station, with additional stops made during the trip. This definition of a product therefore implies that a single train connecting cities A, B and C is treated as a different product for the two connections it makes(A to B and B to C) because it operates for two distinct routes. Since each city may have more than one station, multiple routes could exist, hence even within a given market, consumers face product choice depending on route preferences. It implies that a train running route 1(station A_1 in City 1 to station A_2 in City 2) and the same train running route 2(station B_1 in City 1 to station B_2 in City 2) are treated as different products even if both trains have the same train ID number and both routes are engaged to the same market.

In reality, travelers can transfer from one train to another or change modes of transportation over the course of a single trip. I avoid this problem by defining a product as a combination of a route and a train ID rather than as the complete trip an individual traveler conceptualizes. A single rail trip is therefore a series of products, as defined above, in that a traveler may take different trains for each section of his trip.[6] The dataset provided by Korail(dataset 1) does not contain information regarding individual passengers' itineraries; therefore, that data could not support a trip-based analysis. This is the primary reason I chose to define products as I did.

The characteristics of each product are inherited from the respective product's train and that train's routing. The characteristics of a train are its type, fare, traveled distance, and schedule. The characteristics of a train's routing include distance from station to the city-center, and the number of trains scheduled for the route within a day. Those characteristics of a train's routing attempts to explain the convenience of each route in terms of intra-city transportation.

3.2.2 Notion of Schedule Delay

This paper attempts to explicitly incorporate traveler's heterogeneous preference on travel schedule. Because the fares for a given product, do not vary within a single day, I have assumed that travelers' schedule choices are based entirely on

[6]For example, a traveler may take train 1 from A0 to A5 and transfer to train 2 at A5 to arrive at B7. Then, the product that the traveler purchases is $\{$(A0$\rightarrow A5$,train 1),(A5$\rightarrow B7$,train 2) $\}$.

the schedule themselves. This ignores, however, that travelers might need to travel at times other than those they prefer due to train availability. Douglas and Miller (1974) suggest two reasons why people cannot travel at their preferred times: the difference between a traveler's desired departure time and the closest scheduled departure; and delays due to excess demand during a traveler's preferred travel time. This paper focuses more on the first source of compromise, which was referred to as *frequency delay* by Douglas and Miller (1974). Personal preference is compromised even more if a traveler wants to take a specific type of train because it decreases the likelihood of traveling at a preferred time even further. Thus, the difference between travelers' most preferred travel times and the actual times chosen could cause inconvenience, and it would significantly affect the demand for trains. In order to measure the potential traveler inconvenience, I adopted the notion of *schedule delay* from Douglas and Miller (1974); Miller (1972), which defines it as the absolute difference between the passenger's most preferred time to travel on 24-hour clock and that of his actual time to travel. Each traveler's schedule delay causes disutility. Unlike Douglas and Miller (1974), this paper does not consider capacity constraint as a source of schedule delay, but the train schedules. Therefore, *Schedule Delay* is defined in this paper as the absolute difference between a traveler's most preferred travel time and his actual time to travel.

3.2.3 Traveler's Problem

As I mentioned earlier, the logit model with traveler's heterogeneous preferences with respect to travel time will be adopted in this paper. A traveler i, whose preferred travel time is h^i, faces a choice problem over products given a city-pair m in a time period t: He has to choose how to travel.[7] Traveler i will consider all of the

[7] The model allows the traveler to choose not to travel at all in order to account for the pure increase in the number of travelers.

products in the market mt to choose a product that yields the highest utility. This paper assumes a linear utility(or disutility); hence, the utility function of a traveler i for a product j(a train-route combination), is given by

$$U^i_{jmt} = x_{jmt}\beta + \eta_m + \xi_{jmt} + \gamma \cdot d(a_{jmt}, h^i) + \epsilon^i_{jmt} \tag{3.1}$$

where a vector $x_{j.}$ contains the observed characteristics of each product including fare. Because the heterogeneity of city pairs is huge, the model includes a dummy variable for each city pair m- the coefficient on the dummy variable for city pair m is η_m -in the demand to allow the valuation of inside goods to be different across markets.$\gamma \cdot d(a_{jmt}, h^i)$ measures the inconvenience caused by schedule delay, where $\gamma < 0$. $d(a_{jmt}, h^i)$ is the absolute difference between a_{jmt} and h^i, where h^i is traveler i's preferred travel time of day, and a_{jmt} is his actual time of day to travel specific to product j in market mt.[8]

The product-level unobservable, ξ_{jmt} accounts for a number of product characteristics, which are not observed by econometricians, such as unobserved characteristics of the routes or trains, the facilities inside each train or in the train stations, and the quality of the train attendants. ϵ^i_{jmt} is an additive error term, specific to product j in market mt, which is assumed to follow an extreme value distribution and to be distributed independently across both consumers and products.[9][10] This error term captures each traveler's idiosyncratic tastes in trains or routes, or possibly his physical location or the purpose of his trip.[11]

---

[8] $d(x, y) = \min\{|x - y|, 24 - |x - y|\}$

[9] Berry et al. (2006) consider this as a factor from preference on time to travel. I explicitly include the preference on arrival or departure time in the model

[10] This model was first proposed in McFadden (1973)

[11] As discussed in Section 3.2.1, the purpose of trip can be transferring to another mode or another train.

I explicitly introduced "outside" alternatives in Section 3.1, which include traveling by modes of transportation other than trains as well as not traveling. The outside alternatives have utility

$$U^i_{0mt} = \epsilon^i_{0mt}$$

The mean of this utility is normalized to be zero. The coefficients on city-pair specific dummy variables(η_m) in the utility of "inside goods" are interpreted as being relative to the outside goods.

Given the utility function (3.1), each traveler i purchases one unit of a product j that yields the highest utility. That is, conditional on $(x_{mt}, \eta_m, \xi_{mt}, a_{mt})$ and his preferred time to travel h^i, he will purchase one unit of j if and only if

$$U^i_{jmt} > U^i_{kmt} \qquad \forall k \in J_{mt} \cup \{0\}, \;\; k \neq j$$

where J_{mt} is a set of products available in market mt and $\{0\}$ is a set of outside alternatives.

3.2.4 Market Share

The "market share" of a product is defined as the percentage of travelers using that product out of all potential passengers. The market size is discussed in Section 3.1. Based on the assumption on the distribution of ϵ, the probability that traveler i purchases a product j conditional on $(x_{mt}, \eta_m, \xi_{mt}, a_{mt})$ and i's preferred time to travel, is given by the well-known formula

$$s^i_{jmt}(\delta_{mt}, a_{mt}, \gamma, h^i) = \frac{\exp(\delta_{jmt} + \gamma d(a_{jmt}, h^i))}{1 + \sum_{q \in J_{mt}} \exp(\delta_{qmt} + \gamma d(a_{qmt}, h^i))} \tag{3.2}$$

where $\delta_{jmt} = x_{jmt}\beta + \eta_m + \xi_{jmt},$, and is shared among all travelers in the market.

If the distribution of h^i is known, the market share for each product can be easily obtained from the expectation of (3.2) over h^i. This paper assumes the traveler's preferred time of day to travel to be discrete so that each traveler has his preferred "hour" to travel on a 24-hour clock. This allows the model to be a discrete mixture of logit models. In other words, h^i takes an integer between 1 and 24.[12], and its probability mass function is

$$\text{Prob}(h^i = \tau) = \phi_{\tau mt} \qquad \forall \tau \in B$$

where B is the set of support of h^i, the 24 integers between 1 and 24. The overall market share of product j is

$$s_{jmt}(\delta_{mt}, a_{mt}, \gamma, \phi_{mt}) = \sum_{\tau \in B} \phi_{\tau mt} \cdot s^i_{jmt}(\delta_{mt}, a_{mt}, \gamma, \tau)$$

where $\phi_{\tau mt}$ denotes the percentage of travelers in the potential travelers of market mt whose preferred time to travel is τ.

3.2.5 Distribution of Traveler's Preferred Time

Although this paper does not contain any random coefficient, the model is similar to the mixture model with random coefficients due to the existence of h^i. Ideally, a variable ϕ_τ, defined as the proportion of travelers whose preferred time is τ, can be estimated from the model; however, it is not practical to estimate a different vector of ϕ for each market. Such a task would be impractical even if I assumed that the distribution travelers is common across markets, because estimation is difficult and it is sensitive to small changes in the specification or instruments as Berry and Jia

[12] Although I assume h^i to take an integer, it can be generalized to take one of any 24 real numbers between 0 and 24.

(2010) points out.[13]

To sidestep this issue, this paper uses a proxy for the proportion of the potential travelers with preferred travel time τ, obtained using the following assumptions. First, I assumed that the distribution of traveler's preferred time to travel varies across city-pairs but does not vary across time periods. That is, $\{\phi_{\tau mt}\}_{\tau=1}^{24} = \{\phi_{\tau m}\}_{\tau=1}^{24}, \quad \forall t$. I also assume that the distribution of h^i is same across all the alternatives. Let $w_{\tau m}$ denote the proxy for the proportion of travelers in city-pair m whose preferred time to travel is τ. Replacing $\phi_{\tau mt}$ with the proxy $w_{\tau m}$ allows the overall market share for product j to be rewritten as

$$s_{jmt}(\delta_{mt}, a_{mt}, \gamma) = \sum_{\tau \in B} w_{\tau m} \cdot s^i_{jmt}(\delta_{mt}, a_{mt}, \gamma, \tau) \tag{3.3}$$

Next, it is essential to find a proxy for $\{\phi_{\tau m}\}_{\tau=1}^{24}$ for each m, which reflects the distribution of travelers preferred times of day to travel. The process of constructing the proxy is based on the underlying belief that all travelers will travel at times that is close to their most preferred times. This is a plausible assumption because fares do not vary within a single day. Therefore, preference for a given travel time can be inferred by the number of travelers during that time. Thus, one reasonable candidate for the distribution of h^i is the hourly train ridership in each market taken from the historical data.[14] This assumes that the company schedules trains to support travelers using knowledge of the true distribution of consumers' preferences over the travel schedules; thus the hourly ridership should reflect travelers' true preferences. I obtained the proportion of travelers in each city pair m who actually travel during time period τ using

[13] According to Berry and Jia (2010), the mixture model with more than three types of consumers is difficult to estimate and sensitive to small changes in the specification or instruments.

[14] Figure 3.1(a) displays the mean of the percentage of rail travelers who travel within an hour across city pair with bars.

$$Q_m^\tau = \frac{\sum_t \sum_{j \in J_{mt}^\tau} q_{jmt}}{\sum_t \sum_{j \in J_{mt}} q_{jmt}} \tag{3.4}$$

where J_{mt}^τ is the set of available trains in a market mt with schedule is τ, and q_{jmt} is the number of passengers purchasing product j.[15]

I construct a proxy for $\{\phi_{\tau m}\}_{\tau=1}^{24}$ for each m, smoothing the proportion of travelers in city pair m who actual travel at τ above using Kernel density estimation. In other words,

$$w_{\tau m} = \int_{\tau-1}^{\tau} \frac{1}{Q_m h} \sum_{y=1}^{24} Q_m^y \cdot K\left(\frac{x-y}{h}\right) dx, \quad \tau = 1, 2, \cdots 24 \tag{3.5}$$

where $Q_m = \sum_{y=1}^{24} Q_m^y$ and $K(x) = \frac{1}{\sqrt{2\pi}h} \exp(-\frac{x^2}{2h^2})$. Figure 3.1(a) shows the mean percentage of rail travelers who travel within an hour across city pairs(with bars) and the mean of proxies(with lines) for distribution of travelers' preferred travel time to illustrate the distribution of travelers' preferred time.

3.2.6 Departure Time vs. Arrival Time

This paper assumes that each traveler has a target time in mind for one endpoint of each potential trip that does not vary with mode or schedule choices. In existing literature that discusses preferences over travel schedule, departure time is usually considered instead of arrival time.(Douglas & Miller, 1974; Koppelman et al., 2008) Although it is not common to use preferences over arrival time, this paper adopts arrival time for travel schedule because a traveler normally chooses a departure time and a mode of transportation with a target arrival time in mind. His preferred departure time there depends on how he travels, while his target arrival time re-

[15] This paper uses "hour of arrival time" for train schedule, and discuss the reason in Section. 3.2.6, and thus $J_{mt}^\tau = \{j \in J_{mt} | a_{jmt} = \tau\}$

mains constant during the selection process. In this context, using preference over arrival time instead of departure time is more consistent.[16]

3.2.7 Robustness

In addition to the main analysis that allows travelers to choose to forego travel, I imposed an alternative assumption that do not allow travelers to choose to forego travel. This experiment analyzes how the results vary with the assumption on the market size, and differs from the main analysis in that now the benefits from the introduction of high-speed train are limited to only travelers and not non-travelers. Unlike the definition used in the main specification, the set of outside alternatives is composed of bus, car and domestic flight. Thus, the market size of outside alternatives is calculated by adding the numbers of rail passengers, airline passengers, bus passengers and auto travelers.[17] Using the inferred market size, I compared the changes in consumer surpluses in this specification to those calculated in the main specification, in which the model allows non travelers to switch to traveling by trains.

To examine how robust the results are, this paper considers several different distributions of h^i, based on several assumptions about the distribution of travelers' preferences over travel schedule. I am concerned with the possibility that hourly ridership might distort the distribution of h^i due to train schedules. For example, consider a hypothetical situation where a consumer wants to travel at 10 AM using a Sae-ma-eul train, but there is no such train available. Suppose he has the options of waiting until 12 PM, or taking a KTX train at a higher price. If he chose to wait

[16]I use departure time for a_{jmt}, and adopt preference over departure time instead of arrival time in an alternative specification for the purpose of robustness check. The results are robust. See Column (3) in Table 3.4.

[17]I describe how I calculate the numbers in Section A.

until 12 PM instead paying the higher price, he would be counted as a consumer whose preferred time is 12 PM instead of 10 AM. To examine how robust the results are, this paper considers several different distributions of h^i.

To consider this issue, I first exploit the conjecture that travelers would travel at times around their preferred time of day, then I combine that with another distributional assumption. Specifically, I partition a set of the 24 numbers(denoted by B) into four groups(denoted by B_g, $g = 1, \cdots, 4$) that can be interpreted as Morning, Daytime, Evening, and Night.[18] [19][20] I construct a proxy for the proportion of travelers whose preferred time of day belongs to each time group using actual data. Note that this does not violate the assumption that each traveler would travel at a time that is close to their most preferred time, as I used in the main specification.

In order to take the effects of train availability on the distribution into account, and in attempt to reduce those effects, I assumed a uniform distribution within each time-group(B_g). By extension, this assumption implies that h^i is uniformly distributed within time-group(B_g) but also the train availability induces the observed hourly ridership.[21] Therefore, Prob$(h^i \in B_g) = \sum_{\tau \in B_g} \phi_{\tau m}$ in each city pair m is replaced with the proportion of rail passengers in a city-pair m traveling at time $\tau \in B_g$, and the same number of travelers are located at each point within B_g by assumption. Hence, $\phi_{\tau m}$, the proportion of travelers who prefer to travel at during time period τ, is replaced with $w_{\tau m}$ such that

[18]Thus, it satisfies $B_g \cap B_{g'} = \emptyset$ for any $g \neq g'$, and $B = \cup_{g=1}^{4} B_g$.

[19]The partition is defined based on the observation of actual train schedule. The four groups are defined as 6:00-12:00, 12:00-18:00, 18:00-24:00, and 24:00-6:00 respectively.

[20]This paper experiments different partitions with the length of interval as 4 hours instead of 6 hours, thus the 24 numbers are partitioned into 6 groups- 3:00 7:00, 7:00 11:00, 11:00 15:00 15:00 19:00 19:00 23:00, and 23:00 3:00.

[21]In addition to uniform distribution, I apply Gaussian distribution centered at the median of each time-group and a randomly chosen arbitrary distribution.

$$\begin{aligned} \sum_{\tau \in B_g} w_{\tau m} &= \frac{\sum_t \sum_{j \in J_{mt}^{B_g}} q_{jmt}}{\sum_t \sum_{j \in J_{mt}} q_{jmt}}, \quad \text{and} \\ w_{\tau m} &= \text{Prob}(h^i = \tau | h^i \in B_g) \cdot \sum_{\tau \in B_g} w_{\tau m} \end{aligned} \tag{3.6}$$

where $J_{mt}^{B_g}$ is a set of available trains in a market mt whose schedule belongs to B_g, and q_{jmt} is the number of passengers purchasing j.[22] $\text{Prob}(h^i = \tau | h^i \in B_g)$ is the distributions within time-group.[23]

Figure 3.1 shows the mean of the percentage of rail travelers who travel within an hour across city pairs(with bars) and the mean of proxies(with lines) for the distribution of travelers' preferred travel times under the different assumptions of the time group distribution. Figure 3.1(b) and Figure 3.1(c) show the distribution of h^i based on six time-groups and four time-groups, respectively, combined with the uniform-distribution regarding the within time-group distributions. 3.1(d) display the mean of two different proxies based on the four time-groups, one using a Gaussian(with solid line) and an arbitrary distribution(with dashed line) for the within time-group distribution. The results under these assumptions is discussed in Section 3.5.3.

3.3 Estimation

To estimate the demand parameters (β, γ), I followed the standard BLP procedure due to the presence of the unobserved product characteristics ξ, and of heterogeneous travel time preference h^i.[24] Although the model in this paper does not

[22]In other words, $J_{mt}^{B_g} = \{j \in J_{mt} | a_{jmt} \in B_g\}$

[23]When uniform distribution is assumed for the distribution within time-group, $\text{Prob}(h^i = \tau | h^i \in B_g) = 1/(\text{length of } B_g)$.

[24]proposed in Berry et al. (1995)

include random coefficients, the existence of heterogeneous taste on preferred time to travel makes the model similar to the ones with random coefficients. Therefore, I first inverted the following market share equation for each market to solve for the vector of δ_{mt} as a function of data and the parameters to be estimated

$$s_{mt}(\delta_{mt}, a_{mt}, \gamma) = s^o_{mt} \qquad \forall m, t$$

where $s_{mt}(\delta_{mt}, a_{mt}, \gamma)$ is a vector of market shares in market mt as described in (3.3), and S^o_{mt} is a vector of observed market shares in market mt. As in Berry et al. (1995), this system of equations is nonlinear in the parameters to be estimated ; however, they can be solved numerically using the contraction mapping

$$\delta^{r+1}_{mt} = \delta^r_{mt} + \ln s^o_{mt} - \ln s_{mt}(\delta^r_{mt}, a_{mt}, \gamma)$$

where r denotes the rth iteration. The series was iterated until $\|\delta^r_{mt} - \delta^{r-1}_{mt}\| < \varepsilon$ for a given tolerance $\varepsilon > 0$ to approximate δ_{mt}.[25], δ^R_{mt} where R is the smallest integer where the convergence criteria is satisfied. As described in Nevo (2000), I use two-stage least squares which solve the linear parameters β as a function of the nonlinear parameter γ and limits the nonlinear search in GMM methodology to the nonlinear parameter only.

By assumption, the rail company considers travelers' schedule preference when determining train schedules; therefore, $E(\xi)$ could have non-zero. Accordingly, we must include a set of exogenous instrumental variables to identify the parameters. The moment conditions used in the estimation are derived from

$$E\left[\xi_{mt}|z_{mt}\right] = 0$$

[25]I iterated until the maximum difference between each iteration is smaller than $2 \cdot e^{-25}$

where z_{mt} is a vector of instruments. For any vector of function $h(\cdot)$[26], the moment conditions imply

$$E\left[\xi_{mt} \cdot h(z_{mt})\right] = 0$$

3.3.1 Instrumental Variables

Although strict regulations on pricing mitigate the endogeneity problem from prices, the endogeneity from train schedules is of concern to this research. Since a rail company in Korea has only limited power over pricing, it might have a strong incentive to control product characteristics such as train schedules instead of fares. As a result, the arrival time of product j, a_{jmt} and Schedule Delay, $d(a_{jmt}, h^i)$ might be endogenously determined by the rail company. [27] Therefore, it is necessary to include valid instruments in order to identify the demand model.

The identification strategy used in this paper searches for the variables that affect the rail company's schedule decision, but not those that affect consumer demand, exploiting a special circumstance of the railroad industry. Consider, for example, trains running along a rail line A with stops at stations between A_0 through A_{N+1}(N intermediate stations). When a rail company determines the schedule for those trains, it would ideally consider the demands for each of the individual routes along the railroad. However, a traveler would care only about the routes in the market he travels in. For example, consider two cities, City 1 and City 2. Assume the cities have stations, A_{n_1} and A_{n_2}, respectively, both located on rail line A. Since people who travel from City 1 to City 2 would not care about the routes $A_n \rightarrow A_{n'}, \quad \forall n, n' \neq n_1 \;\&\; n, n' \neq n_2$, the demand for product j, given a train t_1,

[26] For this paper, I transformed z_{mt} using a principal component analysis of a given function $h(\cdot)$ to make the columns of $h(z_{mt})$ orthogonal.

[27] For example, the rail company could schedule more trains at a popular time, thus the schedule delay might be small for high demand products.

of $A_n \rightarrow A_{n'}, \quad \forall n, n' \neq n_1 \;\&\; n, n' \neq n_2$ constitutes valid instrumental variables for j, and let R_{jmt} denote such routes.[28]

3.4 Expected Utility Calculation

The demand estimates based on the model presented in Section 3.2 provide information about how consumers value each of the product characteristics. These results indicate that consumers have significant disutility from traveling at a time other than their preferred time to travel.(See Section 3.5.1.) The next step is to quantify the changes in consumer surplus after high-speed train introduction. In order to evaluate consumer surpluses resulting from the introduction of high-speed trains, I adopt the estimated demand for travel as presented in Section 3.5.1. Since the train schedules changed as a result of the introduction of high-speed trains, I separately considered the changes in consumer surplus caused by train rescheduling and those caused by high-speed train introduction.

The change in consumer welfare can be measured by the difference between the expected utilities in two different situations. I primarily compared the consumers' expected utilities from the set of products offered after the introduction of high-speed trains to those from the products offered before high-speed trains were introduced. To examine the effects of high-speed train introduction separately from other changes such as train reallocation, I considered equilibria under the six different sets of products to evaluate the travelers' surplus, followed by a stepwise comparison to illustrate the effects of situation changes. The six product sets are defined as following :

[28] For example, two of the instrumental variables are $z_{l,jmt} = \sum_{k \in C^l_{jmt}} q_{kmt}$
where $C^1_{jmt} = \{k \in \cup_m J_{mt} | k\text{'s train ID} = j\text{'s train ID} \;\&\; \text{station pair of } k \in R_{jmt}\}$
and $C^2_{jmt} = \{k \in \cup_m J_{mt} | k\text{'s train ID} \neq j\text{'s train ID} \;\&\; a_{kmt} = a_{jmt} \;\&\; \text{station pair of } k \in R_{jmt}\}$

(S1) Train schedules offered to travelers in 2002, before high-speed trains were available, using the prices from 2002.

(S2) Train schedules offered in 2002, before high-speed trains were available, using the prices from 2006.[29]

(S3) High-speed train schedules offered in 2006, including the other types of trains considered in (S2), using the prices from 2006.

(S4) Same as product set in (S3), but excluding the trains that were no longer part of the 2006 schedule, using the prices from 2006.

(S5) Same as product set in (S4), but including the trains that were newly offered in 2006 versus 2002, using the prices from 2006.

(S6) Train schedules offered in 2006, using the prices from 2006.

(S1) and (S6) present actual situations, whereas the others present hypothetical situations. The changes from (S1) to (S2) correspond to the effects of price changes between 2002 and 2006. A comparison between (S2) and (S3) provides the effects of high-speed train introduction on traveler's surplus. The changes from (S3) to (S6) corresponds the effects of schedule changes after the introduction of high-speed trains, and the stepwise comparisons from (S3) to (S6) break down those effects into three components: the effects from the elimination of trains((S3)→(S4)); the effects from the addition of trains((S4)→(S5)); and the effects from the pure reallocation of the existing trains((S5)→(S6)).

To break down the effect of schedule changes into the three components discussed above, it is necessary to group the trains offered in 2002 into those subsequently removed in 2006 and those still remaining in 2006. Since the systems used to assign identification numbers to trains were different in 2002 and 2006, it was not possible to use the train identification number for the sorting. Thus, this pa-

[29] This paper uses the fare and the trains schedules from November 2006 for all 2006 pricing.

per exploits the partition of hours, which is defined in Section 3.2.5 by matching Morning trains offered in 2002 to Morning trains offered in 2006 based on arrival time and train type. For example, if there were five Mu-gung-hwa trains in the Morning group in 2002 and there were six Mu-gung-hwa trains in the Morning group in 2006, I paired the first offered in 2006 with the five trains offered in 2002 and considered them as trains with "adjusted schedules". The one remaining train was then considered as "an added train". Under this sorting rule, a change in the schedule of a train within a time group(B_g) was considered a reallocation, whereas scheduling a train such that it fell into a different time group($B_{g'}, g' \neq g$) was considered a removal of that train from the first time group(B_g) and adding a new train to the second time group($B_{g'}$). Using a different sorting rule could result in a different distribution of consumer welfare changes across "removing trains", "adding trains" and "reallocating trains"; however, the total effects of "schedule changes", which consists of all the three changes, is invariant across different sorting rules.

Given the estimated demand, the expected utility in a market mt is given by

$$EU_{mt} = \sum_{\tau \in B} \phi_{\tau m} \left[E(\max_{j \in J_{mt}} \hat{U}^i_{jmt}) \right]$$

where

$$\begin{aligned} \hat{U}^i_{jmt} &= V^i_{jmt}(\hat{\beta}, \hat{\gamma}, \hat{\eta}_m, h^i) + \epsilon^i_{jmt} \\ V^i_{jmt}(\beta, \gamma, \eta_m, h^i) &= x_{jmt}\beta + \eta_m + \xi_{jmt} + \gamma \cdot d(a_{jmt}, h^i) \end{aligned}$$

To approximate the expected utility, this paper replaces $\phi_{\tau mt}$, the proportion of travelers whose preferred time to travel is τ, with a proxy $w_{\tau m}$, as defined in (3.5). Since ϵ^i_{jmt} in (3.1) is assumed to have the extreme value distribution, the expected utility can be rewritten as[30]

[30] shown in M. E. Ben-Akiva (1973)

$$EU_{mt} = \sum_{\tau \in B} \phi_{\tau m} \left[\log \sum_{j \in J_{mt}} \exp \left(V^i_{jmt}(\hat{\beta}, \hat{\gamma}, \hat{\eta_m}, \tau) \right) \right]$$

From Nevo (2003), a monetary measure of the change in traveler's welfare, EV_{mt} can be constructed by

$$EV_{mt} = -\frac{M_{mt}}{\beta_p}(EU^1_{mt} - EU^0_m) \tag{3.7}$$

where β_p is the price coefficient and M_{mt} is the market size of mt. EU^1_{mt} and EU^0_m represent the expected utilities of situations with high-speed trains and without high-speed trains respectively, thus, (3.7) allows us to compare two different situations with the same demand system.[31]

Developing the remainder of the methodology would be straightforward if ξ was known for both situation 0 and 1. Unfortunately, ξ cannot be observed directly; however, this paper exploits that the trains with the same train type, routing and arrival time, would be more or less homogeneous so as to have similar unobserved product quality.[32] Therefore, this paper uses $\hat{\xi}$, the estimate for ξ, which is obtained from the residuals of an IV regression in the demand estimation, for the unobserved product characteristics in the expected utility calculation. Specifically, $\hat{\xi}$ is directly used in the travelers' surplus calculation for a given data period, and the median of $\hat{\xi}$ among the set of trains of the same train type, routing and arrival time is utilized for the other hypothetical situations analyzed.

[31] While a city pair m is observed for multiple periods in the estimation, the products offered in a counterfactual situation are observed for one period, thus EU^0_m is subscripted only with m. I take mean of EU^1_{mt} over months, t within a city pair m to compare it to EU^0_m.

[32] Because the main source of unobserved product characteristics is a routing.

3.5 Results

This section covers the results of the estimations using the demand model and the expected utility calculations. Section 3.5.1 presents the results of the estimation using the demand model, and contains its own discussion. In Section 3.5.2 I discuss the main findings of this research; The results of analyzing changes in consumer surplus resulting from both high-speed train introduction and train schedule adjustment. Lastly, I summarize results obtained under alternative assumptions regarding market size and the distributions of travelers' preferred times in Section 3.5.3.

3.5.1 Travel Demand

Table 3.4 shows the results of demand estimations based on the main specification that takes both travelers and non-travelers into consideration. Table 3.4 shows the estimated parameters, which include the mean utility parameters(β) and the parameter representing schedule delay(γ). Column (1) shows the parameters using the main specification, and Column (2) shows the same parameters estimated using the same model without employing the excluded instrumental variables. Column (3) shows the parameters resulting from an OLS estimation of $\ln(s_{jmt}/s_{0mt})$ on δ_{jmt}.

As expected, the mean estimated utility of high-speed trains(KTX) was higher than other types of train, and that of Sae-ma-eul trains was lower than KTX but higher than the other two types of train. Schedule Delay has significantly negative impact on demand. In Column (1) of Table 3.4, the estimated coefficient for Schedule Delay is -0.311. The most straightforward method of interpreting this coefficient is to compare it to the price coefficient. The price coefficient(-0.115) and the coefficient for Schedule Delay imply that travelers are willing to pay up to about 2700 KRW to reduce their Schedule Delay by one hour, holding everything else fixed. The coefficient for price shows that consumers are not as sensitive to price. To be more specific, the probability to purchase a product decreases by 9.9% when price increases by 10%.

Examination of the estimated coefficients of the variables that indicate the convenience of each route such as N(Own Type Train/Day), N(Other Type Train/Day) and Station-City Center, reveals that the routes with more trains scheduled provide a higher utility for travelers. The number of a given type of train scheduled within a day affects a traveler's utility more than the schedules of the other types of train. If the number of a given type of train scheduled within a day increases by 10%, travelers choose the corresponding products with 7% higher probability. On the other hand, a 10% increase in the number of other types of trains scheduled within a day results in only a 0.8% higher purchase probability. Distance between station and city center is also an important factor on demand, based on the estimated parameters. If a given station was relocated 10% farther from its city center, consumers would choose the corresponding products with 9.5% lower probability.

3.5.2 Consumer Surplus

I partitioned the markets into three groups based on high-speed train availability in order to consider the heterogeneity of choice sets as well as heterogeneous preferences over travel schedules. This partitioning facilitates an examination of the different effects across heterogeneous consumers. The results for Group 1, which considers consumers in the markets with high-speed train stations, are shown in Column (1) of Tables 3.7, 3.8 and 3.9. Group 1 contains 107 million travelers per month across 107 city pairs. Column (2) of Tables 3.7, 3.8 and 3.9 summarizes the results for Group 2, which consists of the markets that are located along high-speed rail lines without available high-speed trains. Group 2 contains 190.7 million travelers per month across 330 city pairs. The consumers not accounted for in the first two groups belong to Group 3, whose results are shown in Column (3) of Tables 3.7, 3.8 and 3.9. Group 3 covers 615 city pairs with 348.9 million travelers per month. Consumers in Group 1 and Group 2 were expected to experience stronger effects from both introduction of high-speed trains and the resulting schedule adjustments than consumers in Group 3. I summarized the changes in consumer surplus based on these groups, and Table 3.7, 3.8 and 3.9 reflect the main findings of this paper.

Table 3.7 summarizes the expected consumer surplus changes per person for each market. Each subpanel of Panel A in Table 3.7 displays the change in consumer welfare resulting from each of the five different sources described in Section 3.4. The "Price Change" panel shows the estimated change in consumer welfare due to price differences between 2002 and 2006. Since rail fares decreased for 50% of the products available in my dataset, the changes in consumer surplus due to price change is positive. The "Add KTX" panel shows the gains from attributable to the introduction of high-speed trains into the markets. Since high-speed trains became available in the markets of Group 1, only the consumers in Group 1 directly benefited from the new service. The next three subpanels summarize respectively the changes in consumer welfare due to reducing scheduled trains, scheduling additional trains and rescheduling existing trains to another time within same day. The "Total Effect" panel reflects the overall changes in consumer surplus resulting from all the sources of impact.

Each column of Table 3.7 shows the heterogeneous impacts all normalized to be per person on consumers in each of the three groups, described above.[33] The median of the expected per-person change in Group 1 resulting from introducing

[33]How to define a group does not affect the demand estimates and the change of consumer surplus. Welfare analysis by groups facilitates the understanding on how heterogeneous consumers are differentially affected by high-speed train introduction and ensuing schedule changes.

high-speed trains, is 5,600 KRW(see Panel A), but the expected change resulting from train schedule adjustments is -1,900 KRW, offsetting some of those gains.[34] The median of the expected per-person loss in Group 2 resulting from schedule adjustments after high-speed train introduction, is about 11,140 KRW. This loss occurred because some trains that were available before the high-speed train introduction became unavailable after the introduction. Group 3 consumers experienced only minor changes overall compared to consumers in other groups. The median of the expected per-person change in consumer welfare in Group resulting from schedule adjustments after high-speed train introduction, is about 1,900 KRW. Total effect summarizes the changes of consumer welfare compared to that in 2002. The median of the expected consumer surplus change per person in Group 1 is 4,000 KRW, while that in Group 2 is -8,500 KRW.

Table 3.8 summarizes the expected consumer surplus changes in each market, taking into consideration market sizes and the magnitudes of impact per person.[35] The results obtained using the main specification(shown in Panel A) demonstrate that both the introduction of high-speed train and the ensuing changes in train schedule had substantial effects on consumer welfare, and that the size of the impact varied across consumers. The fact that the median and mean impacts are substantially different suggests that the changes in consumer surplus are heterogeneous across markets. Although the mean of the expected *per-person* consumer surplus change in Group 1 resulting from reallocating trains is positive, the mean calculated *per market* is negative. This implies that the losses resulting from reallocating trains occurred in larger markets, which tended to also be more strongly

[34] The expected change from schedule adjustment after high-speed train introduction is the sum of the changes from removing trains, adding trains and rescheduling trains.

[35] Table 3.8 shows the summary statistics of (The per-person expected surplus changes in each market) $\times$ (Market Size)

affected by high-speed train introduction directly, while some other markets in Group 1 benefited.

Table 3.9 summarizes the gross changes of consumer surplus in each of the three groups. As I pointed out earlier, rail fares decreased for 50% of the products available in my dataset, thus the overall changes in consumer surplus due to price change was positive. The second row in each panel shows the gains from introducing high-speed trains to the markets. Since high-speed trains became available in the markets of Group 1, only the consumers in Group 1 benefited from the new high-speed rail service. More concretely, the introduction of high-speed trains caused an estimated 10 trillion KRW increase in consumer surplus per month(see Panel A). The net gains for travelers in Group 1 are not as large as superficially anticipated, however, since schedule changes such as reallocation and reduction of non-high-speed trains caused sizable losses that offset 50% of the direct gains resulting from the introduction of high-speed trains.

The next three rows(rows 3-5) summarize the changes in consumer welfare due to rescheduling such as reducing the number of scheduled trains, scheduling additional trains and reallocating existing trains to another time slot within same day. The consumer welfare change due to schedule adjustments in Group 1 markets was about -560 billion KRW. Consumers in Group 2 suffered a considerable amount of loss, -2.4 trillion KRW, due to changes in the set of products offered because train schedules in the corresponding markets were reduced by more than 50%. Without any added benefits from new high-speed services, consumers in the markets of Group 2, experienced losses three times higher than the gains of Group 1 resulting from the introduction of high-speed trains. On the other hand, consumer welfare in the markets of Group 3, increased by about 2 trillion KRW. Although some trains were removed from the original schedules, the gains resulting from additional non-high-speed trains and from reallocated trains outweighed the losses resulting from removed trains. Unlike consumers in Groups 2 and 3, consumers in Group 1 suffered a loss of 73 billion KRW resulting from trains rescheduled to other time slots. This is because KTX trains are primarily scheduled at peak times and non-high-speed trains are primarily scheduled away from those times.

Overall, the gains from having high-speed train are substantial. However, the losses from schedule adjustments that consumers were subjected to in the markets that are located along high-speed rail lines without high-speed trains scheduled, outweighed those gains. Overall changes in consumer surplus were about 317 billion KRW, however, the positive changes are led by the gains from schedule adjustment in Group 3 markets, but the gains from high-speed trains do not exceed the losses that occurred due to schedule reductions in Group 2 markets.

To summarize, introducing high-speed trains substantially raised consumer surplus in markets where they were actually made available. The changes in the set of products offered to consumers offset 50% of the gains, however. Moreover, it resulted in greater losses of consumer surplus in markets located along high-speed rail lines but not connected by high-speed trains, and those losses outweighed the gains directly from introduction of high-speed trains. The overall change in consumer surplus after the introduction of high-speed train was positive because the gains resulting from schedule adjustments in markets that are not located along high-speed rail lines made up for the losses in markets that are located along high-speed rail lines without available high-speed trains. I also found that there are substantial differences in the magnitudes of the consumer welfare changes across heterogeneous consumers. The benefit gained directly from high-speed trains is concentrated in some of the markets, although changes in the choice sets affected a broader range of consumers.

3.5.3 Alternative Specifications

Tables 3.5, 3.10 and 3.11 provide the results under alternative assumptions. Table 3.5 provides the coefficients estimated under alternative assumptions and Table 3.11 compares the respective changes in consumer welfare.

Table 3.5 provides the coefficients estimated under alternative assumptions and Table 3.11 compares the respective changes in consumer welfare. Column (2) presents the results from the specification that use departure time instead of arrival time. Therefore, travel time of day a_{jmt} is hour of product j's departure time, and preference of travel schedule h^i is also defined over departure time. Column (3)-(6) of Table 3.5 and Panels B-E of 3.11 present the results from the specification that adopts $w_{\tau m}$ shown in (3.6) as a proxy for the distribution of h^i. Column (3) and Panel B assume that B is partitioned into 6 time groups with 4-hour intervals as defined in 3.2.7 and h^i is uniformly distributed within each time-group. Columns (4),(5) and (6) and Panels C, D and E assume that B is partitioned into 4 time groups with 6-hour intervals as defined in 3.2.7 with different within-group distributional assumptions for h^i. Column (4) and Panel C utilizes a uniform distribution, and Column (5) and Panel D use a normal distribution centered at the median of each time-group. Column (6) and Panel E employ a randomly-chosen arbitrary distribution, which is shown in Figure 3.1(d). Since most of the losses resulting from schedule changes are due to the reduced number of scheduled trains and not due to reallocations, the implications regarding consumer welfare are still consistent with the findings from the main specification. They are robust across the assumptions on the distribution of h^i.

Table 3.10 and 3.11 provide the results under the alternative assumptions. Table 3.11 compares the respective changes in consumer welfare. These results are based on the estimated coefficients presented in Table 3.5.

Table 3.10 and Panel A of Table 3.11 adopt the estimated coefficients shown in Column (1) of Table 3.5 to show the results under the assumption that does not allow non-travelers to travel. Panel A of Table 3.10 shows the heterogeneous impacts all normalized to be per person on consumers in each of the three groups. Panel B of Table 3.10 summarizes the expected consumer surplus changes in each market, taking into consideration market sizes and the magnitudes of impact per person. Panel A of Table 3.11 displays the nationwide total changes in consumer welfare resulting from each of the five different sources. The per-person impacts from each source(shown in Panel A of 3.10) are similar shown in Panel A of Table 3.7, whether consumers are allowed to forego travel or not. However, the changes of consumer surplus *per market* reflected in Panel B of 3.10 are different from those in Table 3.8 despite the similar magnitudes of per-person impact. Moreover, the nationwide total effect became negative because these results are based on the assumption that the changes in consumer surplus from the introduction of high-speed trains are limited to travelers and the estimated changes are understated. One general conclusions to be made regardless of the assumed market size, is that the gains from high-speed trains introduction are not as substantial as superficially anticipated due to the losses resulting from the reduced schedule frequency in Group 2. These results highlight the importance of accounting for changes in existing products when analyzing the impact of new product entry on consumers.

Panels B-E of 3.11 present the results from the specification that adopts $w_{\tau m}$ shown in (3.6) as a proxy for the distribution of h^i, using the estimated results shown in Column (3)-(6) of Table 3.5. Panel B assume that B is partitioned into 6 time groups with 4-hour intervals as defined in 3.2.7 and h^i is uniformly distributed within each time-group as in Column (3) of Table 3.5.

Panels C, D and E assume that B is partitioned into 4 time groups with 6-hour intervals as defined in 3.2.7 with different within-group distributional assumptions for h^i. Panel C utilizes a uniform distribution with the estimated result shown in Column (4) of Table 3.5, and Panel D use a normal distribution centered at the median of each time-group with the estimated result shown in Column (5) of Table 3.5. Panel E employ the estimated result shown in Column (6) of Table 3.5 and a randomly-chosen arbitrary distribution, which is shown in Figure 3.1(d). Since most of the losses resulting from schedule changes are due to the reduced number of scheduled trains and not due to reallocations, the implications regarding consumer welfare are still consistent with the findings from the main specification. They are robust across the assumptions on the distribution of h^i.

3.5.4 Limitation

A critical limitation of these results is an implicit assumption on the stability of demand system. This approach presumes that consumers had the same demand over product characteristics regardless of the existence of the new product. The results are derived based on the estimates of indirect utility function for the period after the innovation although ex ante and ex post welfare calculations provide quantitatively different measures.(Trajtenberg, 1989) Since the estimated demand is only based on the revealed preferences observed for the periods after the introduction, the counterfactual consumer surplus is valid only if the functional form of the demand is stable as we move away from the center of the data.

More serious problem arises due to the distribution of travelers' preferred time. First, we cannot guarantee that the distribution over travelers' preferred time is time invariant. The assumption imposed when the proxy for ϕ is constructed could lead to the bias in the results. I used the hourly train ridership in each market from the historical data for the proxy, assuming that the train schedule and the hourly ridership reflect travelers' true preference. However, this could lead to a biased result if the preference over travel schedule changed after the introduction because scheduling trains in a different way from the one observed in 2006 will result in welfare losses. I believe that this bias is not serious because i) the welfare implication is robust under other distributions(see Table 3.11), ii) the proportion of welfare changes due to schedule preference is relatively smaller than those coming from schedule frequencies.

Lastly, I want to suggest a potential extension of this research. In the model, I focus heterogenous preference over travel schedule rather than heterogeneous sensitivity to fare and schedule delay. The model suggested in this chapter can be generalized so as to allow for the random coefficient on price and schedule delay. The heterogeneity of sensitivity to schedule delay is another dimension of heterogeneity though it is potentially correlated with traveler's preferred time of traveling. In reality, the sensitivity would affect consumer's modal choice together with the sensitivity to prices.[36] Therefore, in the generalized model, one needs to consider potential correlation between preference on travel schedule and sensitivity to fare and schedule delay.[37]

3.6 Conclusion

In this chapter I addressed the effect on consumer surplus resulting from the introduction of high-speed trains and the ensuing changes in train schedules. I examined the impacts of introducing high-speed trains on consumer welfare using Korean transportation industry data, taking changes in rail company's product se-

[36]For example, travelers whose disutility from schedule delay is severe are more likely to take high-speed trains, which are scheduled more frequently than conventional trains.

[37]For example, business travelers who are likely to be more sensitive to schedule delay than fare would have different preference over travel schedule from leisure travelers'.

lection into account. With this data, I estimate a model of travel demand, that incorporates consumers' heterogeneous preferences over travel schedules into a standard discrete choice model. My analysis treated the rail company's choice of train schedules as endogenous. After comparing the consumer surplus resulting from a set of products offered to consumers before and after high-speed train introduction, this paper yields the implications in consumer surplus. I discussed in detail a rich analysis of consumer welfare changes after the introduction of high-speed trains and of the indirect welfare changes resulting from changes in the firm's product selection.

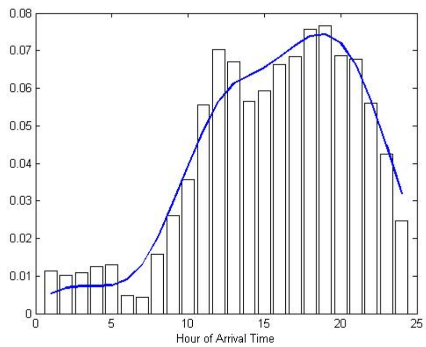
My results show that consumers newly introduced high-speed trains had differential effects on consumers, and that the ensuing changes in train schedules also indirectly affects consumer surplus. The changes in consumer surplus within a market depended on availability of high-speed train. In order to investigate the effect, which varies across heterogeneous consumers, I partitioned markets into three groups based on the availability of high-speed trains in consumers' choice sets. Group 1 consumers who travel between two cities connected by high-speed trains benefited from the new product, but 50% of the gains were offset by the changes in the set of products offered to those consumers. On the other hand, Group 2 consumers, who travel along high-speed rail lines but do not have high-speed trains in their choice sets, suffered significant welfare losses from a reduction in frequency of non-high-speed trains. Group 3 consumers who travel between two cities that are not located along high-speed rail lines, experienced an increased number of trains scheduled, thus substantially increasing consumer surplus. Overall, the losses for Group 2 consumers outweighed the gains resulting from high-speed trains being made available to Group 1 consumers. However, the consumer surplus for Group 3 consumers increased due to the increased schedule frequencies; the increase incidentally made up for the losses for Group 2. The overall consumer surplus after high-speed train introduction increased; however, that increase was not nearly as substantial as the gains directly resulting from the introduction of high-speed trains because of the losses incurred by groups to which high-speed trains were not made available. These results highlight the importance of accounting for changes in existing products when analyzing the impact of new product entry on consumers.

There is, as always, a caveat. This approach presumes that consumers had the same demand over product characteristics regardless of the existence of the new product. Since the estimated demand is only based on the revealed preferences observed for the periods after the introduction, the counterfactual consumer surplus is valid only if the functional form of the demand is stable as we move away from the center of the data. One weakness of my work is that the structural model does not incorporate the changes occurred in other industries due to data constraints although other transportation industries are closely related according to Chapter 2. The lack of the supply side model which is desired for the full-blown cost benefit analysis is also a limitation of this work, and this leaves a scope for improvement and further studies.

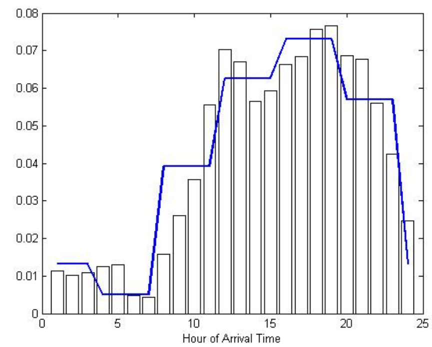
Since my work is limited to the impact only within the transportation industry, there is an opportunity for future research regarding wider impacts of transportation improvement. Since transportation system is a key infrastructure for the development, the improvement in transportation connections could bring much larger impact, particularly on the growth in industries other than transportation.

Table 3.1: Variable Definition

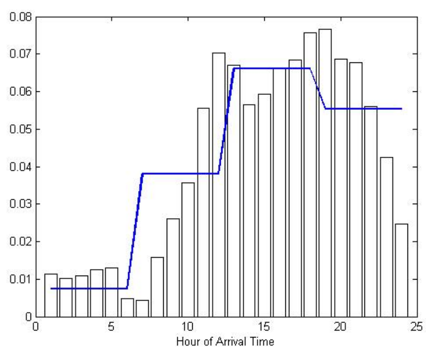
Variable	Definition
City Pair	Directional pair of two cities
Market	A combination of a pair of two cities and a month
j	A combination of train ID and a pair of stations(routing)
Average Population	Geometric average of two cities population
Rail,Air,Bus Passengers	Total number of passengers in buses, domestic flights, and trains who travel for a directional pair of two cities
Car Ownership	Geometric average of household counts with cars in two cities
Market Size	Average Population
Market Share(j)	Q_j / Market Size
Q_j	Number of passengers in a month who travel for a directional pair of stations using a specific train
Price(10^3 KRW)	Real price of each product(10^3 KRW $\approx$ 1 USD)
Distance(Km)	Distance of each routing
N(Own Type Train/Day)	Number of same type of trains as j running for a pair of stations in a day
N(Other Type Train/Day)	Number of different types of trains from j's running for a pair of stations in a day
Station-City Center(Km)	Distance from departure station to departure city center + Distance from arrival station to arrival city center
Schedule Delay	Absolute difference between traveler's preferred time and actual time to travel



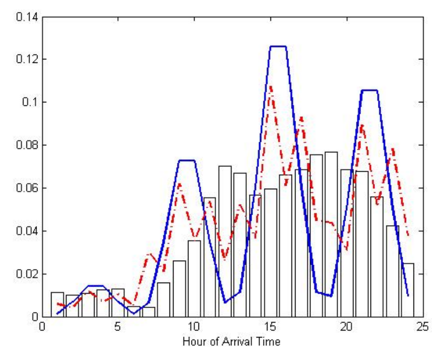
(a) Hourly Ridership and Distribution used in the Estimation



(b) Hourly Ridership and Distribution with 6 Time-Groups and Uniform Distribution



(c) Hourly Ridership and Distribution with 4 Time-Groups and Uniform Distribution



(d) Hourly Ridership and Distribution with 4 Time-Groups and Gaussian, Arbitrary Distribution

Figure 3.1: Hourly Ridership and Distribution of Travelers' Preferred Time

Table 3.2: Summary Statistics

Variable	Mean	STDEV	Median	Min	Max
N=13,347					
Average Population	616,195	724,814	358,772	53,353	7,796,378
Rail,Air,Bus Passengers + Car Ownership	100,172	126,147	56,331	6,874	1,366,424
N=392,459					
Market Share(j)	0.0002	0.0004	0.0001	0.0000	0.0242
Q_j	182	447	49	0	15041
Price(10^3KRW)	8.6	6.7	6.5	1.9	47.0
Distance(Km)	126.0	97.3	96.6	2.9	506.4
N(Own Type Train/Day)	12.5	11.3	9.0	1.0	68.0
N(Other Type Train/Day)	5.7	10.3	1.0	0.0	92.0
Station-City Center(Km)	13.8	9.4	11.4	1.0	82.3

Table 3.3: Summary Statistics by Train Types

Variable	Mean	Median	STDEV	Mean	Median	STDEV
		KTX (N=15,334)			Sae-ma-eul (N=48,436)	
Market Share(j)	0.0004	0.0003	0.0003	0.0002	0.0001	0.0002
Q_j	1115	618	1533	170	84	234
Price(10^3KRW)	17.8	16.3	10.7	14.7	11.6	8.6
Distance(Km)	163.9	133.0	100.5	161.2	133.5	110.3
N(Own Type Train/Day)	25.9	18.0	21.1	6.7	5.0	5.4
N(Other Type Train/Day)	22.1	23.0	11.0	18.1	14.0	14.9
Station-City Center(Km)	9.8	8.9	3.7	10.6	8.9	6.4
		Mu-gung-hwa (N=328,689)				
Market Share(j)	0.0002	0.0001	0.0004			
Q_j	140	41	281			
Price(10^3KRW)	7.3	5.3	5.2			
Distance(Km)	119.1	89.0	93.5			
N(Own Type Train/Day)	12.8	10.0	10.6			
N(Other Type Train/Day)	3.1	0.0	6.9			
Station-City Center(Km)	14.5	12.3	9.8			

Table 3.4: Estimated Coefficients of Demand Model

	(1) Main Model	(2) Without Instruments	(3) OLS
Schedule Delay(Hour)	-0.311***	-4.613***	-
	(0.004)	(0.474)	-
Price(10^3KRW)	-0.115***	-0.113***	-0.118***
	(0.002)	(0.002)	(0.002)
N(Own Type Train)	0.056***	0.058***	0.056***
	(3.0E-4)	(3.2E-4)	(2.7E-4)
N(Other Type Train)	0.015***	0.017***	0.014***
	(2.4E-4)	(2.8E-4)	(2.8E-4)
Station-City Center	-0.069***	-0.071***	-0.067***
	(2.9E-4)	(3.0E-4)	(2.6E-4)
I(KTX)	-1.240***	-1.262***	-1.204***
	(0.029)	(0.032)	(0.033)
I(Sae-ma-eul)	-0.434***	-0.502***	-0.348***
	(0.017)	(0.018)	(0.018)
I(KTX)*Distance	0.017***	0.015***	0.017***
	(3.4E-4)	(3.8E-4)	(3.8E-4)
I(Sae-ma-eul)*Distance	0.008***	0.007***	0.008***
	(1.7E-4)	(1.8E-4)	(1.9E-4)
I(KTX)*Distance^2	-8.8E-6***	-6.5E-6***	-8.8E-6***
	(8.4E-7)	(9.2E-7)	(9.1E-7)
I(Sae-ma-eul)*Distance^2	-6.2E-6***	-4.5E-6***	-7.1E-6***
	(4.3E-7)	(4.6E-7)	(4.7E-7)
Distance	0.011***	0.010***	0.011***
	(2.4E-4)	(2.4E-4)	(2.2E-4)
Distance^2	-2.1E-5***	-2.1E-5***	-2.1E-5***
	(6.3E-7)	(6.2E-7)	(5.6E-7)
Constant	-8.872***	-7.030***	-9.187***
	(0.036)	(0.029)	(0.020)
R^2	0.578	0.584	0.536
City Pair FE	YES	YES	YES

N=392,459; N(Markets)=13,347; N(City Pairs)=1,114

***Significant at p=0.01;**Significant at p=0.05;*Significant at p=0.1

Table 3.5: Estimated Coefficients of Demand Model under Alternative Distribution of h^i

	(1)	(2)	(3)	(4)	(5)	(6)
Distribution of h^i	Kernel	Kernel				
N(Time-groups)	N\A	N\A	6 Time-Groups	4 Time-Groups	4 Time-Groups	4 Time-Groups
Within Group Distribution	N\A	N\A	Uniform	Uniform	Gaussian	Arbitrary
Schedule Delay(Hour)	-0.322***	-0.329***	-0.361***	-0.480***	-0.497***	-0.481***
	(0.004)	(0.004)	(0.005)	(0.008)	(0.009)	(0.008)
Price(10^3KRW)	-0.106***	-0.114***	-0.115***	-0.115***	-0.114***	-0.115***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
N(Own Type Train)	0.057***	0.056***	0.057***	0.057***	0.058***	0.057***
	(3.0E-4)	(3.0E-4)	(3.0E-4)	(3.0E-4)	(3.1E-4)	(3.0E-4)
N(Other Type Train)	0.016***	0.015***	0.015***	0.015***	0.015***	0.015***
	(2.4E-4)	(2.4E-4)	(2.4E-4)	(2.4E-4)	(2.5E-4)	(2.4E-4)
Station-City Center	-0.068***	-0.069***	-0.069***	-0.069***	-0.069***	-0.069***
	(2.9E-4)	(2.9E-4)	(2.9E-4)	(2.9E-4)	(3.0E-4)	(2.9E-4)
I(KTX)	-1.259***	-1.241***	-1.236***	-1.240***	-1.231***	-1.242***
	(0.029)	(0.029)	(0.030)	(0.030)	(0.030)	(0.030)
I(Sae-ma-eul)	-0.466***	-0.435***	-0.417***	-0.408***	-0.369***	-0.406***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.017)
I(KTX)*Distance	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***
	(3.4E-4)	(3.4E-4)	(3.4E-4)	(3.5E-4)	(3.5E-4)	(3.5E-4)
I(Sae-ma-eul)*Distance	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(1.7E-4)	(1.7E-4)	(1.7E-4)	(1.7E-4)	(1.8E-4)	(1.7E-4)
I(KTX)*Distance2	-9.1E-6***	-8.3E-6***	-8.6E-6***	-8.3E-6***	-7.9E-6***	-8.2E-6***
	(8.2E-7)	(8.3E-7)	(8.4E-7)	(8.4E-7)	(8.6E-7)	(8.4E-7)
I(Sae-ma-eul)*Distance^2	-6.7E-6***	-6.1E-6***	-6.1E-6***	-6.1E-6***	-6.3E-6***	-6.1E-6***
	(4.4E-7)	(4.3E-7)	(4.4E-7)	(4.4E-7)	(4.5E-7)	(4.4E-7)
Distance	0.010***	0.011***	0.011***	0.011***	0.011***	0.011***
	(2.4E-4)	(2.4E-4)	(2.4E-4)	(2.4E-4)	(2.5E-4)	(2.4E-4)
Distance^2	-2.1E-5***	-2.1E-5***	-2.1E-5***	-2.1E-5***	-2.0E-5***	-2.0E-5***
	(6.3E-7)	(6.4E-7)	(6.3E-7)	(6.3E-7)	(6.4E-7)	(6.3E-7)
Constant	-7.306***	-8.834***	-8.750***	-8.476***	-8.440***	-8.472***
	(0.036)	(0.036)	(0.036)	(0.038)	(0.039)	(0.038)
R^2	0.573	0.578	0.580	0.577	0.568	0.575
City Pair FE	YES	YES	YES	YES	YES	YES

In all the specifications above except (1), a market is consist of travelers and non-travelers.
In all the specifications above except (2), travel time preference is defined over arrival time.
In Column (2), it is defined over departure time
N=392,459; N(Markets)=13,347; N(City Pairs)=1,114
***Significant at p=0.01;**Significant at p=0.05;*Significant at p=0.1

Table 3.6: Number of Products Available in Each Group of Markets

		Data in 2002			Data in 2006		
		(1)	(2)	(3)	(4)	(5)	(6)
		Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
N(City Pairs)		107	330	644	108	330	676
KTX	N	-	-	-	108	-	-
	Mean	-	-	-	14.6	-	-
	Median	-	-	-	8	-	-
	STDEV	-	-	-	19.0	-	-
Sae-ma-eul	N	105	127	257	108	260	246
	Mean	15.9	26.4	4.6	8.4	6.7	7.6
	Median	8	25	4	5	4	5
	STDEV	22.2	18.2	3.4	9.0	8.7	7.3
Mu-gung-hwa	N	107	330	637	107	330	669
	Mean	80.1	62.0	23.3	41.7	31.8	20.5
	Median	50	42	13	32	25	15
	STDEV	76.1	50.9	23.8	33.0	23.1	21.3

N: the number of city pairs in each group where each type of trains is available.

Group 1 : City pairs with high-speed connection

Group 2 : City pairs on high-speed rail lines without available high-speed trains

Group 3 : City pairs that are not located along high-speed rail lines

Table 3.7: Changes of Consumer Surplus Per Person Across Markets(10^3KRW)

			(1)	(2)	(3)
			Group 1	Group 2	Group 3
	N(City Pairs)		107	330	615
Panel A	Price change	Mean	0.54	0.33	0.96
		Median	-0.18	-0.39	0.23
		STDEV	1.44	1.42	1.82
	Add KTX	Mean	5.64	0.00	0.00
		Median	3.65	0.00	0.00
		STDEV	5.46	0.00	0.00
	Remove Trains	Mean	-6.20	-13.81	-1.62
		Median	-5.74	-13.09	-1.66
		STDEV	4.04	8.42	12.42
	Add Trains	Mean	1.71	2.47	3.93
		Median	0.87	0.68	2.26
		STDEV	2.16	7.06	11.38
	Reschedule Trains	Mean	2.29	2.52	2.16
		Median	1.91	1.40	1.56
		STDEV	3.79	3.87	4.67
	Total Effect	Mean	3.98	-8.50	5.43
		Median	3.74	-10.68	3.01
		STDEV	7.81	11.41	8.07

Panel A is based on the estimates shown in Column (1) of Table 3.4
Group 1 : City pairs with high-speed connection
Group 2 : City pairs on high-speed rail lines without available high-speed trains
Group 3 : City pairs that are not located along high-speed rail lines

Table 3.8: Change of Consumer Surplus Across Markets(10^6 KRW)

			(1)	(2)	(3)
			Group 1	Group 2	Group 3
	N(City Pairs)		107	330	615
Panel A	Price change	Mean	-400.29	120.38	359.80
		Median	-79.15	-65.07	67.78
		STDEV	2136.95	1336.25	1525.64
	Add KTX	Mean	9930.24	0.00	0.00
		Median	1359.58	0.00	0.00
		STDEV	22579.08	0.00	0.00
	Remove Trains	Mean	-6317.17	-9868.11	-1106.28
		Median	-1879.11	-4427.52	-578.20
		STDEV	11722.20	19636.37	11437.76
	Add Trains	Mean	1790.29	1274.05	3079.74
		Median	333.52	217.88	672.13
		STDEV	3870.33	3994.22	12537.64
	Reschedule Trains	Mean	-683.35	1333.59	1262.32
		Median	562.99	438.95	455.32
		STDEV	4782.99	2850.51	5191.70
	Total Effect	Mean	4319.72	-7140.08	3595.59
		Median	1340.29	-2668.40	833.82
		STDEV	16219.36	19151.81	7673.53

Panel A is based on the estimates shown in Column (1) of Table 3.4
Group 1 : City pairs with high-speed connection
Group 2 : City pairs on high-speed rail lines without available high-speed trains
Group 3 : City pairs that are not located along high-speed rail lines

Table 3.9: Gross Change of Consumer Surplus in Each Group of Markets(10^9 KRW)

		(1) Group 1	(2) Group 2	(3) Group 3	National Gross
	N(City Pairs)	107	330	615	1052
Panel A	Price change	-42.83	39.73	221.28	218.17
	Add KTX	1062.54	0.00	0.00	1062.54
	Remove Trains	-675.94	-3256.47	-680.36	-4612.77
	Add Trains	191.56	420.44	1894.04	2506.04
	Reschedule Trains	-73.12	440.08	776.33	1143.30
	Total Effect	462.21	-2356.23	2211.29	317.27

Panel A is based on the estimates shown in Column (1) of Table 3.4

Group 1 : City pairs with high-speed connection

Group 2 : City pairs on high-speed rail lines without available high-speed trains

Group 3 : City pairs that are not located along high-speed rail lines

Table 3.10: If non-travelers are excluded from the consideration

			(1)	(2)	(3)
			Group 1	Group 2	Group 3
	N(City Pairs)		107	330	615
Changes of Consumer Surplus Per-Person Across Markets(10^3 KRW)					
Panel A	Price change	Mean	0.54	0.33	0.96
		Median	-0.18	-0.39	0.23
		STDEV	1.44	1.42	1.82
	Add KTX	Mean	6.22	0.00	0.00
		Median	4.03	0.00	0.00
		STDEV	6.03	0.00	0.00
	Remove Trains	Mean	-6.79	-15.20	-2.14
		Median	-6.23	-14.36	-1.82
		STDEV	4.43	8.82	11.21
	Add Trains	Mean	1.87	2.75	4.66
		Median	0.95	0.75	2.49
		STDEV	2.36	7.21	10.71
	Reschedule Trains	Mean	2.41	2.75	2.40
		Median	2.04	1.50	1.73
		STDEV	4.23	4.23	4.72
	Total Effect	Mean	4.24	-9.36	5.89
		Median	4.08	-11.83	3.27
		STDEV	8.59	12.49	8.81
Change of Consumer Surplus Across Markets(10^6 KRW)					
Panel B	Price change	Mean	-57.74	35.09	59.61
		Median	-13.21	-10.39	11.89
		STDEV	310.35	239.03	249.15
	Add KTX	Mean	1887.60	0.00	0.00
		Median	236.08	0.00	0.00
		STDEV	4286.02	0.00	0.00
	Remove Trains	Mean	-1119.94	-1842.63	-230.97
		Median	-344.91	-804.49	-97.34
		STDEV	2173.30	4066.60	1850.27
	Add Trains	Mean	373.02	215.18	494.61
		Median	59.87	35.96	120.20
		STDEV	858.81	687.04	1912.64
	Reschedule Trains	Mean	-207.59	219.79	228.43
		Median	85.57	75.17	75.89
		STDEV	1082.54	524.21	840.52
	Total Effect	Mean	875.35	-1372.56	551.68
		Median	224.52	-496.24	141.45
		STDEV	3313.28	3978.88	1109.29

Panels A and B are based on the estimates shown in Column (2) of Table 3.5
Group 1 : City pairs with high-speed connection
Group 2 : City pairs on high-speed rail lines without available high-speed trains
Group 3 : City pairs that are not located along high-speed lines

Table 3.11: Change of Consumer Surplus in Each Group of Markets(10^9 KRW)

		(1) Group 1	(2) Group 2	(3) Group 3	National Gross
Panel A	Price change	-6.18	11.58	36.66	42.06
	Add KTX	201.97	0.00	0.00	201.97
	Remove Trains	-119.83	-608.07	-142.05	-869.95
	Add Trains	39.91	71.01	304.19	415.11
	Reschedule Trains	-22.21	72.53	140.48	190.80
	Total Effect	93.66	-452.95	339.28	-20.00
Panel B	Price change	-42.74	39.74	220.67	217.68
	Add KTX	1072.59	0.00	0.00	1072.59
	Remove Trains	-681.41	-3299.42	-688.46	-4669.29
	Add Trains	193.48	434.63	1949.32	2577.43
	Reschedule Trains	-72.51	464.37	836.58	1228.44
	Total Effect	469.41	-2360.68	2318.11	426.84
Panel C	Price change	-42.88	39.68	219.18	215.99
	Add KTX	1086.82	0.00	0.00	1086.82
	Remove Trains	-686.53	-3350.46	-718.88	-4755.87
	Add Trains	196.25	467.69	2116.00	2779.94
	Reschedule Trains	-66.38	517.71	915.54	1366.88
	Total Effect	487.29	-2325.38	2531.84	693.75
Panel D	Price change	-42.74	39.65	219.34	216.24
	Add KTX	1079.29	0.00	0.00	1079.29
	Remove Trains	-686.33	-3362.60	-728.49	-4777.42
	Add Trains	198.34	456.26	2126.00	2780.60
	Reschedule Trains	-74.86	489.85	901.31	1316.30
	Total Effect	473.69	-2376.84	2518.16	615.00
Panel E	Price change	-42.93	39.69	218.99	215.75
	Add KTX	1083.39	0.00	0.00	1083.39
	Remove Trains	-684.86	-3344.87	-715.69	-4745.41
	Add Trains	196.15	462.47	2110.71	2769.32
	Reschedule Trains	-66.99	513.11	911.36	1357.48
	Total Effect	484.77	-2329.59	2525.35	680.53

Panel A is based on the estimates shown in Column (1) of Table 3.5
Panel B is based on the estimates shown in Column (3) of Table 3.5
Panel C is based on the estimates shown in Column (4) of Table 3.5
Panel D is based on the estimates shown in Column (5) of Table 3.5
Panel E is based on the estimates shown in Column (6) of Table 3.5
Group 1 : City pairs with high-speed connection
Group 2 : City pairs on high-speed rail lines without available high-speed trains
Group 3 : City pairs that are not located along high-speed rail lines

Chapter 4

Concluding Remarks

In this thesis, I examined the changes in consumers' choices and their welfare due to the introduction of high-speed rail service. My work adds to the existing literature by considering the impact not only on the rail industry but also on the competing modes of transportation such as domestic airline or intercity bus industries. It also takes the ensuing changes in product characteristics or the set of products offered to consumers after the new product entry into consideration, and investigates how those changes affect consumer welfare. Through this thesis, I emphasized the importance of considering all the related industries which are under potential influence and accounting the subsequent changes in existing products when we evaluated new product introduction.

I performed empirical analyses with Korean transportation industry data to evaluate the impact of high-speed train introduction on the demand for passenger travel, applying a difference in differences model and a fixed effect model in Chapter 2. My results from the fixed effect model show that high-speed train introduction caused significant changes not only in the ridership of rail industry but also in that of the entire Korean transportation industry. In addition, the impact from high-speed train introduction was not restricted to the routes where high-speed train has been made available. The routes without high-speed train service have been also affected, and how each route was affected depends on whether it is connected to high-speed rail service.

To be specific, for the analyses in Chapter 2, I partitioned all the routes into four groups based on high-speed train availability. Group 1 consists of the routes with high-speed connections. Group 2 and Group 3 contain the routes that are not connected by high-speed trains but are under potential influence. To be specific, the routes in Group 2 can be partially replaced with high-speed trains, thus the ridership for those routes are expected to decrease. In contrast, the routes in Group 3 are along the branch lines of a high-speed rail line, thus the ridership for these routes are expected to increase due to the travelers who transfer at a high-speed train station. All other routes are considered to be included in the control group. According to the results from the most general specification which allows individual specific intercepts and group specific linear trends, I found that the ridership of rail service increased 32% in the routes with high-speed trains. In contrast, the train ridership for the routes which can be partially replaced with high-speed trains, decreased by 44%. On the contrary, the routes that are located along the branch lines of a high-speed rail line attracted more passengers, thus the ridership in these routes increased by 70%.

At the same time, Korean intercity bus and domestic airline industries also experienced large impact after the introduction of high-speed trains. I found that the ridership of domestic airlines for the routes where they directly compete with high-speed trains lost more than 30% of their customers. These losses are concentrated in the Group 1 routes unlike in the rail industry. Intercity bus companies were also affected. The number of passengers for intercity buses for Group 1 routes decreased by 15%, which is less severe than the decrease in the airline industry. The utilization of buses for Group 2 route, which can be partially replaced with high-speed trains, decreased by 13%, while there is no evidence suggesting that the airline industry lost in those routes. These results imply that consumers who take flights for the routes with direct air connections, are not willing to switch to high-speed trains if they have to transfer, but some consumers who take buses for the routes with direct bus connections, are willing to switch to high-speed trains, putting up with a transfer.

I also found that the impact of high-speed trains varies within rail services. Each type of rail service was differentially affected. Since high-speed trains are more closely substitutable to Sae-ma-eul trains, the ridership of Sae-ma-eul trains has stronger negative impact from high-speed train introduction in the routes where Sae-ma-eul trains directly compete with high-speed trains. Therefore, Sae-ma-eul trains' ridership in those routes decreased by 30%, while Mu-gung-hwa train's ridership for the routes with high-speed train connections remained the same. Mu-gung-hwa and Sae-ma-eul trains' ridership for the routes which can be partially replaced with high-speed trains, decreased by 18% and 14% respectively. Mu-gung-hwa and Sae-ma-eul trains ridership for the routes that are located along branch lines of a high-speed rail line increased by 32% and 43% respectively.

A critical limitation of this model is that it requires caution when interpreting the results. Unlike typical natural experiments, the introduction of high-speed train may affect broader range of regions and industries, thus it is challenging to clearly define the control group and the treatment group. According to the evidence shown in Chapter 3, some routes without any high-speed train connections or potential direct and indirect influence, underwent schedule changes after high-speed train introduction. In this sense, it might be appropriate to interpret the estimated coefficients based on the reduced form models as the effect on "the treatment groups" relative to "the control groups" rather than the causal effect.

Another limitation of these results is that it shows only the overall impact of high-speed train introduction. After high-speed trains were introduced, other changes such as service schedule adjustment ensued, and the mixture of these changes resulted in the overall changes in utilization of the entire transportation industry. However, the fixed effect model or the difference in difference cannot disentangle the impact of high-speed train introduction itself from that of ensuing changes, which raises the necessity of structural model for travel demand.

In order to identify the welfare changes attributable to high-speed train introduction itself, in Chapter 3, I discussed the structural demand model and examine the impact of introducing high-speed trains on consumer welfare, taking the effect from the ensuing changes into account. My analysis treated the rail company's choice of train schedules as endogenous. After comparing the consumer surplus resulting from a set of products offered to consumers before and after high-speed train introduction, this thesis yields the implications in consumer surplus. I discussed in detail a rich analysis of consumer welfare changes after the introduction of high-speed trains and of the indirect welfare changes resulting from changes in the firm's product selection.

In Chapter 3, I addressed the demand for passenger travel and the effect on consumer surplus resulting from the introduction of high-speed trains and the ensuing changes in train schedules, taking changes in rail company's product selection into account. I estimated a model of travel demand, which incorporates consumers' heterogeneous preferences over travel schedules into a standard discrete choice model. In order to reflect the fact that the rail company may control product characteristics such as train schedules, I treated rail company's choice of train schedule as endogenous and instrument for it in the estimation. To study the effects of both new product introduction and the ensuing changes in product characteristics on consumer welfare, I performed counterfactual analyses to separately quantify the gains resulting from introducing high-speed trains and the welfare changes resulting from the rail company's schedule adjustments. The comparison of consumer surplus resulting from a set of products offered to consumers before and after high-speed train introduction leads to the implications in consumer surplus.

My results show that newly introduced high-speed trains had differential effects on consumers, and that the ensuing changes in train schedules also indirectly affects consumer surplus. The changes in consumer surplus within a market depended on availability of high-speed train. In order to investigate the effect, which varies across heterogeneous consumers, I partitioned markets into three groups based on the availability of high-speed trains in consumers' choice sets. Group 1 consumers who travel between two cities connected by high-speed trains benefited from the new product, but 50% of the gains were offset by the changes in the set of products offered to those consumers. On the other hand, Group 2 consumers, who travel along high-speed rail lines but do not have high-speed trains in their choice sets, suffered significant welfare losses from a reduction in frequency of non-high-speed trains. Group 3 consumers who travel between two cities that are not located along high-speed rail lines, experienced an increased number of trains scheduled, thus substantially increasing consumer surplus. Overall, the losses for Group 2 consumers outweighed the gains resulting from high-speed trains being made available to Group 1 consumers. However, the consumer surplus for Group 3 consumers increased due to the increased schedule frequencies; the increase incidentally made up for the losses for Group 2. The overall consumer surplus after high-speed train introduction increased; however, that increase was not nearly as substantial as the gains directly resulting from the introduction of high-speed trains because of the losses incurred by groups to which high-speed trains were not made available. These results highlight the importance of accounting for changes in existing products when analyzing the impact of new product entry on consumers.

These results are consistent with the findings in Chapter 2. The increases in rail ridership for Group 1 route came from high-speed trains despite the decreased ridership of Mu-gung-hwa and Sae-ma-eul. This is consistent with the results suggesting that consumer surplus increased in Group 1 due to high-speed train introduction itself despite the losses due to reduced schedule frequency. As expected, the schedule reduction of conventional trains in Group 2 led to the decreased ridership, which is shown in Chapter 2. Note that Group 3 in Chapter 3 contains both Group 3 and the control group defined in Chapter 2. Thus the changes of train schedule frequency shown in Table 3.6 reflect the changes of train schedule frequency in Group 3 and the control group defined in Chapter 2. Although the changes for Group 3 is less noticeable in Table Table 3.6 due to the differences in a way to define groups, we can conclude that the increased train frequency contributed to the increased ridership for Group 3 routes shown in Chapter 2.

There is, as always, a caveat. This approach presumes that consumers had the same demand over product characteristics regardless of the existence of the new product. Since the estimated demand is only based on the revealed preferences observed for the periods after the introduction, the counterfactual consumer surplus is valid only if the functional form of the demand is stable as we move away from the center of the data. One weakness of my work is that the structural model does not incorporate the changes occurred in other industries due to data constraints although other transportation industries are closely related according to Chapter 2. The lack of the supply side model which is desired for the full-blown cost benefit analysis is also a limitation of this work, and this leaves a scope for improvement and further studies.

Another limitation of this study is that it only considers the change within the transportation industry. Since transportation system is a key infrastructure for development, the improvement in transportation connections could bring much larger impact, which is not necessarily limited to the transportation industry. However, the potential contribution on various industries cannot be accounted in my current work due to the nature of the framework. The impact from high-speed train in a broader sense can also be explored in future studies.

My research calls attention to the impact on consumer welfare from new product introduction and the subsequent changes. Such subsequent changes may be due to the reactions of economic agents in the related industries or the industrial circumstance. The impact from those changes is neither limited to one industry nor restricted to a specific group of consumers. Although the subsequent changes may result in substantial influence on consumer surplus, the scope of investigation can be easily restricted to one specific industry or a particular group of consumers. Such restricted scope or the understated impact from the subsequent changes can lead to the biased results of welfare implication. My results emphasize the importance of accounting the impact of ensuing changes after new product introduction and urge more careful investigation regarding the benefit of new product introduction when one evaluates a new product entry.

This study also provokes a discussion regarding government spending. As expected, the construction of high-speed rail lines was costly and Korean government allocated enormous budget, which was levied from the entire tax payers. However, the benefits tend to be concentrated in a few markets despite diffused costs. The findings of this research can be applied to government's investment decision on other industries too. Whether an investment decision is appropriate depends not only on the direct impact from the investment but also on its indirect impact. Therefore, a thorough investigation regarding the benefit of government spending and its wider impact and an in-depth discussion is essential for better decisions on government's investment. The influence on the people who are seemingly unaffected should also be taken into account.

Appendix A

Alternative Assumption on Market Size

The numbers of airline passengers for each route within a month and the numbers of rail passengers for each route within a month are accurately observed and provided by Korea Airports Corporation(KAC) and Korail respectively.

I did not observe the number of inter-city bus passengers and auto travelers for each route, which I did for domestic flights. Instead, I took the monthly-aggregated numbers of inter-city bus passengers throughout the country from the Statistical Yearbook of Land, Transport & Maritime Affairs, and combined them with the numbers of households per city from from Korean Statistical Information Service, KOSIS to infer the number of travelers using inter-city buses or cars. First, to allow disaggregation of the numbers of bus passengers at the city-pair level, I imposed two assumptions: i) inter-city buses are available between all pairs of cities ii) the number of passengers is proportional to the geometric average of two respective cities' populations. [1][2] Assumption (ii) implies that the percentage of travelers using buses among the geometric average of two cities population is constant for all the city pairs.[3] Second, I inferred the number of auto travelers using the geometric averages of the number of cars owned in the two respective cities.

The assumptions discussed above are very limiting, and they may be unrealistic since the geometric averages of populations might not have a strong linear relationship with the respective numbers of bus travelers. It is also true that the proportion of bus travelers in a given market mt, among all bus travelers in period t, only depends on the populations of two cities', although other factors such as distance between two cities or convenience of bus connection could also be important. Similarly, the number of cars owned might not have a strong linear relationship with the number of car travelers when considering all routes. I imposed these assumptions and use the sum of the monthly aggregated number of rail and airline passengers for each route, bus travelers disaggregated at city pair level and auto travelers constructed above as the market size for the secondary specification. Rail, Air, Bus Passengers + Car Ownership in Table 3.2 shows the summary of this market size variable.

---

[1] Data used in the estimation covers 86 cities and there are more than 150 bus terminals throughout the country.

[2] obtained from Korean Statistical Information Service, KOSIS

[3] number of travelers using bus in $mt =$ (number of travelers using bus throughout the country in t) $\times \frac{\text{geometric average of two cities population in } mt}{\sum_m \text{geometric average of two cities population in } mt}$

Bibliography

Acutt, M., & Dodgson, J. (1996). Cross-elasticities of demand for travel. *Transport policy*, *2*(4), 271-277.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics*. Princeton University Press.

Ben-Akiva, M., & Watanatada, T. (1981). Structural analysis of discrete data with econometric analysis. In C. F. Manski & D. McFadden (Eds.), (p. 320-342). The MIT Press.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models*. Unpublished doctoral dissertation, Department of Civil Engineering, MIT, Cambridge, MA.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *Rand journal of economics*, *25*(2), 242-262.

Berry, S., Carnall, M., & Spiller, P. (2006). Airline hubbing, costs and demand. In D. Lee (Ed.), (Vol. vol1. Competition Policy and Anti-Trust, p. 183-214). Elsevier Press.

Berry, S., & Jia, P. (2010, August). Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, *2*(3), 1-43.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *60*(4), 889-917.

Berry, S., Levinsohn, J., & Pakes, A. (2004). Estimating differentiated products demand systems from a combination of micro and macro data: The market for new vehicles. *Journal of political economy*, *112*(1), 68-105.

Borenstein, S., & Netz, J. (1999). Why do all the flights leave at 8am?: Competition and departure-time differentiation in airline market. *International journal of industrial organization*, *17*, 611-640.

Brueckner, J. K., & Flores-Fillol, R. (2006). Airline schedule competition: Product quality choice in a duopoly model. *CESifo working paper No.1731*.

Chen, Y., & Riordan, M. H. (2008). Price-increasing competition. *The Rand journal of economics*, *39*(4), 1042-1058.

Douglas, G. W., & Miller, J. C. (1974). Quality competition, industry equilibrium, and efficiency in the price-constrained airline market. *The American Economic Review*, *64*(4), 657-669.

Eizenberg, A. (2011). Upstream innovation and product variety in the u.s. home pc market. *SSRN Working Paper Series*.

Fan, Y. (2009). *Market structure and product quality in the u.s. daily newspaper market*. Unpublished doctoral dissertation, Yale University.

Gabszewicz, J. J., Shaked, A., Sutton, J., & Thisse, J.-F. (1986). Segmenting the market: The monopolist's optimal product mix. *Journal of economic theory*, *39*, 273-289.

Gagnepain, P., & Ivaldi, M. (2002). Incentive regulatory policies: the case of public transit systems in France. *Rand journal of economics*, *33*(4).

Gaudry, M. (1975). An aggregate time-series analysis of urban transit demand: The Montreal case. *Transportation Research*, *9*, 249-258.

Hausman, J. (1996). Valuation of new goods under perfect and imperfect competition. In T. F. Bresnahan & R. J. Gordon (Eds.), *The economics of new goods*

(p. 209-247). University of Chicago Press.

Ho, K. (2006). The welfare effects of restricted hospital choice in the u.s. medical care market. *Journal of Applied Econometrics*, *21*, 1039– 1079.

Ivaldi, M., & Vibes, C. (2005). Intermodal and intramodal competition in the long-haul passenger transport market. *IDEI Report*(9).

Jones, I., & Nichols, A. (1983). The demand for inter-city rail travel in the United Kingdom. *Journal of transport economics and policy*.

Koppelman, F. S. (2006). Handbook of transport modeling. In D. A. Hensher & K. J. Button (Eds.), (chap. Closed Form Discrete Choice Models). Elsevier Science Ltd.

Koppelman, F. S., Coldren, G. M., & Parker, R. A. (2008). Schedule delay impacts on air-travel itinerary demand. *Transporation Research Part B*, *42*, 263-273.

Mandel, B., Gaudry, M., & Rothengatter, W. (1994). Linear or nonlinear utility functions in logit models? the impact on German high-speed rail demand forecasts. *Transportation research part B*, *28B*(2), 91-101.

McFadden, D. (1973). Frontiers of econometrics. In P. Zarembka (Ed.), (p. 105-142). Academic Press.

Mcfadden, D. (1978). The theory and practice of disaggregate demand forecasting for various modes of urban transportation. In W. F. Brown, R. B. Dial, D. S. Gendell, & E. Weiner (Eds.), *Emerging transportation planning methods* (p. 1-28). Department of Transportation DOT-RSPA-DPB-50-78-2.

Mcfadden, D. (2001). Disaggregate Behavioral Travel Demand's RUM Side A 30-Year Retrospective. In D. Henshner (Ed.), *The leading edge of travel behaviour research.* Pergamon Press.

Miller, J. C. (1972). A time of day model for aircraft scheduling. *Transportation Science*, *6*(3), 221-246.

Morrison, S. A., & Winston, C. (1985). An econometric analysis of the demand for intercity passenger transportation. In T. E. Keeler (Ed.), *Research in transportation economics* (Vol. 2). JAI Press Inc.

Nevo, A. (2000). A practitioner's guide to estimation of random coefficients logit models of demand. *Journal of economics & management strategy*, *9*(4), 513-548.

Nevo, A. (2003). New products, quality changes, and welfare measures computed from estimated demand systems. *Review of Economics and Statistics*, *85*(2), 266-275.

Park, Y., & Ha, H.-K. (2006). Analysis of the impact of high-speed railroad service on air transport demand. *Transportation research part E*, *42*, 95-104.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political economy*, *110*(4).

Rhoades, D., Williams, M., & Green, D. (2006). Imperfect substitutes:competitive analysis failure in U.S. intercity passenger rail. *World review of intermodal transportation research*, *1*(1), 82-93.

Richards, M. G., & Ben-Akiva, M. E. (1975). *A disaggregate travel demand model*. Saxon house and Lexington books.

Small, K. A., & Winston, C. (1999). The demand for transportation: Models and Applications. In J. A. G.-I. nez, W. B. Tye, & C. Winston (Eds.), *Essays in transportation economics and policy.* Brookings Institution Press.

Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *The Review of economic studies*, *43*(2), 217-235.

Tirole, J. (1988). *The theory of industrial organization*. Cambridge, Massachusetts: The MIT Press. (section2.1-2.2)

Train, K. (2009). *Discrete choice methods with simulation* (Second ed.). Cambridge University Press.

Train, K. E. (1993). *Qualitative choice analysis: Theory econometrics, and an application to automobile demand*. MIT Press.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge University Press.

Trajtenberg, M. (1989). The welfare analysis of product innovations, with an application to computed tomography scanners. *The journal of political economy*, *97*(2), 444-479.

Walker, J. K., & Ben-Akiva, M. (2011). Advances in discrete choice: mixture models. In A. de Palma, R. Lindsey, E. Quinet, & R. Vickerman (Eds.), *A handbook of transport economics* (p. 160-187). Edward Elgar.

Wardman, M. (1997). Inter-urban rail demand, elasticities and competition in Great Britain: Evidence from direct demand models. *Transportation research part E*, *33*(1), 15-28.

Wardman, M. (2006). Demand for rail travel and the effects of external factors. *Transportation research part E*, *42*, 129-148.