

Analyzing the Difference between the Stated Preference and the Revealed Preference before/after the High-speed Rail Service in Korea

Jang-Ho Lee[†]

Abstract

The Korean high-speed rail (HSR) began its commercial service in 2004. This service has been created significant changes in the system of intercity passenger travels of Korea. However, the actual ridership was approximately half of the estimated one in the planning stage. In this background, this paper presents the difference between the stated preference (SP) before the HSR service and the revealed preference (RP) after it using the intercity travel mode choice models. Several meaningful differences are found in terms of the factors affecting the travel mode choice, the estimation results of model, the monetary values of time, and elasticities. While the access/egress travel time of high-speed rail is less important than in-vehicle travel time in the SP sample, they have same weight in the RP sample. Also the RP models show that the probability of choosing HSR can be decreased by the increase of the number of vehicles in household contrary to the results from the SP models. The monetary values of travel time are relatively high and the direct and cross elasticities in response to changes in level-of-service of HSR are relatively low in the RP sample. This Korean case is expected to offer referable material for preparing high-speed rail services in other countries by showing the difference between the SP and RP before/after the actual service, identifying the importance of access/egress travel time and lower direct elasticities of HSR demand.

Keywords : State preference, Revealed preference, High-speed rail, Mode choice model

1. Introduction

The Korean high-speed rail (HSR) that was referred to as KTX (Korean Train eXpress) began its commercial service on April 1, 2004. KTX runs on both the Gyeongbu (Seoul-Busan) and the Honam (Seoul-Mokpo) corridors; these two corridors form the main transportation line in Korea. The first stage of the service completed in 2004 involved electrified existing rail links in addition to newly constructed links. The second stage was completed in 2010 with newly constructed links to most lines except for some sections in the Seoul metropolitan area.

The average daily KTX users have increased from 72,300 passengers in 2004 to 154,700 passengers in

2012 [1]. This would be a result of the adjustment in the operational scheme, the increased service frequency of KTX, and the newly constructed stations in the line such as Ulsan, Gyeongju, Gimcheon, etc. This service has been created significant changes in the system of intercity passenger travels of Korea. However, the actual ridership was approximately half of the estimated one in the planning stage. Some possible causes can be explained from the prediction in the total intercity passenger travel demand and the mode choice behavior. For further reference detailed, see Lee and Chang [2].

For the estimation of intercity travel mode choice model, the revealed preference (RP) data is generally used because it depicts the current equilibrium among the travel modes and has high reliability and validity. However, the RP data only has existing alternatives. Therefore the SP data is widely used for the model estimation with new mode such as HSR in spite of the lack of reliability and validity. [3]

Hensher and Bradley [4] used a collected data as part of

[†] Corresponding author: Department of Railroad Facility Engineering, Korea National University of Transportation, Korea
E-mail : transwho@ut.ac.kr

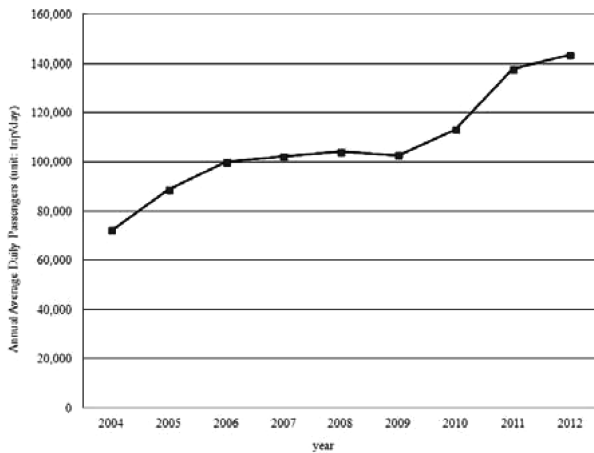


Fig. 1 Average Numbers of KTX Passengers

a study of HSR and illustrated the advantages of the RP-SP dual data strategy. Hensher [5] also used the SP data for identifying the market potential for HSR in the Sydney-Canberra corridor. Burge et. al. [6] surveyed the SP data for providing values for the different service components in the mode choice modeling process, quantifying where HSR fits in the modal choice hierarchy, and collecting background information on travelers' socio-economic characteristics, attitudes and travel preferences, and quantify the impact of these on demand for HSR in Great Britain. Lee et. al. [7] used the combined RP-SP data for the demand analysis of the Honam (Seoul-Mokpo Corridor) HSR in Korea. Park and Ha [8] analyzed the impact of HSR service on air transport using the SP data in Korea. Chou and Fu [9] also used the SP method to collect preferences of passengers regarding airline operations currently and the HSR carrier in the future in Taiwan. Willigers and Wee [10] presented for a stated choice experiment for location choices of offices in the Netherlands. They focused on how the level-of-service characteristics of railway station, in particular the presence of HSR service, influence the attractiveness of locations for specific types of offices. Outwater et. al. [11] developed the integrated statewide models that offer a comprehensive tool to forecast long and short distance travel in California using RP and SP data. It provided a robust and accurate assessment of multimodal travel at the statewide level. Also there are several research reports used the SP data for high speed rail assessment in Norway, Canada, etc. [12,13]

Most of relating studies have been targeted at investigating the demand forecasting of HSR using combined RP-SP data or SP data. In author's opinion, researchers should pay more attention to the difference between the estimation results from SP and RP data for reducing the

error in the HSR demand forecasting. In these backgrounds, this paper presents the differences between the stated preference before the HSR service in 2002 and the revealed preference after the HSR service in 2011 using the intercity travel mode choice models in terms of the factors affecting the travel mode choice, the estimation results of model, the monetary values of time, and elasticities.

2. Model Formulation

This paper formulates two types of inter-city travel mode choice models: the multinomial logit model and the mixed logit model that accommodates the observed and unobserved (to an analyst) heterogeneity across individuals in response to level-of-service (LOS) factors because an individual's response to LOS attributes affects his/her travel mode choice for a trip, and the response varies across individuals based on observed and unobserved individual characteristics (e.g., the purpose of trip, vehicle ownership, etc.).

Hensher and Button [14] defined that the heterogeneity effects refer to observed and unobserved differences across decision-makers in the intrinsic preference for a choice alternative (Preference heterogeneity) and in the sensitivity to characteristics of the choice alternatives.(Response heterogeneity) The unobserved heterogeneity effects can be accommodated by using the random-parameter structure. Bhat [15] estimated an inter-city travel mode choice model that accommodates variations in response to LOS measures due to both observed and unobserved individual characteristics. The model was applied to examine the impact of improved rail service on weekday business travel in the Toronto-Montreal corridor. Bhat [16] also formulated a mixed logit model of multi-day urban travel mode choice that accommodates variations in mode preference and response to LOS factors. The model was applied to examine the travel mode choice of workers in the San Francisco Bay area. Hensher and Greene [17] accommodated unobserved response heterogeneity, along with inter-alternative correlation, in a RP-SP study on vehicle type choices among conventional, electric, and uncompressed natural gas/liquid natural gas (UNG/LNG) vehicles in single vehicle households.

In the mixed logit model framework, the random utility term is made up of two components: a probit-like random component with a multivariate distribution, and an IID type I extreme value distributed random component. The probit-like random component captures the interdependencies among alternatives and/or response heterogeneity. Walker [18] showed that the mixed logit model with the

factor-analytic structure is a general formulation that can be used to specify all known (additive) error structures, including heteroscedasticity, nested (cross-nested) error structures, random parameters, and auto-regressive process. Thus, we specify the heterogeneity using a factor-analytic structure.

In a compact vector form, the utility function at the t -th choice occasion of individual n accommodating the heterogeneity can be written as follows:

$$U_{tn} = W_n \alpha + X_{nt} \beta_n + \varepsilon_{nt} \quad (1)$$

where

U_{tn} : a $(J_n \times 1)$ vector of utilities,

W_n : a $(J_n \times K)$ matrix of socio-demographic and trip characteristic variables including alternative specific variables,

α : a $(K \times 1)$ vector of unknown parameters,

X_{nt} : a $(J_n \times L)$ matrix of LOS variables,

β_n : a $(L \times 1)$ vector of unknown parameters,

ε_{nt} : a $(J_n \times 1)$ vector of IID type I extreme value random variables

L : the number of LOS variables,

K : the number of socio-demographic and trip characteristic variables including alternative specific variables

The second term of the Eq.(1) represents the response heterogeneity. Random parameter structure (Eq. (2)) can accommodate the observed heterogeneity as well as the unobserved heterogeneity as:

$$\beta_n = -(\beta + \gamma Z_n + T \nu_n) \quad (2)$$

where

β : a $(L \times 1)$ vector of constants in the LOS variables,

γ : a $(L \times K)$ matrix of unknown parameters,

Z_n : a $(K \times 1)$ vector of socio-demographic and trip characteristic variables for the observed heterogeneity,

T : the $(L \times L)$ lower Cholesky matrix, such that $TT' = \Sigma_\beta$ (variance-covariance matrix)

ν_n : a $(L \times 1)$ vector of IID random variables with zero mean and unit variance.

The unobserved heterogeneity model is estimated with normal distributions. T is usually specified as diagonal except when the LOS variables are closely correlated. The unknown parameters in this model are α, β, γ , and those in T . The explanatory variables W_n and X_n are observed, whereas ν_n , and ε_n are unobserved. If the factors ν_n are known, the model corresponds to a multinomial logit formulation. The conditional choice probability of alternative i at the t -th choice occasion of individual n , given ν_n is as follows.

$$L_{nt}(i|\nu_n) = \frac{\exp[\mu_t(W_n \alpha + X_{nt} \beta_n)]}{\sum_{j \in C_n} \exp[\mu_t(W_n \alpha + X_{nt} \beta_n)]} \quad (3)$$

The unconditional choice probability of alternative i at the t -th choice occasion of individual n is:

$$P_{nt}(i) = \int_{\nu} L_{nt}(i|\nu_n) dF(\nu) \quad (4)$$

where

$F(\cdot)$: the multivariate cumulative normal distribution.

For maximum likelihood estimation we need the probability of each sampled individual's sequence of observed choices. The probability of individual n 's observed sequence of choices is the product of the conditional choice probability. (Eq.(3))

$$L_n(i|\nu_n) = \prod_t L_{nt}(i|\nu_n) \quad (5)$$

The unconditional probability for the sequence of choices is:

$$P_n(i) = \int_{\nu} L_n(i|\nu_n) dF(\nu) \quad (6)$$

3. Model Estimation

The Maximum Simulated Likelihood (MSL) estimation method is generally used to estimate the mixed logit model. The choice probability of alternative i is replaced with the unbiased smooth tractable simulator:

$$\hat{P}_n(i|\theta) = \frac{1}{D} \sum_{d=1}^D L_n(i|\theta, \nu_n^d) \quad (7)$$

where

$L_n(i|\theta, \nu_n^d)$: the choice probability of alternative i given θ, ν_n^d ,

θ : the vector of unknown parameters,

ν_n^d : the d^{th} draw from the distribution of ν ,

D : the number of draws (repetitions).

This process is repeated D times for the given value of the parameter vector to be estimated, and the integrand (Eq. (6)) is approximated by averaging the computed choice probabilities in the different draws. The simulated log-likelihood function can be written as:

$$SLL(\theta) = \sum_{n=1}^N \ln[\hat{P}_n(i|\theta)] \quad (8)$$

where

N : the total number of individuals who in the sample

The parameter vector is estimated as the vector values that maximize the simulated log-likelihood function. Under rather weak regularity conditions, the maximum simulated likelihood estimator is consistent, asymptotically efficient, and asymptotically normal-distributed [19]. The MSL estimator will generally be a biased simulation of the maximum likelihood estimator because of the logarithmic transformation of the choice probabilities in the log-likelihood function. The bias of the MSL estimator decreases as the number of repetitions (D) increase [20]. Brownstone and Train [21] have shown the bias to be rather negligible with 250 repetitions. Since the mixed logit model in this paper used the normal distributions, the empirical analysis was carried out for 2500 times.

Bhat [22] has shown that the Halton simulation method outperforms the pseudo random Monte Carlo simulation method for mixed logit model estimation. Bhat [23] also described a problem that the standard Halton sequence defined by large primes can be highly correlated with each other over large portions of the sequence for simulation of high-dimensional integrals and suggested an effective solution, the scrambled Halton sequence. In this paper, the standard Halton sequence is used because relatively low dimensional integrals are simulated.

To decrease estimation time, we analytically programmed the gradients of the simulated log-likelihood function with respect to the parameter and approximated the Hessian (the second derivatives) with the BHHH technique [20] :

$$B = \sum_{n=1}^N \left(\frac{\partial \ln P_n(\theta)}{\partial \theta} \right) \left(\frac{\partial \ln P_n(\theta)}{\partial \theta} \right)', \quad (9)$$

The score was evaluated per individual and computed with the simulated scores for the MSL estimation. The estimation and computations in the paper were carried out using the Microsoft Visual C++ programming language on a personal computer

4. Estimation Results

4.1 Data Sources

The SP data used in the analysis were assembled by the Korean Society of Transportation in 2002 before the HSR service to develop travel demand models to forecast the future HSR demand in the HoNam corridor (Seoul - Iksan, Gwangju, MokPo corridor) with choice-based sampling method [24]. Also, the RP data were collected by the Korea Transport Institute in 2011 after the HSR service to

develop travel demand models to modify the guidance for the feasibility survey of railway construction project with same sampling method [25]. The sample used consists of 2,078 SP data in 2002 and 2,538 RP data in 2011.

All the data include socio-demographic and general trip-making characteristics of the traveler. (e.g., the purpose of trip, travel group size, origin, and destination cities, choice set, children or baggage dummy, income level, number of vehicles in household, gender, age, etc.) The LOS variables are access/egress travel time, access/egress travel cost, in-vehicle travel time, and travel cost (fare). The universal choice set in the SP and RP sample consists of five travel modes: car, bus, high-speed rail (HSR), conventional rail (CNR), and air. Each traveler has his/her own choice set.

Especially, the SP survey was to obtain the data to model the travel behavior responses of the HoNam corridor travelers to changes in the HSR in-vehicle time and its cost. The other travel conditions were hypothesized as the same as the current travel conditions in 2002. Thus, the access time and cost of the HSR alternative were assumed to be equal to those of the existing conventional rail alternative. The experimental design for the SP survey generated 6 scenarios of the HSR alternative. (2 levels of fare \times 3 levels of in-vehicle time)

- HSR fare: equivalent to 65% and 90% of the airfare.
- HSR in-vehicle time: each case of journey speed 150 km/hr, 180 km/hr, and 210 km/hr.

The RP survey was to obtain the information on the current trip such as perceived travel times and costs of available alternatives in 2011 after the HSR service in Korea. So the access/egress travel time and cost of the HSR alternative could be different to those of the conventional rail alternative in contrast with the SP data in 2002.

Table 1 presents the availability and choice share of each mode by data. The share of choosing the HSR in the RP

Table 1 Availability and Choice Shares of Alternatives

Alternatives	SP Sample (Before the HSR service)		RP Sample (After the HSR service)	
	Availability Shares	Choice Shares	Availability Shares	Choice Shares
Car	0.782	0.351	1.000	0.424
Bus	0.934	0.147	1.000	0.202
HSR	1.000	0.315	0.734	0.134
CNR	1.000	0.137	1.000	0.182
Air	0.355	0.050	0.134	0.059
Sample size	2,078		2,538	

Table 2 Socio-demographic and Trip Characteristics in the SP Sample

Average Monthly Household Income (Unit: million Won)		Travel Group Size		Purpose of Trip		Number of Vehicles per Household		Age	
Range ¹	Share	Size	Share	Purpose	Share	Number	Share	Range	Share
< 1.5	0.146	1	0.479	Business	0.366	0	0.085	20-29	0.342
1.5-2.0	0.190	2	0.243	Visit	0.391	1	0.651	30-39	0.330
2.0-2.6	0.219	3	0.097	Tour	0.093	2	0.221	40-49	0.226
2.6-3.5	0.245	4	0.111	Others	0.151	3+	0.044	50 <	0.102
3.5 <	0.201	5+	0.070						

¹ < 1.5: lower 25%, 1.5-2.0: lower 25-45%, 2.0-2.6: upper 35-55%, 2.6-3.5: upper 15-35%, 3.5 <: upper 15% of the population

Table 3 Socio-demographic and Trip Characteristics in the RP Sample

Average Monthly Household Income (Unit: million Won)		Travel Group Size		Purpose of Trip		Number of Vehicles per Household		Age	
Range ¹	Share	Size	Share	Purpose	Share	Number	Share	Range	Share
< 2.0	0.017	1	0.240	Business	0.365	0	0.121	20-29	0.098
2.0-3.0	0.044	2	0.360	Visit	0.286	1	0.832	30-39	0.315
3.0-4.0	0.370	3	0.134	Tour	0.319	2	0.044	40-49	0.376
4.0-5.0	0.367	4	0.209	Others	0.030	3+	0.003	50 <	0.211
5.0 <	0.202	5+	0.057						

¹ < 1.5: lower 25%, 1.5-2.0: lower 25-45%, 2.0-2.6: upper 35-55%, 2.6-3.5: upper 15-35%, 3.5 <: upper 15% of the population

Table 4 Data-fit Statistics

	SP-MNL	SP MMNL	RP MNL	RP MMNL
Log-likelihood at zero	-2887.64	-2887.64	-3400.16	-3400.16
Log-likelihood at convergence	-2199.96	-2176.44	-2873.32	-2862.50
Number of Parameters	26	25	15	17
Rho-squared	0.2381	0.2463	0.1549	0.1581
Rho-bar-squared	0.2291	0.2376	0.1505	0.1531

sample after the HSR service is considerably lower than that in the SP sample before the HSR service, and the shares of choosing the car and the bus in the RP sample are slightly higher than that in the SP sample.

Table 2 and Table 3 present the descriptive statistics for socio-demographic and trip characteristics in the SP and RP sample. The sample share of traveling alone in the RP sample is lower than that in the SP sample, while the sample share of traveling with 4 people in the RP sample is higher than that in the SP sample. Also the share of sample for sightseeing is higher in the RP sample.

Table 4 provides the data-fit statistics of the various models. There is an improvement in data-fit statistics when one introduces the heterogeneity. The rho-bar-squared value increases from 0.2291 to 0.2376 in the SP sample, from 0.1505 to 0.1531 in the RP sample. These results show the need to accommodate the heterogeneity.

4.2 Models with SP Sample

Table 5 shows the estimation results of the two models with the SP sample before the actual HSR service. The first model is the multinomial logit (SP-MNL) model, the second is the mixed logit (SP-MMNL) model accommodating observed and unobserved heterogeneity.

The effects of socio-demographic variables indicate that individuals in a household with a high number of vehicles are not to prefer the conventional rail alternative in both models. In the SP-MNL model, they are likely to prefer the car and the air. While those who travel on business relatively prefer the air and not prefer the bus and the conventional rail in the SP-MNL model, there are no statistical significances in the SP-MMNL model. An individual traveling in a group does not prefer the bus and the conventional rail. The bus, HSR, and air are less attractive to the individuals who have baggage or children due to the

Table 5 Estimation Results in the SP Sample

Variables	SP MNL		SP MMNL	
	Parameter	t-statistics	Parameter	t-statistics
Alternative Specific Constants				
Car	1.8682	5.38	1.8325	5.44
Bus	1.8949	4.56	1.4730	3.55
HSR	1.5252	5.19	1.3151	5.38
CNR	3.3825	8.42	2.8892	7.08
Socio-demographic Variables				
Vehicles in household				
Car	0.1325	1.57	-	-
CNR	-0.4492	-3.71	-0.4449	-3.55
Air	0.2490	1.77	-	-
Travel group size				
Bus	-0.6256	-6.88	-0.6678	-7.73
CNR	-0.5288	-6.54	-0.5472	-6.25
Baggage/Children				
Bus	-0.5296	-2.77	-0.5712	-2.54
HSR	-0.8837	-6.56	-0.9848	-6.20
Air	-2.4209	-3.88	-3.3535	-3.75
Business trip				
Bus	-0.5845	-3.81	-	-
CNR	-1.2629	-6.89	-	-
Air	0.9127	3.72	-	-
Higher income group ¹				
Car	0.9335	4.03	1.1284	3.94
Bus	0.4643	1.74	0.5908	1.98
HSR	1.2145	5.46	1.3571	5.69
Level of Service Variables				
Access/egress time (in 10 mins.)				
Constant	-0.0351	-2.63	-0.1006	-4.27
Standard Deviation	-	-	0.2501	6.91
Access/egress cost (in 1000 Won)				
Bus	-0.1540	-5.03	-0.1774	-5.35
HSR and Air	-0.0417	-2.70	-0.0844	-2.74
CNR	-0.0774	-3.10	-0.0487	-2.60
In-vehicle time of Car (in 10 mins.)				
Constant	-0.0844	-10.37	-0.1074	-8.12
Business trip	-	-	-0.0615	-6.05
In-vehicle time of Bus (in 10 mins.)				
Constant	-0.0572	-5.09	-0.0553	-4.01
Business trip	-	-	-0.0850	-8.34
In-vehicle time of HSR, CNR, and Air (in 10 mins.)				
Constant	-0.1282	-11.56	-0.1234	-9.28
Business trip	-	-	-0.1302	-8.36
Travel cost (Fare) (in 1000 Won)	-0.0452	-12.72	-0.0638	-11.79

¹Upper 15% income group (Over 3.5 million Won in average monthly household income)

²(-): Data not applicable or Parameter not statistically significant

burden of transferring at the bus terminal, station, and airport. An individual in the higher income group (Upper 15% of the population) prefers the car, HSR, and air in both models.

Among the LOS variables, the results show the expected

negative effects of travel time and cost. All the standard deviation parameters representing the unobserved heterogeneity are not statistically significant. In particular, the travelers' response to access/egress travel time factor varies across the individuals. The parameters representing the

Table 6 Estimation Results in the RP Sample

Variables	RP MNL		RP MMNL	
	Parameter	t-statistics	Parameter	t-statistics
Alternative Specific Constants				
Car	-4.1250	-5.41	-5.8885	-3.55
Bus	-3.1373	-4.10	-4.8518	-2.93
HSR	-2.1867	-2.89	-3.7173	-2.28
CNR	-3.3441	-4.39	-4.4405	-2.69
Socio-demographic Variables				
Vehicles in household				
Car	0.6004	4.89	0.7609	5.12
HSR	-0.9591	-6.43	-0.9333	-5.80
Travel group size				
Car	0.6068	10.52	0.7062	8.64
Bus	0.2123	3.83	0.3521	4.47
Air	-0.7552	-5.44	-1.2736	-3.88
Business trip				
CNR	-0.2139	-1.88	-0.2395	-1.26
Air	1.2871	4.41	1.9607	3.22
Level of Service Variables				
Total travel time of Car, Bus, and HSR (in 10 mins.)				
Constant	-0.0517	-6.58	-0.0882	-6.88
Standard deviation	-	-	0.0875	4.19
Total travel time of CNR (in 10 mins.)				
Constant	-0.0246	-3.49	-0.1092	-4.51
Standard deviation	-	-	0.0667	3.22
Total travel time of Air (in 10 mins.)				
Constant	-0.1369	-2.65	-0.2356	-2.22
Total travel cost (Fare) (in 1000 Won)				
Constant	-0.0091	-3.61	-0.0126	-3.75

1 (-): Data not applicable or Parameter not statistically significant

observed heterogeneity are also statistically significant. Business travelers are more sensitive to the in-vehicle travel time. However, the income level does not affect the observed heterogeneity. The reason can be guessed that the individuals who are in the sample tend to avoid telling their own income.

Among the LOS variables, the meaningful differences can be found between the SP models and the RP models. The access/egress travel time variables are less important than in-vehicle travel time variables in the SP sample. However, they have same weight in the RP sample. Also, there is no statistical difference between the access/egress cost variable and the travel cost variable. In the SP models, the in-vehicle travel time variable of the HSR alternative is not different with those of the conventional rail and air, while the total travel time of the HSR is not different with those of the car and the bus in the RP models. Also, it is higher than that of the conventional rail and considerably lower than that of the air. The standard deviation

parameters representing the unobserved heterogeneity in the total travel time of the car, bus, HSR and conventional rail are statistically significant. All the parameters representing the observed heterogeneity are not statistically significant.

4.3 Models with RP Sample

Table 6 presents the estimation results of the two models with the RP sample after the actual HSR service. The first model is the multinomial logit (RP-MNL) model, and the second is the mixed logit (RP-MMNL) model accommodating unobserved heterogeneity.

The effects of travelers who have baggage or children and those of individuals in the higher income group (Upper 25% of the population) are not statistically significant in the RP sample in contrast with the results from the SP sample. Individuals in a household with a high number of vehicles prefer the car alternative but not prefer the HSR alternative. The non-preference of the conventional

Table 7 Monetary Values of Travel Time

Monetary Value of Time ¹	Car	Bus	High-speed Rail	Conventional Rail	Air	
SP-MNL	14,916 (\$13.6)	10,109 (\$9.2)	22,656 (\$20.6)	22,656 (\$20.6)	22,656 (\$20.6)	
SP-MMNL	Business trip	21,146 (\$19.2)	17,565 (\$16.0)	31,752 (\$28.9)	31,752 (\$28.9)	31,752 (\$28.9)
	Non-business trip	13,446 (\$12.2)	6,924 (\$6.3)	15,450 (\$14.0)	15,450 (\$14.0)	15,450 (\$14.0)
RP-MNL	34,054 (\$31.0)	34,054 (\$31.0)	34,054 (\$31.0)	16,193 (\$14.7)	90,182 (\$82.0)	
RP-MMNL	42,036 (\$38.2)	42,036 (\$38.2)	42,036 (\$38.2)	52,062 (\$47.3)	112,319 (\$102.1)	

¹Unit: Won/hr (U.S. \$1.0 is about 1,100 Won)

rail in the SP sample is not significant in the RP sample. Also, an individual traveling in a group tends to prefer the car and the bus, not to prefer the air alternative. It can be explained by the cost for each person. The bus is the cheapest mode and the air is the most expensive mode. Furthermore, the cost of the car can be reduced by increasing the travel group size. Those who travel on business relatively tend to prefer the air but not to prefer the conventional rail. The preference of the HSR in the SP sample is not significant in the RP sample.

Among the LOS variables, the meaningful differences can be found between the SP models and the RP models. The access/egress travel time variables are less important than in-vehicle travel time variables in the SP sample. However, they have same weights in the RP sample. Also, there is no statistical difference between the access/egress cost variable and the travel cost variable.

In the SP models, the in-vehicle travel time variable of the HSR alternative is not different with those of the conventional rail and air, while the total travel time variable of the HSR is not different with those of the car and the bus in the RP models. Also, it is higher than that of the air. The standard deviation parameters representing the unobserved heterogeneity in the total travel time of the car, bus, HSR, and conventional rail are statistically significant. All the parameters representing the observed heterogeneity are not statistically significant.

4.4 Policy Implication

Table 7 presents the monetary values of time from the various models. The monetary values of time in the SP models are escalated using the inflation rate from 2002~2011 for the comparison at the same level. The monetary values of time in the SP-MMNL model are classified into “Non-business trip” and “Business trip”. The monetary values of time on a business trip are almost

twice as much as those of a non-business trip. Those from the RP models are higher than those from SP models except for that of the conventional rail.

It can be concluded that the monetary values of the conventional rail can be overestimated and those of the air can be underestimated using the SP sample. Also, the monetary values of the HSR can be lower than those of the air and higher than those of the conventional rail.

The objective of the original survey was to examine the effects of the newly constructing HSR mode. Consequently, an examination of the aggregate-level direct and cross-elasticities of the changes in the LOS attribute of the HSR alternative is focused from the estimation results. The aggregate elasticities provide the proportional changes in the expected market shares of each mode in response to a uniform percentage change in the LOS attribute of the HSR alternative.

Table 8 shows the aggregate direct and cross-elasticities of the four models in the Seoul-Busan corridor. The simulation for the aggregate forecasting was carried out for 1000 times. The aggregate direct and cross-elasticities from the RP models are less sensitive than those from the SP models. It indicates that the SP model leads to an overestimation in the choice probability of the HSR in response to an improvement in the level-of-service of HSR. Also, it shows the possibility of the cognitive incongruity with the actual response in the SP survey.

Furthermore, in the SP models, the HSR direct elasticities indicate that a reduction in the fare is a more effective means of increasing the HSR market share than a reduction in the in-vehicle travel time.

On the contrary, a reduction in the total travel time including access/egress travel time is a more effective means in the RP models. It can be supported by the empirical results of Korean case. A considerable increase of average daily KTX users from 2004 to 2012 is mainly

Table 8 Direct and Cross Elasticities in Response to Changes in Level-of-Service of HSR

		SP-MNL	SP-MMNL	RP-MNL	RP-MMNL
Decrease in Travel Time ¹ of the HSR	HSR	0.9490	1.1442	0.3837	0.4736
	Car		-1.1889		-0.2386
	Bus		-1.1758		-0.2292
	CNR	-0.9665	-1.0303	-0.3911	-0.6722
	Air		-1.2216		-0.5450
Decrease in Fare of the HSR	HSR	1.1799	1.5682	0.2387	0.2439
	Car		-1.5986		-0.2533
	Bus		-1.7619		-0.5210
	CNR	-1.2017	-1.7704	-0.2434	-0.2364
	Air		-1.5059		-0.2026

¹In-vehicle travel time in the SP models, Total travel time in the RP models

caused by the reduction of access/egress travel time from the opening of the newly constructed stations in the line and the reduction of in-vehicle travel time from the completion of the second stage in 2010.

While the SP-MMNL model predicts higher percentage decrease in the air choice probabilities in response to the reduction of the HSR in-vehicle travel time, the RP-MMNL model predicts relatively higher percentage decrease in the conventional rail choice probabilities. The same model predicts higher percentage decrease in the bus choice probabilities in response to the reduction of the HSR fare.

5. Conclusions

Most of studies for investigating the demand forecasting of high speed rail have been used the combined RP-SP data or the SP data because the high speed rail was not an existing mode before it's service in spite of the lack of reliability and validity. Therefore, it is necessary to analyze the differences between the estimation results from SP data before the actual HSR service and the RP data after it for reducing the error in the high-speed rail demand forecasting.

This paper formulates two types of models: the multinomial logit model and the mixed-logit framework that accommodates the observed and unobserved heterogeneity across individuals in response to LOS attributes. The mixed logit formulation is estimated using the maximum simulated likelihood estimation method that employs the quasi-random Halton draws.

From the estimation results of intercity travel mode choice model, the meaningful differences between the SP and the RP are presented. First, the results support the

importance of access/egress travel time and confirm the effect of the number of vehicles in household. The parameter of access/egress travel time has same weight as that of in-vehicle travel time in the RP sample unlike the relatively lower weight in the SP sample. The probability of choosing HSR can be decreased by the increase of the number of vehicles in household. Second, the results show that the monetary values of the conventional rail can be overestimated and those of the air can be underestimated in the SP sample contrary to the results from the RP sample. Therefore, the monetary values of the HSR can be lower than those of the air and higher than those of the conventional rail. Third, this paper indicates that the SP model can lead to an overestimation in the choice probability of the HSR in response to an improvement in the level-of-service of HSR. It also supports that a reduction in the total travel time including access/egress travel time can be more effective means of increasing the HSR market share than a fare reduction.

This Korean case is expected to offer referable material for preparing high-speed rail services in other countries by showing the differences between SP and RP sample, identifying the importance of access/egress travel time and lower elasticities of HSR demand.

References

1. Korea Railroad Corporation. (2012). "Statistical Yearbook of Railroad 2012."
2. Lee, J. and Chang, J. S. (2006). "Effects of High-Speed Rail Service on Shares of Intercity Passenger Ridership in South Korea", *Transportation Research Record*, No. 1943, pp.31-42.
3. Louviere, J., Hensher, D. A., and Swait, J. (2000). "Stated

- Choice Methods: Analysis and Application," Cambridge University Press.
4. Hensher, D. A. and Bradley, M. (1993). "Using Stated Response Choice Data to Enrich Revealed Preference Discrete Choice Models," *Marketing Letters*, Vol. 4, No. 2, pp. 139-151.
 5. Hensher, D. A. (1997). "A Practical Approach to Identifying the Market Potential for High-speed Rail: A Case Study in the Sydney-Canberra corridor," *Transportation Research Part A*, Vol. 31, No. 6, pp. 431-446.
 6. Burge, P., Kim, C.W., and Rohr, C. (2011). "Modeling Demand for Long-Distance Travel in Great Britain: Stated Preference Survey to support the modeling of demand for high-speed rail," RAND http://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR899.sum.pdf. Accessed July 21, 2013.
 7. Lee, J., Chon, K. S., and Park, C. (2004). "Accommodating Heterogeneity and Heteroscedasticity in Intercity Travel Mode Choice Model: Formulation and Application to HoNam, South Korea, High-Speed Rail Demand Analysis," *Transportation Research Record*, No. 1898, pp. 69-78.
 8. Park, Y. and Ha, H. (2006). "Analysis of the Impact of High-speed Railroad Service on Air Transport Demand," *Transportation Research Part E*, Vol. 42, No. 2, pp. 95-104.
 9. Chou, H. and Fu, C. (2007). "A Study for Domestic Air Passenger's Preference for High-Speed Rail Mode in Taiwan," *The Journal of Global Business Management*, Vol. 3, No.2.
 10. Willigers, J. and Wee, B. V. (2010). "High-speed Rail and Office Location Choices: A Stated Choice Experiment for the Netherlands," *Journal of Transport Geography*.
 11. Outwater, M., Tierney, K., Bradley, M., Sall, E., Kuppamm, A., and Modugula, V. (2010). "California Statewide Model for High-Speed Rail," *Journal of Choice Modeling*, Vol. 3, No. 1, pp.58-83.
 12. TEMS Inc. (2008). "Market Assessment of High Speed Rail Service in the Calgary-Edmonton Corridor."
 13. ATKINS (2012). "Norway High Speed Rail Assessment Study: Phase III Model Development Report."
 14. Hensher, D. A., and Button, K. (2000). "Handbook of Transport Modeling," Elsevier Science, pp. 5-6.
 15. Bhat, C. R. (1998). "Accommodating Variations in Responsiveness to Level of Service Measures in Travel Mode Choice Modeling," *Transportation Research Part A*, Vol. 32, pp. 495-507.
 16. Bhat, C. R. (2000). "Incorporating Observed and Unobserved Heterogeneity in Urban Work Travel Mode Choice Modeling," *Transportation Science*, Vol. 34, pp. 228-238.
 17. Hensher, D.A. and Greene, W. (2000). "Choosing Between Conventional, Electric, UNG/LNG Vehicles in Single Vehicle Households," Technical paper, Institute of Transport Studies, University of Sydney, Australia.
 18. Walker, J. (2001). "Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables," Ph. D. Dissertation, Massachusetts Institute of Technology.
 19. McFadden, D., and Train, K. (2000). "Mixed MNL models for Discrete Response," *Journal of Applied Econometrics*, Vol. 15, No. 5, pp. 447-470.
 20. Train, K. (2003). "Discrete Choice Methods with Simulation," Cambridge University Press.
 21. Brownstone, D. and Train, K. (1999). "Forecasting New Product Penetration with Flexible Substitution Patterns," *Journal of Econometrics*, Vol. 89, pp. 228-239.
 22. Bhat, C. R. (2001). "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research Part B*, Vol. 35, pp. 677-693
 23. Bhat, C. R. (2002). "Simulation of Mixed Discrete Choice Models using Randomized and Scrambled Halton Sequences," The 81st Annual Meeting of the Transportation Research Board (CD-ROM).
 24. The Korean Society of Transportation (2002). "Research and Study for the HoNam High Speed Rail Basic Construction Plan: Route and Station Selection - Survey Report."
 25. The Korea Transport Institute (2012), "A Study for Upgrading Assessment System of Railway Facility Investment (4th Ed.)," Korea Railway Network Authority.