Estimating Airport Ground Transport Mode Choice Using Market Aggregate Data

Wenyi Xia [1], University of British Columbia
Jong Hae Choi, Incheon International Airport Corporation
Kun Wang, University of International Business and Economics
Anming Zhang, University of British Columbia

1. Introduction

Airport ground transport attributes and their effects on passengers' airport access mode choice have been extensively studied in the past decades (Harvey, 1986; Ishii et al., 2009; Keumi and Murakami, 2012; Koster et al., 2011; Pels et al., 2003; Tam, et al., 2011; Windle and Dresner, 1995). However, all the existing empirical studies on airport ground transport mode choice rely exclusively on micro-level survey data. With the stated preference (SP) or revealed preference (RP) survey data, multinomial or nested-logit models are used to estimate the effects of airport ground transport attributes on passengers' mode choice. This approach could potentially be flawed in several aspects. First, surveys could be costly to conduct as a large amount of disaggregate-level data has to be collected. One has to trade off between the financial cost and estimation efficiency (with more observations). Second, the measurement error may be present, especially with SP survey where the respondents may not truly reveal their attributes and choices. Last, the survey sample may not always be representative for the general population. Selection bias may manifest if the estimation is based on a selective sample which exhibits heterogeneous characteristics.

In this study, we propose a new estimation approach which uses aggregate market-level operational data to estimate air passengers' ground transport mode choice. Unlike the abovementioned traditional survey approach, this method does not need any individual-level attributes and preferences. The estimation simply replies on those easily observed market-level data, namely the ridership, fare, schedule, in-vehicle travel time of each airport ground transport mode, and airport passenger traffic. This significantly lowers the data requirement. In addition, this empirical approach utilizes the information on the continuous airport passenger-flow and discretely scheduled airport ground transport modes, which substantially improves estimation efficiency. Specifically, the indirect utility of a choice alternative to access or egress the airport is developed to account for the effects of fare, in-vehicle travel time, schedule delay and other service quality characteristics. We estimate the choice probability of each individual passenger, and aggregate these individual probabilities into a total predicted ridership for each scheduled airport ground transport service. We estimate the utility function using nonlinear least squares, which minimizes the empirical differences between the estimated and true ridership for each scheduled airport ground transport service.

We focus on airport egress mode choice of arriving air passengers in this study, instead of airport access mode choice, because this helps to obtain unbiased estimates. For airport egress mode choice, passengers' arrival time is determined by the arrival flight, which helps to rule out passengers' strategical behavior in choosing their departure time to take airport ground transport. To implement the proposed econometric approach, data from Incheon International Airport (ICN) in Seoul are collected. The daily flight arrival data was retrieved, with information on the operating airlines, number of passengers onboard, scheduled and actual flight arrival time at ICN, and the departure airport. Three major egress ground transport modes are considered, namely the bus, KTX (Korea Train eXpress), and outside options, including private car or taxi. KTX is a high-speed rail (HSR) operated by Korail, providing fast and direct transport linking ICN with major Korean cities, such as Daejeon, Daegu and Busan. This study generates both methodological contributions and policy implications that could improve airport egress transport scheduling.

[1] Presented at the 54th Annual Meetings of the *Canadian Transportation Research Forum*, May 26-29, 2019 at Vancouver, BC

2. Econometric Model

Consider the following indirect utility function of a passenger at an airport choosing egress transport modes:

$$U_{ijmd} = V_m + \theta_1 T_{jmd} + \theta_2 f_{ijmd} + \theta_3 P_{jmd} + \theta_4 X_i + \epsilon_{ijmd} \quad (1)$$

where i: arrival flight; j: final destination of a passenger; m: airport ground transport mode; d: scheduled time of the next available mode; U_{ijmd}: indirect utility of a passenger on arriving flight i to destination j choosing mode m with the next available service scheduled at time d; V_m: mean utility of transport mode m; T_{jmd}: in-vehicle time from the airport to destination j by mode m with the next available service scheduled at time d; f_{ijmd}: schedule delay of a passenger on arriving flight i to destination j choosing mode m with the next available service scheduled at time d; P_{jmd}: ticket price to destination j by egress mode m with the next available service scheduled at time d; X_i: characteristics of arriving flight i; ϵ_{ijmd}: unobserved idiosyncratic utility shock; $\Theta = V_m, \theta_1, \theta_2, \theta_3, \theta_4$: parameters to estimate.

Assume ϵ_{ijmd} to be independent and follows Type I Extreme Value distribution, and normalize the mean utility of outside options (taxi and private cars, etc.) to zero. The closed-form probability of a passenger on arriving flight i choosing egress mode m with the next available service at time d to final destination j is defined as S_{ijmd}, with the following expression:

$$S_{ijmd} = \frac{exp(V_m + \theta_1 T_{jmd} + \theta_2 f_{ijmd} + \theta_3 P_{jmd} + \theta_4 X_i)}{1 + \sum_M exp(V_m + \theta_1 T_{jmd} + \theta_2 f_{ijmd} + \theta_3 P_{jmd} + \theta_4 X_i)} \quad (2)$$

where M: the set of airport ground transport alternatives. The estimated ridership $\hat{Q}_{jmd}$ of airport egress mode m scheduled at time d to destination j can be estimated as:

$$\hat{Q}_{jmd}(\Theta) = \sum_{i \in I} A_i \, r_j S_{ijmd} \quad (3)$$

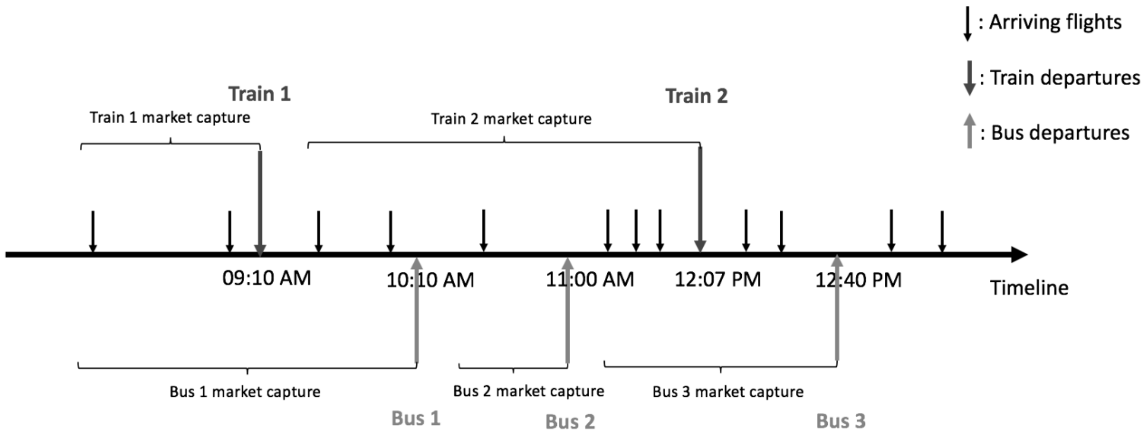
where A_i: the number of arrival passengers on flight i; r_j: the share of arrival passengers heading to destination j; I: the set of arriving flights that can take egress mode m at scheduled time d. The non-linear least squares method is adopted to minimize the sum of squared errors between the estimated ridership $\hat{Q}_{jmd}(\Theta)$ and the actual ridership Q_{jmd} on mode m to destination j at scheduled time d.

$$M_n(\Theta) = \min_{\{\Theta\}} \frac{1}{n} \sum_{j,m,d} \left(\hat{Q}_{jmd}(\Theta) - Q_{jmd}\right)^2 \quad (4)$$

where n is the total scheduled frequencies of the available egress modes to all destinations considered.

To illustrate the above proposed method, consider a simple hypothetic example as shown in Figure 1. This is an airport with two egress modes (bus and train), with fixed daily departure schedules. Fights arrive at this airport during a specific time interval as manifested in Figure 1. Compared to the discrete departure times of buses and trains, the flight arrivals are much more frequent and resemble a continuous process. For one particular arriving flight, its passengers have the chance to choose the nearest departing bus and train on the right-hand side of the flight arriving time along the timeline. Thus, the capture market (which is denoted as I in Eq. (3)) by each bus and train can be defined: it includes all the flights to the left of the departure time of each bus/train but to the right of the previous departing bus/train in Figure 1. A passenger's schedule delay is defined as the time gap between the arriving time at the bus/train platform and the next available departing bus/train. For each arrival passenger on flight i, we first derive his/her probability to choose the next bus or train by using Eq. (2) in order to obtain the variable S_{ijmd}. Next, using Eq. (3), we can aggregate each individual's probability within the bus or KTX's captive markets, weighted by the number of passengers on each arrival flight i, to obtain the estimated ridership for each scheduled bus and KTX. This is essentially to derive the variable $\hat{Q}_{jmd}(\Theta)$ in Eq. (3). This approach converts unobserved individual's choice probability at a relatively continuous dimension to the observable counts at a discrete dimension.

Figure 1. A hypothetic example to illustrate the proposed econometric approach



3. Data Description

The dataset covers an eight-day period from July 15 (Saturday) to July 22 (Saturday), 2017, which is the peak season for ICN. For each day over the study period, detailed information is collected for the arriving flights at ICN and for the airport egress modes (bus and KTX) to three destinations (Daegu, Daejeon, and Busan), which are the three largest cities along the KTX line. Seoul is excluded as an egress destination for this analysis because the ground transport between ICN and Seoul is characterized by various modes (e.g., express train, all-stop train, subway, bus, taxi, etc.) and high frequencies. For each arriving flight, data is available on the airline, flight number, origin airport, scheduled and actual flight arrival time, arriving concourse, number of passengers on board, and flight distance. For the two egress modes (i.e., bus and KTX), data is available on the schedule of each departure from ICN to the three destinations, the in-vehicle travel time, ticket price, and the ridership on each scheduled service. Table 1 summarizes the characteristics of arriving flights during the sample period. It can be seen that Northeast Asia and Southeast Asia rank the first and second in terms of the number of arriving flights, with their total share summing up to 80.7%. In addition, low-cost carriers (LCCs) account for about 28.3% arrivals.

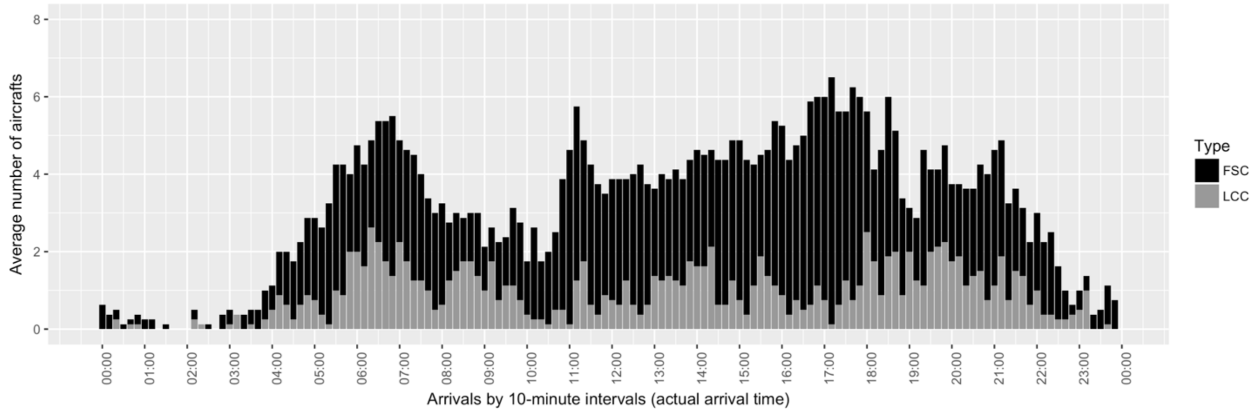
Table 1. Characteristics of arriving flights at ICN (July 15 to July 22, 2017)

	Total (sample period)	Average (per day)
No. of arriving flights	3623	453
Departure Region:		
Asia: Northeast Asia	2063 (56.9%)	258
Asia: Southeast Asia	863 (23.8%)	108
Asia: Central + South Asia	53 (1.5%)	7
Europe	200 (5.5%)	25
North + Central America	271 (7.5%)	34
Middle East	38 (1.0%)	5
Southwest Pacific	131 (3.6%)	16
Africa	4 (0.2%)	0.5
Flight Type:		
LCCs	1026 (28.3%)	128

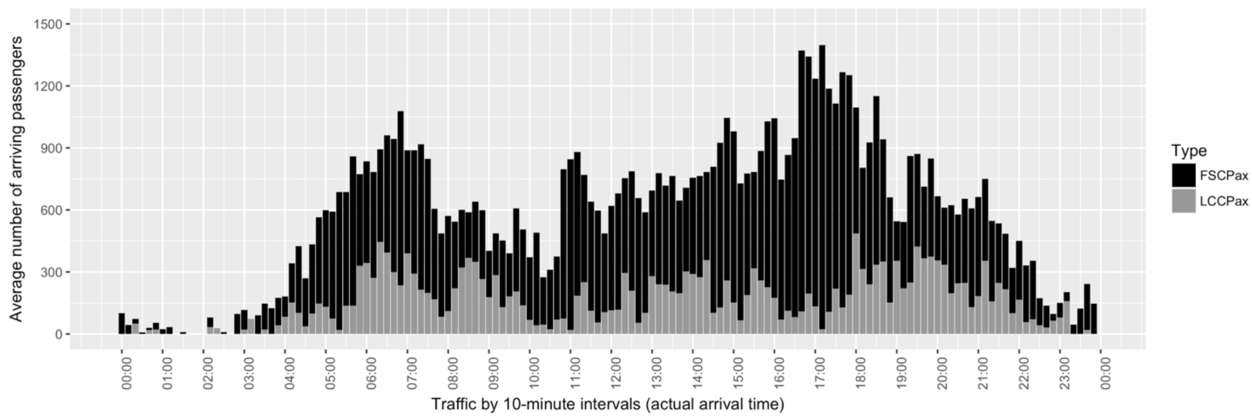
We further plot the average arriving flights and passengers along the 24-hour daily window with 10-minute intervals over the sample period in Figure 2. Peak hours at ICN occur in the mornings and late afternoons. To avoid airport congestion, LCCs use slots in unfavorable periods such as early morning and late evening to avoid peak hours for full-service carriers (FSCs) such as in late afternoons. Distribution of the passenger arrival rates at ICN over the clock time is likely to affect the schedule delay for each egress transport mode.

The flight type (LCC vs. FSC) may imply the passenger type and suggests possible heterogeneous tastes for egress modes.

Figure 2. Arrival of aircrafts and passengers at ICN with 10-minute intervals



(a). Arrival of aircrafts



(b). Arrival of passengers

Table 2 summarizes the operation statistics of egress bus and KTX at ICN. It can be seen that KTX has significant speed advantage over bus, while bus has much higher frequency, resulting in a much smaller average schedule delay for bus passengers. Thus, an average passenger has to trade-off between the travel time and schedule delay when choosing his/her egress mode. KTX, due to its superior speed advantage and higher operating cost, charges higher price than bus.

Table 2. Operation characteristics of the egress modes (bus and KTX) at ICN to the three destinations (Daejeon/Daegu/Busan). Note: The $ in this paper refers to US$.

	Earliest service	**Latest service**	**Frequencies per day**	**Avg. travel time**	**Avg. fare**	**Ridership on each service**
Bus						
Daejeon	06:00	23:30	57	3hr20min	$21	23
Daegu	06:20	23:30	42	3hr50min	$32	11
Busan	07:00	23:30	18	5hr10min	$38	9
KTX						
Daejeon	06:55	20:38	7	1hr50min	$32	22
Daegu	06:55	20:38	7	2hr40min	$49	49
Busan	06:55	20:38	6	3hr30min	$64	28

To estimate the indirect utility of passengers' egress mode choice in Eq. (1), we explicitly consider the ticket price P_{jmd}, in-vehicle travel time T_{jmd} and the schedule delay f_{ijmd} incurred by each arriving passenger. To accurately calculate a passenger's schedule delay, we try to account for detailed time segments between a passenger's landing and getting onboard the bus or the train. In practice, an expected time point needs to be estimated for a passenger to reach the bus or train platform after claiming luggage (if any) and clearing customs (for international passengers). Practically, a scheduled bus or KTX departing before a passenger arrives at the bus or KTX platform is not in the choice set of the passenger. As flights arrive at two different concourses in ICN, passengers' transfer time from the airport concourse to the bus or KTX platform are different. A passenger arrives at Concourse B (Concourse A) on average takes 47 minutes (62 minutes) after landing to reach the bus platform, and 62 minutes (77 minutes) to KTX platform. These estimated times are obtained by consulting with the terminal operation staff of ICN. The schedule delay is defined as the gap between the passenger's actual arrival time at bus or KTX platform and the scheduled departure time of the next bus or KTX. Therefore, the schedule delay is passenger-specific, depending on the flight arrival time, the concourse, and the bus or KTX schedules. In the estimation, we assume that the egress mode with longer than one-hour schedule delay will not be chosen by the arriving passenger, as it seems to be unrealistic for passengers to wait so long for a bus or KTX. However, changing the threshold will not affect the estimates qualitatively. Table 3 shows the summary statistics of the variables included in the estimation.

For the other control variables X_i, we try to capture some demographic characteristics of the passengers and their heterogeneous preferences on different egress modes. Specifically, we control for the region of the departure airport for each arriving flight. It is conjectured that passengers from different regions may have various preferences over egress modes, affected by their awareness on the egress characteristics at ICN and their habits of transport mode choices formed in their home countries. In addition, passengers are also differentiated by their airlines, providing some clues on their types. For example, passengers flying LCCs are likely to be leisure travelers, thus may prefer public transport over private modes. In addition, passengers of the Korean domestic airlines might have better information on ICN's egress transport services, which may result in different choice behaviors compared to passengers of foreign airlines. Thus, we control for the potential impact for passengers taking Korean domestic airlines, namely Korean Air (KE), Asiana Airlines (OZ) and Jeju Air (7C). The flight distance is also controlled as long-haul and short-haul passengers may have different sensitivities on the travel time and cost of airport egress modes.

4. Estimation Results

To estimate the model, we tried several alternative specifications as shown in Table 4. Specifically, Model 1 includes all the variables described in Section 3. As expected, the in-vehicle travel time, schedule delay and fare all have negative and statistically significant impacts on a passenger's airport egress transport mode choice. In particular, passengers incur substantially higher utility loss from schedule delay than in-vehicle travel. Using the compensation variation approach, the value of time (VOT) for in-vehicle travel and schedule delay can be estimated. VOT for in-vehicle travel is 46.62 \$/hour, and the cost of schedule delay is 121.83 \$/hour. That said, on average, passengers incur about 2.6 times higher cost of schedule delay than the value of in-vehicle travel time. This seems reasonable as it could be more stressful waiting at the platform than being onboard. In addition, the alternative specific constants for KTX and bus are estimated to capture the effect of passengers' preference associated with a particular mode. The estimation indicates, *ceteris paribus*, KTX is slightly preferred to bus.

The estimated VOT of in-vehicle time and schedule delay for airport egress mode seems to be comparable to the existing empirical findings. Specifically, the value of egress time can be benchmarked with the value of access time (VOAT) to the airport or the value of airborne time. Koster et al. (2011) estimate the VOAT to be 40.05 euro/hour for business passengers and 30.02 euro/hour for leisure passengers. The US DOT (Department of Transportation) also issues guidance about air passengers' VOT, and it suggests the value is between \$36.1 and \$54.1/hour as of 2015. Schumer and Maloney (2008) suggest the VOT for airborne of US air passengers to be $37.6/hour. For the value of schedule delay, there are, however, very few direct comparisons. It could be sensible for us to benchmark with the value of flight delay in that these two could be similar in nature from a passenger's perspective. For example, Yan and Winston (2014) estimate airport delay value to be $104 /hour. Landau et al. (2016) estimate the flight delay cost in the US to be $123.3/hour for leisure passengers and $286.3/hour for business passengers.

Table 3. Descriptive statistics for the sample data used in estimation

Variable	Destination	No. of Obs.	Mean	Std. Dev.	Min	Max
In-vehicle travel time by KTX (hr)	to Daejeon	56	1.907	0.071	1.816	2.050
	to Daegu	56	2.690	0.074	2.583	2.783
	to Busan	48	3.591	0.077	3.450	3.700
In-vehicle travel time by bus (hr)	to Daejeon	456	3.333	0	3.333	3.333
	to Daegu	336	3.853	0.252	3.666	4.666
	to Busan	144	5.111	0.157	5.000	5.333
Schedule delay by KTX (hr)	to Daejeon	1384	0.488	0.289	0	1
	to Daegu	1384	0.488	0.289	0	1
	to Busan	1215	0.509	0.293	0	1
Schedule delay by bus (hr)	to Daejeon	3498	0.174	0.144	0	1
	to Daegu	3465	0.274	0.210	0	1
	to Busan	2720	0.514	0.287	0	1
Fare by KTX ($)	to Daejeon	56	31.94	0	31.946	31.946
	to Daegu	56	49.38	0	49.380	49.380
	to Busan	48	63.80	0	63.805	63.805
Fare by bus ($)	to Daejeon	456	20.62	0.576	20.442	22.477
	to Daegu	336	31.85	0.855	30.973	34.867
	to Busan	144	37.55	1.138	36.991	40.708
Flight distance (1000 km)		3623	2.960	2.815	0.2620	12.095
Domestic airlines (KE, OZ, 7C)		3623	0.516	0.499	0	1
LCC		3623	0.283	0.450	0	1
Southeast Asia		3623	0.238	0.426	0	1
Europe		3623	0.055	0.228	0	1
America		3623	0.074	0.263	0	1
Middle East		3623	0.010	0.101	0	1
Southwest Pacific		3623	0.036	0.186	0	1
Mainland China		3623	0.229	0.420	0	1
Japan		3623	0.207	0.405	0	1
Other Northeast Asia		3623	0.132	0.339	0	1
South and Central Asia		3623	0.014	0.120	0	1

Note: As we restrict the schedule delay to be less than 1 hour, the maximum observed schedule delay

As shown in Table 4 Model 1, the two regional dummies (other Northeast Asia and Southwest Pacific) are marginally statistically significant. As the regional dummies are benchmarked with Mainland China, the negative sign for the two variables indicates that passengers from other Northeast Asia and Southwest Pacific are less likely to use public airport egress transport modes than passengers from Mainland China. However, for the other control variables, the coefficients are not statistically significant. This suggests that the flight distance, flight type (LCC vs. FSC), and the region of the flight origin do not have significant effects on passengers' egress mode choice.

Table 4. Estimation results

	Model 1	Model 2	Model 3	Model 4	Mode 5	Model 6
Bus constant	7.396***	7.293***	7.525***	6.651***	6.612***	6.774***
	(1.251)	(1.229)	(1.141)	(0.858)	(0.773)	(0.742)
KTX constant	7.582***	7.471***	7.704***	6.824***	6.771***	6.908***
	(1.309)	(1.273)	(1.194)	(1.024)	(0.883)	(0.864)
In-vehicle travel time	-1.719***	-1.721***	-1.740***	-1.614***	-1.622***	-1.624***
(hr)	(0.236)	(0.235)	(0.237)	(0.177)	(0.179)	(0.178)
Schedule delay	-4.492***	-4.497***	-4.641***	-4.191***	-4.174***	-4.175***
	(1.466)	(1.450)	(1.484)	(1.559)	(1.458)	(1.484)
Fare	-3.687***	-3.685***	-3.710***	-3.593***	-3.582***	-3.644***
($100)	(0.780)	(0.776)	(0.794)	(0.743)	(0.731)	(0.755)
Flight Distance	0.133	0.1362		-0.006		
(1000 km)	(0.189)	(0.188)		(0.049)		
Domestic airlines	-0.150			-0.052		
	(0.273)			(0.271)		
LCC	0.522	0.559	0.604*	(0.270)	0.471**	
	(0.373)	(0.372)	(0.360)		(0.201)	
Southeast Asia	-0.638	-0.621	-0.298			
	(0.482)	(0.486)	(0.375)			
Europe	-2.141	-2.077	-1.103			
	(1.390)	(1.383)	(0.868)			
America	-1.747	-1.822	-0.529			
	(1.806)	(1.795)	(0.528)			
Middle East	-0.272	-0.223	0.669			
	(1.472)	(1.479)	(0.855)			
Southwest Pacific	-1.171*	-1.181*	-0.895			
	(0.675)	(0.679)	(0.680)			
Japan	-0.352	-0.315	-0.367			
	(0.638)	(0.632)	(0.606)			
Other Northeast Asia	-0.917*	-0.884*	-0.754			
	(0.472)	(0.471)	(0.491)			
South and Central Asia	-1.078	-1.160	-0.586			
	(1.063)	(1.051)	(0.790)			
In-vehicle time value ($/hr)	46.62	46.70	46.90	44.92	45.28	44.57
Schedule delay cost ($/hr)	121.83	122.04	125.09	116.64	116.53	114.57

Note: 1. the coefficient of the Mainland China region is normalized to zero; 2. The number in the parenthesis is the estimated standard error of the coefficient; 3. *. **. *** represent 10%, 5% and 1% significance levels, respectively.

As a robustness check, we also tried to exclude some of these variables and re-run the estimation. Specifically, Model 2 excludes the dummy variable for domestic airlines; Model 3 excludes the flight distance variable and the variable for domestic airlines; Model 4 excludes all regional dummy variables; Model 5 excludes the flight distance variable, domestic airlines variable, and all regional dummy variables, while Model 6 only keeps the key variables. The main empirical findings are robust in terms of the effects of in-vehicle travel time, schedule delay and fare of the egress mode. The LCC variable is found to be statistically significant in Model 3 and Model 5. The positive sign of the LCC variable suggests that LCC passengers are more likely to take bus or KTX (public transport) than FSC passengers.

6. Conclusion

Airport ground transportation is an essential part of an airport's service quality. Among the existing empirical studies on passengers' airport ground transport choice, the focus has been exclusively placed on individual-level survey data and choice on airport access modes, while this study utilizes aggregate market-level data and focuses on airport egress mode choice. This approach overcomes the limitations of the traditional SP survey approach, which may include measurement errors, sample selection bias, limited sample size, and the high cost of carrying out questionnaires. The proposed empirical approach also has the advantage of improving estimation efficiency, as the method tries to match passengers' mode choice with each scheduled service (bus, train etc.), enlarging the sample size. In addition, with the estimates from the model, counterfactuals can be conducted to see if the current schedules of KTX are optimal. With the proposed empirical method that applies non-linear least squares estimation, air passengers' willingness-to-pay for the in-vehicle travel time and the cost of schedule delay (waiting for the next available service) can be obtained. we find that the cost of schedule delay is three times as large as the value of time for in-vehicle travel, indicating that airport passengers incur larger disutility when waiting for the next available egress service than being onboard. This may provide tools for ICN and airport ground transport operators to evaluate economic impacts of improving service quality and make pricing decision.

References

Harvey, G. (1986). Study of airport access mode choice. Journal of transportation Engineering, 112(5), 525-545.

Ishii, J., Jun, S., & Van Dender, K. (2009). Air travel choices in multi-airport markets. Journal of Urban Economics, 65(2), 216-227.

Keumi, C., & Murakami, H. (2012). The role of schedule delays on passengers' choice of access modes: a case study of Japan's international hub airports. Transportation Research Part E: Logistics and Transportation Review, 48(5), 1023-1031.

Koster, P., Kroes, E., & Verhoef, E. (2011). Travel time variability and airport accessibility. Transportation Research Part B: Methodological, 45(10), 1545-1559.

Landau, S., Gosling, G. D., Small, K., & Adler, T. (2016). Measuring air carrier passengers' values of time by trip component. Transportation Research Record: Journal of the Transportation Research Board, (2569), 24-31.

Pels, E., Nijkamp, P., & Rietveld, P. (2003). Access to and competition between airports: a case study for the San Francisco Bay area. Transportation Research Part A: Policy and Practice, 37(1), 71-83.

Schumer, C.E., Maloney, C.B. (2008). Your flight has been delayed again: flight delays cost passengers, airlines, and the US economy billions. The US Senate Joint Economic Committee. Retrieved from https://www.jec.senate.gov/public/_cache/files/47e8d8a7-661d-4e6b-ae72-0f1831dd1207/yourflighthasbeendelayed0.pdf

Tam, M. L., Lam, W. H., & Lo, H. P. (2011). The impact of travel time reliability and perceived service quality on airport ground access mode choice. Journal of Choice Modelling, 4(2), 49-69.

Windle, R., & Dresner, M. (1995). Airport choice in multiple-airport regions. Journal of Transportation Engineering, 121(4), 332-337.

Yan, J., & Winston, C. (2014). Can private airport competition improve runway pricing? The case of San Francisco Bay area airports. Journal of Public Economics, 115, 146-157.

U.S. Department of Transportation, Office of the Secretary of Transportation. (2016). Revised Departmental Guidance on Valuation of Travel Time in Economic Analysis. Washington, D.C., September 27, 2016