Article:

Huan, N., Hess, S. orcid.org/0000-0002-3650-2518, Yamamoto, T. et al. (1 more author) (2024) Modelling intermodal traveller behaviour in mega-city regions: simultaneous versus sequential estimation frameworks. Transportation. ISSN 0049-4488

https://doi.org/10.1007/s11116-024-10489-2

Modelling intermodal traveller behaviour in mega-city regions: Simultaneous versus sequential estimation frameworks

Ning Huan[a,b], Stephane Hess[c], Toshiyuki Yamamoto[a], Enjian Yao[b,*]

[a] Institute of Material and Systems for Sustainability, Nagoya University, Nagoya 464-8603, Japan

[b] Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing 100044, China

[c] Institute for Transport Studies & Choice Modelling Centre, University of Leeds, Leeds LS2 9JT, United Kingdom

Abstract

The sustained expansion of mega-city regions and the development of multimodal transport networks have catalysed intercity mobility, thereby restructuring regional travel demand patterns. This study aims to interpret the behaviour of intermodal travellers in a short-haul intercity context within mega-city regions. A comparative modelling framework, utilising both simultaneous and sequential estimation methods, is proposed based on stated preference survey data collected in the Beijing-Tianjin-Hebei region, China. The simultaneous estimation framework examines the integrated measurement of the perceived utility of multiple stages of travel using cross-nested logit models. In contrast, the sequential estimation framework systematically investigates the bidirectional interactions associated with the intercity mode decision and decisions related to access and egress modes in a stepwise manner. The latter quantifies the accessibility of transport hubs and destinations to assess the implicit cost of feeder trips in the intercity mode decision. It validates the sequential impact on feeder mode choice preferences. In addition to identifying behavioural determinants, the models' relative performance is assessed regarding behaviour prediction accuracy for diverse groups of travellers categorised by travel purpose, fellow traveller, baggage size, and travel frequency. Statistically, the weighted prediction errors for access, intercity, and egress mode choices are 1.12%, 1.33%, and 0.89% under the simultaneous estimation framework. In contrast, under the sequential estimation framework, these errors are reduced to 0.81%, 0.63%, and 0.50%, respectively. The results suggest the superior applicability of the latter in interpreting intermodal mobility patterns.

Keywords: intercity travel; mode choice; cross-nested logit; accessibility; behaviour prediction; urban agglomerations

1. Introduction

In the past decade, urban mobility has experienced noteworthy transformations attributed to the sustained expansion of mega-city regions and the rapid evolution of multimodal transport networks, particularly in numerous Asian, North American, and European countries (Gottmann 1961; Hall and Pain 2006). The literature typically considers mega-city regions as agglomerations of adjacent cities that are highly integrated and exhibit significant economic strength (Hall and Pain 2006). Given that mega-city regions are highly developed spatial concentrations of cities (Fang and Yu 2017), rapid urbanization is typically accompanied by a swift expansion of railway and road networks. The extensively connected transport networks significantly enhance service levels for intercity travel and facilitate the formation of new intercity mobility patterns. Within this context, understanding intermodal travel behaviour becomes foundational for forecasting the demand for regional transport systems and is crucial for expediting the operational integration of intermodal passenger transport.

This study aims interpret intermodal travel behaviour in the context of mega-city regions, with a dual focus. Firstly, it aims to identify behavioural determinants at each stage of intermodal travel. Secondly, it seeks an appropriate choice model estimation approach to characterise the multiple decisions, ultimately achieving improved predictive accuracy of behavioural outcomes. The literature typically defines intermodal travel as a journey involving two or more modes of transport, with park and ride (P&R) being a representative example. This concept has been extensively explored in the context of urban areas (Cheng and Tseng 2016; Huang et al. 2022; Meyer de Freitas et al. 2019; Wang et al. 2023). As travellers' accessibility increases in mega-city areas, intermodal travel plays a crucial role in intercity mobility services by providing more seamless solutions to enhance the traveller experience (Huan et al. 2023; Yang et al. 2022). This has attracted growing attention for analysing intercity mobility patterns and exploring the operational integration of multimodal passenger transport (Bai et al. 2021; Luo et al. 2021). Hence, this study employs a broader definition of intermodal travel to include intercity travel and its access and egress trips, interpreting it as a multi-stage mode choice behaviour, as illustrated in Figure 1.

Typically, the entire intermodal travel can be divided into three stages: the access stage from the origin to the departure transport hub, the intercity stage between the two cities, and the egress stage from the arrival transport hub to the destination. Therefore, travellers are expected to make three sets of travel mode choices, except for private car travellers who avoid feeder trips. It is noteworthy that the distance of intermodal travel within the mega-city area is usually less than

500km. Within this range, rail transport, particularly high-speed rail (HSR), holds a distinct advantage over air transport (Dobruszkes et al. 2014; Zhang et al. 2019). Hence, this study considers only rail and road transport as intercity travel modes.

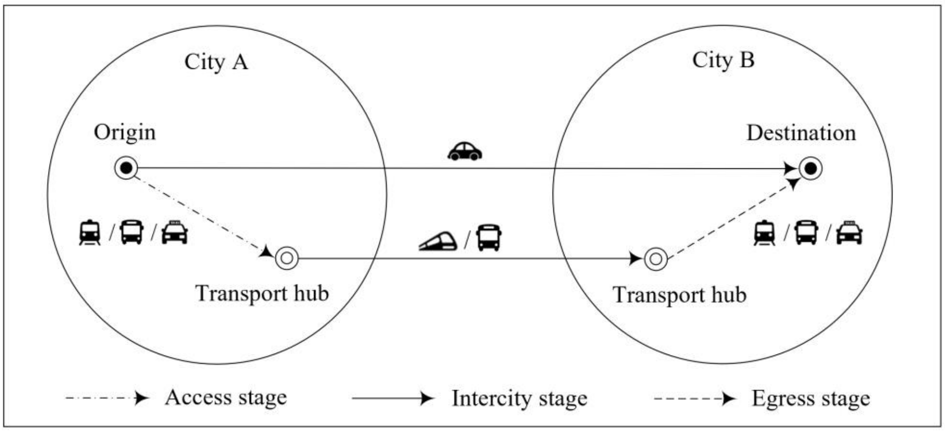


Figure 1 Illustration of multiple stages of intermodal travel

The recent emergence and development of Mobility as a Service (MaaS) further emphasises the need to understand intermodal travel behaviour (Bushell et al. 2022), especially the correlated choices between different stages of travel. Accurate prediction of travellers' intermodal choices serves as a crucial basis for providing on-demand mobility services. Although numerous studies have investigated intercity mode choice (Hess et al. 2018; Román et al. 2014) and feeder mode choice (Wen et al. 2012; Yang et al. 2019; Yang et al. 2015), past studies have generally lacked the integrated consideration of multi-stage choice behaviour. Hence, the preference differences among stages of travel are still underexplored, impeding the high-precision forecasting of intermodal travel demand patterns. To address this issue, this study implements a stated-preference (SP) survey to investigate the underlying behavioural determinants of intermodal travellers and performs model comparison analysis to demonstrate the empirical applicability of simultaneous and sequential estimation frameworks, as depicted in Figure 2.

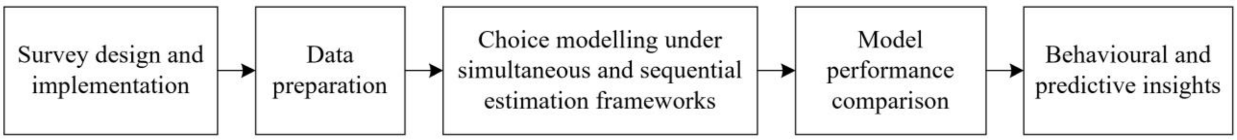


Figure 2 Illustration of research process

The remainder of this paper is structured into seven sections. Section 2 reviews past studies on analysing intercity travel behaviour. Section 3 introduces the SP survey designed for this study. Section 4 describes the survey data collected and relevant data preparation processes. Section 5 presents model specifications tailored for simultaneous and sequential estimation approaches. Section 6 reports the results of model estimation and comparative analysis. Section 7 summarises the conclusions of this study.

2. Literature review

Extensive research efforts have been dedicated to analysing the behaviour of intercity travellers over the past decade. Given that intercity travel is less frequent, more purpose-driven, and flexible in timing than intra-city travel, previous studies have predominantly concentrated on mode choice within the context of multimodal corridors within the context of multimodal corridors (Bergantino and Madio 2020; Capurso et al. 2019; Hess et al. 2018; Román et al. 2014; Zhou et al. 2020). A minority of studies have investigated travel demand generation (Llorca et al. 2018; Lu et al. 2014), destination choice (Wang et al. 2016; Yao and Morikawa 2005), and departure time choice (Chaichannawatik et al. 2019). Since the intercity mode significantly influences the route, with the exception of private cars using the road network, there are consequently few studies that address the route choice problem (Wang et al. 2014).

Regarding intercity mode choice models, previous studies have generally relied on SP surveys (Capurso et al. 2019; Hess et al. 2018; Zhou et al. 2020), revealed preference surveys (Román et al. 2014), and the combination of both (Bergantino and Madio 2020; Wong and Habib 2015). Discrete choice models, such as the multinomial logit (MNL) model, mixed logit model, and nested logit (NL) model, have been widely recognised as practical tools for analysing intercity mode decisions.

Concerning feeder modes, the literature also offers substantial references for modelling access and egress travel behaviour with urban transit stations (Rahman et al. 2022; Yang et al. 2015), railway stations (Wen et al. 2012; Yang et al. 2019; Zhen et al. 2019), and airports (Gokasar and Gunay 2017; Tam et al. 2011) being the objects of connection. Some studies have incorporated the feeder mode choice into intercity mode choice modelling. The first approach is to regard feeder modes as alternatives parallel to intercity modes. For instance, Waerden P and Waerden J (2018) developed a mixed MNL model encompassing three train-based intermodal alternatives and a private car alternative. However, the distinction between access and egress mode choice behaviour cannot be obtained, resulting in a lack of interpretation for egress mode decisions. Moreover, relevant models have consistently assumed that travellers simultaneously perceive the total utility of the intercity mode and its access mode, aligning with the essence of the simultaneous choice modelling approach. However, a demonstration of the method's applicability remains outstanding.

Instead of treating feeder modes as independent alternatives, the second approach incorporates the perceived effects of feeder trips into intercity mode decisions. The most common method is to introduce feeder trip-related explanatory variables into the utility functions of intercity modes. For instance, using continuous variables to represent travel time (Bergantino and Madio 2020; Capurso et al. 2019; Hess et al. 2018; Román et al. 2014; Wong and Habib 2015) and travel distance (Miskeen M A et al. 2013) of feeder trips, and dummy variables to represent feeder mode choice outcomes (Ranjbari et al. 2017; Wang et al. 2014; Wong and Habib 2015). However, the above approach has distinct limitations and drawbacks: firstly, feeder trip-related explanatory variables were mostly determined by the shortest time or distance route of a particular travel mode, and therefore, they cannot fully reflect the service levels of feeder trips; secondly, using access and egress mode choices as a prerequisite implies, in effect, that travellers make the intercity mode choice after deciding on feeder modes. This is undoubtedly an implicit assumption of sequential decision-making with feeder modes as the predecision, but its rationality has not been adequately justified.

Given that both simultaneous and sequential decision-making hypotheses have been embodied in existing models by default, it is valuable to evaluate the model performance from an outcome-oriented perspective: specifically, the effectiveness of behavioural models in predicting disaggregate travel demand. Theoretically, in a multiple-decision problem, simultaneous decision-making refers to making multiple decisions simultaneously, while sequential decision-making involves making successive decisions in a process (Diederich 2001). In behavioural studies, particularly in consumer behaviour research (Simonson 1990), the former is conceptually similar to simultaneous choice, in contrast to sequential or multi-stage choice. In addition to interpreting intermodal behaviour per se, this study examines the rationality of the decision-making hypothesis by comparing various model specifications and estimation approaches. The literature on choice modelling estimation techniques has suggested that, in the case of hybrid discrete choice models, simultaneous and sequential estimation results in differences in forecasting and policy evaluation (Bierlaire 2016; Raveau et al. 2010). The model estimation approach, while not precisely representing the corresponding decision-making mechanism, allows for assumptions about the decision-making sequence to be incorporated into model specifications. This enables the numerical comparison of model performance under different decision-making hypotheses, providing a feasible approach for experimental validation.

Therefore, this study proposes a comparative framework involving simultaneous and sequential estimation methods to interpret intermodal travellers' multi-stage choice behaviour comparatively. The research contribution is twofold: First, the determinants of each stage of mode decision are identified based on SP observations. Second, the applicability of two decision-making hypotheses is examined in terms of behaviour prediction performance. The findings provide implications for reducing biases in intermodal travel demand forecasts by suggesting the most statistically sound model specifications.

3. Stated-preference experiment

The SP experiments are designed based on actual intercity travel within the research area, as shown in Figure 3. This region, with a resident population of over 110 million, is one of the largest mega-city regions in China. Beijing serves as the central city, while Tianjin and Shijiazhuang, the capital of Hebei Province, are the other two key cities. The corridors from Beijing to Tianjin and Shijiazhuang are selected as intercity travel scenarios in this study.

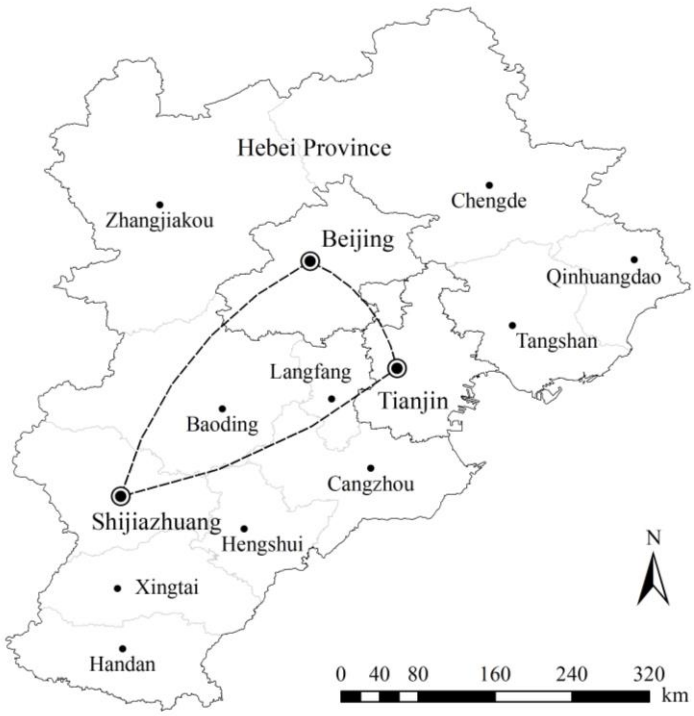


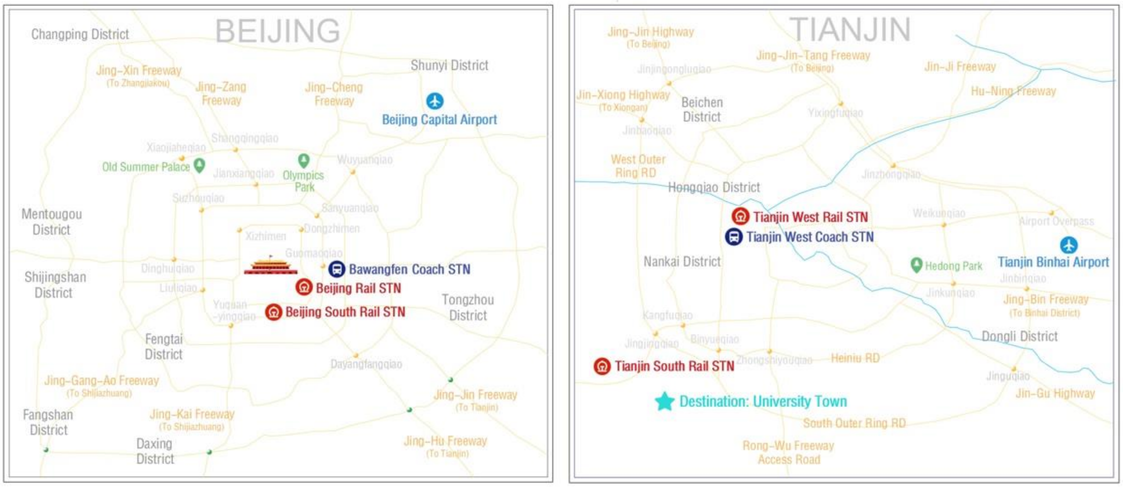
Figure 3 Spatial layout of Beijing-Tianjin-Hebei mega-city region

In the actual survey, the respondents are initially asked to specify their place of residence and their most and second most familiar travel scenarios characterised by travel purpose, fellow traveller and baggage size. Using the SP choice tasks, respondents' three-stage travel mode choices are collected in their familiar scenarios. More explicitly, Figure 4 presents a set of SP choice tasks as an example. In each SP choice task, respondents are first provided with maps of origin and destination cities to help them understand the location of transport hubs. The alternative modes for the access and egress stages include public transport, i.e., metro, bus and intermodal public transport (IPT, a combination of metro and bus), and taxi or e-hailing (ToE).

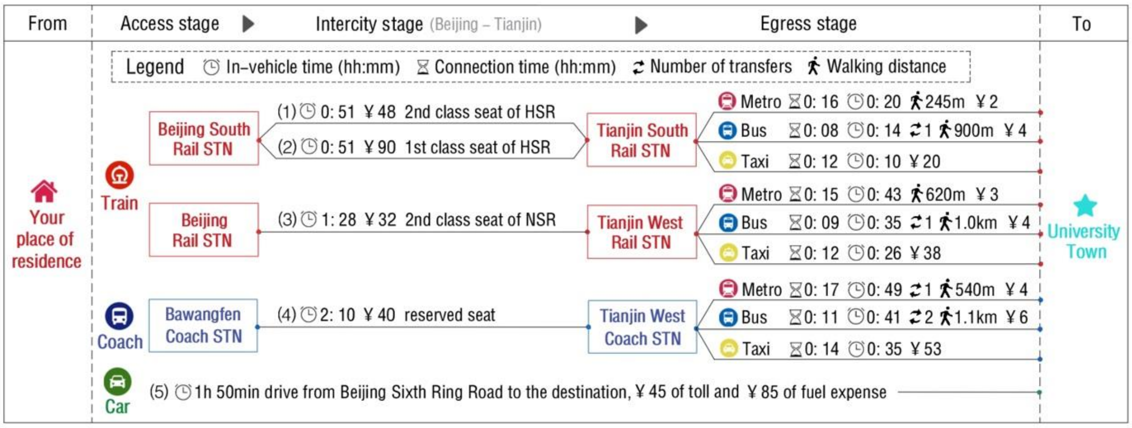
For the intercity stage, depending on the actual situation in the study area, the second- and first-class seat of HSR (HSRa and HSRb), the second-class seat of normal-speed rail (NSR), intercity coach, and private car are taken into account. Each respondent is required to complete four sets of SP choice tasks with different destinations in different cities, contributing to at most eight valid SP observations.

Please read the pictures below and complete the following questions.

The maps below depict the positions of transport hubs in Beijing and Tianjin, as well as the destination—University town in Tianjin.



The picture below displays the options for the three stages of intercity travel.



1. In the first case, *taken from the response regarding the most familiar travel scenario, please specify your preferred choice for each travel stage.

Access stage in Beijing	Intercity stage (Beijing to Tianjin)	Egress stage in Tianjin
Choose a travel mode according to your travel experience: [▾] The drop-down list includes metro, bus, metro & bus, and taxi (incl. e-hailing).	Choose a travel mode based on the provided information: [▾] The drop-down list includes (1), (2), (3), (4), and (5).	Choose a travel mode based on the provided information: [▾] The drop-down list includes metro, bus, and taxi (incl. e-hailing).

2. In the second case, *taken from the response regarding the second most familiar travel scenario, please specify your preferred choice for each travel stage.

Same as the table above.

Figure 4 Example of intermodal travel SP scenario

To prevent overwhelming respondents with excessive information, respondents are solely provided with LOS attributes related to intercity and egress travel. Given that the origin of intercity travel is set to their place of residence, respondents are expected to be more acquainted with the access trip to transport hubs and relatively capable of making informed choices, even without detailed LOS information. Consequently, respondents are tasked with selecting an access mode based on the provided map and their experience. As the access mode choice has been simplified, an additional alternative of IPT is included in the alternative set at this stage. However, IPT is set as unavailable at the egress stage to prevent overcomplicating the entire SP scenario.

The SP choice tasks are generated using an orthogonal experimental design. To avoid an excessive number of variables in the design process, hypothetical scenarios for intercity and egress travel are independently formulated. Two sets of situational variables are then randomly combined to create complete SP scenarios. In line with the study's practical design principle to enhance survey quality, hypothetical values for situational variables are assigned, with actual LOS attributes being the baseline, such as applying a 15% increase or decrease in travel time. Consequently, the variable ranges are significantly broadened, mitigating potential multicollinearity issues and enhancing the overall fit of the models. Tables 1 and 2 summarise the variables and corresponding levels used in intercity and egress travel scenarios, respectively.

Table 1 Levels of situational variables for intercity travel scenarios

<table>
<tr><th rowspan="2">Variables</th><th rowspan="2">Constraints</th><th colspan="5">Alternatives</th></tr>
<tr><th>HSRa</th><th>HSRb</th><th>NSR</th><th>Intercity coach</th><th>Private car</th></tr>
<tr><td>In-vehicle travel time (TT)</td><td>N/A</td><td colspan="2">(1) Baseline; (2) ±15%</td><td>(1) Baseline; (2) ±15%</td><td>(1) −15%; (2) 15%</td><td>(1) −15%; (2) 15%</td></tr>
<tr><td rowspan="3">Travel expense (TE)</td><td>Baseline TT</td><td>(1) −10%; (2) 10%</td><td>(1) −10%; (2) 10%</td><td>(1) −10%; (2) 10%</td><td>N/A</td><td>N/A</td></tr>
<tr><td>Low-level TT</td><td>(1) 10%; (2) 20%</td><td>(1) 10%; (2) 20%</td><td>(1) 10%; (2) 20%</td><td>(1) 10%; (2) 20%</td><td>(1) 10%; (2) 20%</td></tr>
<tr><td>High-level TT</td><td>(1) −10%; (2) −20%</td><td>(1) −10%; (2) −20%</td><td>(1) −10%; (2) −20%</td><td>(1) −10%; (2) −20%</td><td>(1) −10%; (2) −20%</td></tr>
</table>

Note. The baseline values for TT and TE are obtained from the Baidu Map Travel Planning and Navigation APIs, train and intercity coach official ticketing websites during the survey period.

The intercity travel scenario involves two situational variables for five alternatives. For the private car alternative, the sum of toll and fuel expenses is considered a single variable, even though they are separately displayed in the SP choice tasks. Several design principles are applied as constraints in defining variable levels: (1) HSRa and HSRb share the same levels for in-vehicle travel time (TT). (2) Baseline values for TT are exclusively set for rail transport based on timetables, excluding road traffic-based alternatives due to the uncertainty in TT. (3) To ensure the rationality of level crossings, the levels for travel expense (TE) depend on the levels of TT. Following general pricing principles for rail and road passenger transport services, reduced TT corresponds to low and high levels of increases in TE, while increased TT corresponds to low and high levels of decreases in TE. Baseline TT corresponds to a low level of increase or decrease in TE. Regarding the selection of an appropriate orthogonal array, TT for HSR (HSRa and HSRb) and NSR each requires two binary variables. The first binary variable signifies whether baseline or adjusted values are chosen, while the second one indicates the adjustment magnitude. TT determination for the intercity coach and private car necessitates one binary variable each. TE determination for each alternative also demands a single binary variable. Thus, a total of eleven variables, each with two levels, is needed. Consequently, twelve intercity travel scenarios are generated using the orthogonal array L12.2.11 to cover the various combinations of these variables effectively.

Table 2 Levels of situational variables for egress travel scenarios

Variables	Alternatives		
	Metro (M)	Bus (B)	ToE
Connection time	(1) M > B > ToE; (2) M > ToE > B; (3) B > M > ToE; (4) B > ToE > M; (5) ToE > M > B; (6) ToE > B > M		
In-vehicle travel time	(1) Baseline; (2) −15%; (3) 15%	(1) Baseline; (2) −15%; (3) 15%	(1) Baseline; (2) −15%; (3) 15%
Walking distance (W) Number of transfers (T) Travel expense (E)	(1) W = 1, T = 1, E = 1; (2) W = 1, T = 1, E = 0; (3) W = 1, T = 0, E = 0; (4) W = 1, T = 0, E = 1; (5) W = 0, T = 1, E = 1; (6) W = 0, T = 1, E = 0; (7) W = 0, T = 0, E = 1; (8) W = 0, T = 0, E = 0;		N/A

Note. The baseline values for the variables are obtained from Baidu Map Travel Planning and Navigation APIs during the survey period. ‘W = 1’ indicates that the walking distance for the metro is shorter than that for the bus. ‘T = 1’ indicates that the number of transfers for the metro is fewer than that for the bus. ‘E = 1’ indicates that the travel expense for the metro is lower than that for the bus.

The egress travel scenario involves five situational variables for three alternatives. Notably, the connection time encompasses the time spent walking from the train platform/intercity coach stand to the public transit stand, as well as the waiting time for egress travel modes. The waiting time component is influenced by the departure intervals of the metro and bus or the queue length for taxis at different transport hubs. In defining variable levels for connection time, walking distance, number of transfers, and travel expense, the experimental design aims to reduce complexity while preserving distinctiveness by only constraining the order of service levels for different egress modes. The specific values undergo adaptive modifications based on the acquired baseline values. As such, connection time is represented by a single six-level variable. TT for each alternative is expressed through a three-level variable. Walking distance, number of transfers, and travel expenses are characterised by an eight-level variable. The orthogonal array L18.3.6.6.1 is thus adopted. Specifically, three three-level variables are allocated to represent the eight-level variable, with one redundant experiment, obtaining a total of seventeen egress travel scenarios.

In addition to the aforementioned situational variables, the experimental design includes two implicit background variables: the number of destination cities and final destinations. Two levels are set for each of these variables. Accordingly, twelve intercity travel scenarios are divided into three groups, each comprising four scenarios corresponding to different destinations for each respondent. Furthermore, depending on the number of transport hubs involved in the intercity travel scenarios, a suitable number of egress travel scenarios are randomly drawn from the generated scenarios. As illustrated in Figure 4, respondents make two sets of choices in each SP scenario, aligning with their self-reported most and second-most familiar travel scenarios. Namely, each respondent contributes to a maximum of eight sets of intermodal choice observations.

4. Data description

This section provides a preliminary analysis of the collected samples and introduces the data preparation process for collecting LOS attributes related to respondents' access trips.

4.1 Descriptive analysis

A web-based survey was conducted from January to March 2020. A total of 2,216 questionnaires were obtained, resulting in 13,551 valid SP samples. Table 3 reports the statistical results for travel characteristics and socio-demographics.

Table 3 Descriptive statistics of respondent information

Attributes	Levels	Sample sizes	Proportions (%)
Travel purpose*	Business	664	29.96
	Non-business (tourism, family visits, medical treatments, and others)	1,552	70.04
Fellow traveller*	Alone	1,141	51.49
	Accompanied	909	41.02
	With vulnerable groups	166	7.49
Baggage size*	Carry-on baggage	1,731	78.11
	Checked baggage	485	21.89
Reimbursement of business travel expenses	1 (yes)	898	82.08
	0 (no)	196	17.92
Gender	Male	1,077	48.60

	Female	1,139	51.40
Age	≤ 25	565	25.50
	(25-50]	1,541	69.54
	> 50	110	4.96
Education	Secondary, technical schools or below	522	23.56
	Bachelor (obtained/in progress)	1,223	55.19
	Master or above (obtained/in progress)	471	21.25
Monthly income (CNY)	≤ 6k	609	27.48
	(6k-15k]	1,224	55.23
	> 15k	383	17.28
Employment	Processing and manufacturing machine operators and related production workers	36	1.62
	Clerical workers (e.g., sales clerk, hotel front desk clerk, data entry clerk)	146	6.59
	Employees in private/foreign/state-owned enterprises in professional or administrative positions	1,139	51.40
	Government or public institution employees	348	15.70
	Self-employed workers	62	2.80
	Freelancer, retiree, and others	485	21.89
Annual intercity travel frequency	≤ 2	1,010	45.58
	[3, 6)	668	30.14
	[6, 9)	264	11.91
	⩾ 9	274	12.36
Car ownership	1 (yes)	1,333	60.15
	0 (no)	883	39.85

*The statistics of these variables are based on respondents' self-reported most familiar travel scenario.

Further, the conditional probabilities of intermodal choices are computed using the collected SP observations to investigate the relationships between multi-stage travel mode choices. A comparison between Figures 5(a) and 5(b) reveals noticeable differences in intermodal travellers' access and egress mode decisions. A distinctive feature is the higher modal share of ToE in the egress trip compared to the access trip (approximately 40% versus 20%). This is likely attributed to the unfamiliarity with the arrival city and the preference for a more comfortable travel mode at the end of the journey to alleviate travel fatigue. Similar phenomena have been observed in other studies. For instance, in-vehicle travel time exhibits higher time values in the access trip than in the egress trip (Hensher and Rose 2007), and walk access is less influenced by distance compared to walk egress (Yamamoto and Komori 2010). This underscores the importance of distinguishing between access and egress mode decisions in choice modelling.

Additionally, significant variations in feeder mode preferences are identified among travellers using different intercity modes. For example, first-class seat passengers of HSR show a preference for choosing ToE in feeder trips compared to second-class seat passengers. NSR users tend to opt for public transport more than HSR users. While conditional probabilities indicated correlations in multi-stage decisions, the determinants of such behaviour remain unclear, warranting the need for further exploration through choice modelling.

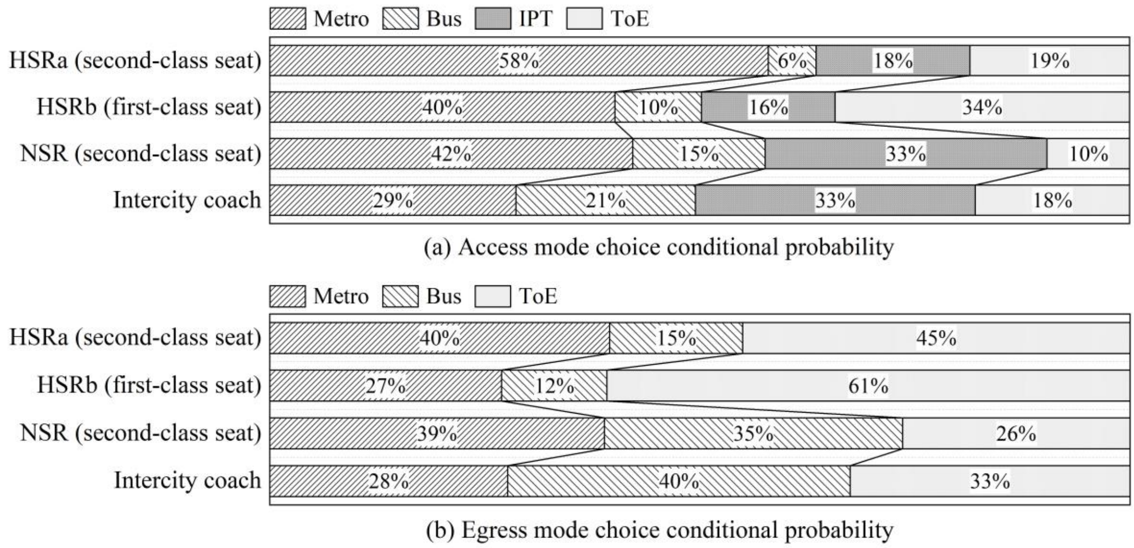


Figure 5 Conditional probabilities of multi-stage mode choices

4.2 Level-of-service attributes collection

Since the LOS information is not provided to respondents in the access part of the SP scenario, it is thus essential to collect corresponding situational attributes for modelling use. Utilising respondents' self-reported residence information and the locations of transport hubs, i.e., NSR and HSR stations, and intercity coach stations, LOS attributes are collected through the Baidu Map Travel Planning and Navigation APIs (Application Programming Interface). This process was carried out in March 2020, amid and immediately after the survey data collection was completed, to minimise potential biases introduced by disparities in data collection timing. The detailed procedures for data collection are presented in Figure 6.

Specifically, the recommended public transport routes, optimised for a balance of travel time, expenses, and the number of transfers, as well as the driving route with the shortest travel time, are employed to extract LOS attributes for the alternatives. In the case of private car travellers, pertinent attributes for the access trips from respondents' residences to the motorway entrances on the Sixth Ring Road are also taken into account.

Based on the residential data of the 2,216 respondents, the current levels of shuttle services to major transport hubs in Beijing can be inferred in terms of mean travel time, expenses, and the number of transfers, as detailed in Table 4. The statistics encompass five transport hubs, comprising three railway stations and two coach stations. The average access time by public transport ranges from approximately one to two hours, with the metro exhibiting the least travel time compared to the bus and IPT. The ToE is undoubtedly the fastest access mode, saving nearly twenty minutes on average to transport hubs, but it is also the costliest access mode.

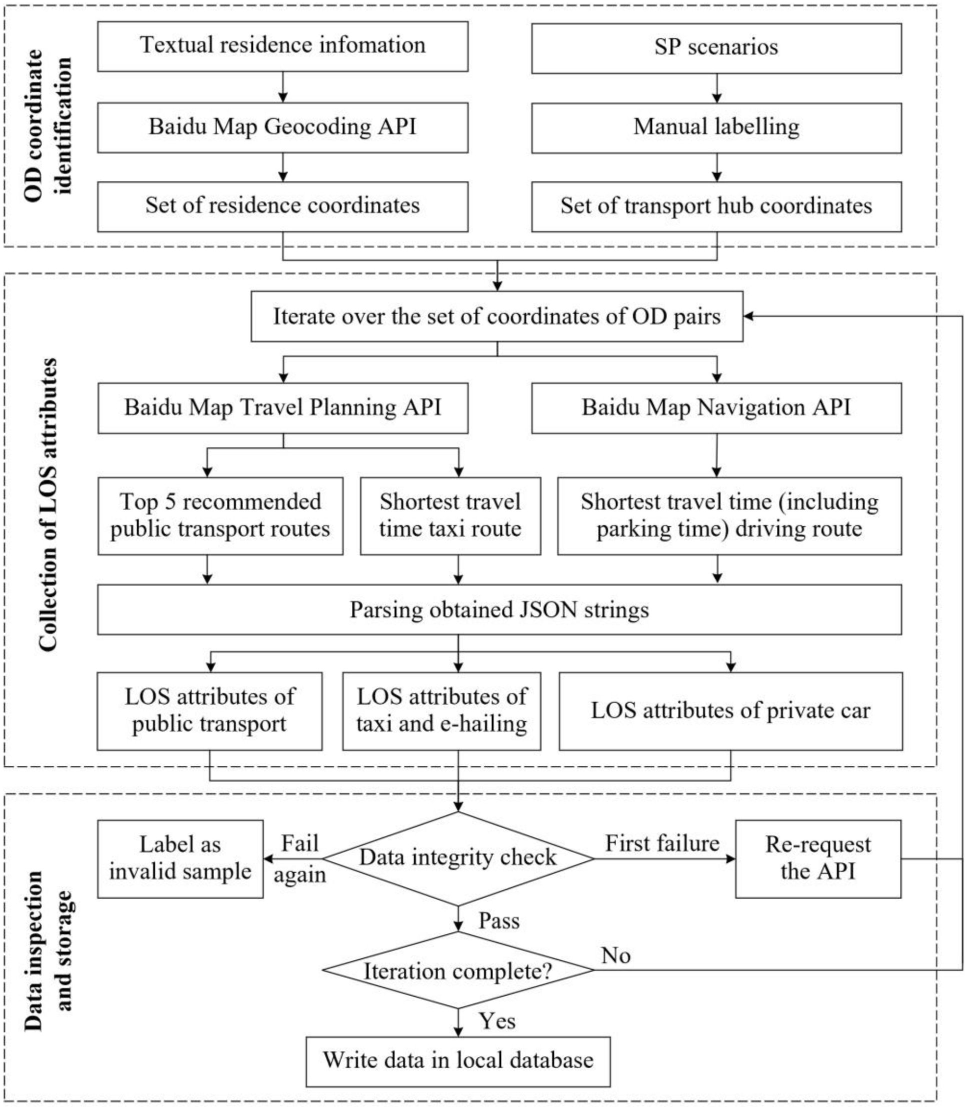


Figure 6 Procedures for LOS attributes collection

Table 4 Statistics of LOS attributes for access trips to major transport hubs

Attributes	Access modes	Transport hubs				
		Beijing railway station	Beijing South railway station	Beijing West railway station	Liuliqiao coach station	Bawangfen coach station
Mean travel time (s)	Metro	3,489.78	3,416.80	3,476.44	3,817.95	4,154.00
	Bus	5,770.99	6,268.19	5,199.52	5,418.48	6,049.51
	IPT	4,734.07	5,034.84	4,721.85	4,990.24	4,979.92
	ToE	2,402.98	2,486.36	2,670.93	2,160.20	2,306.19
Mean travel expense (CNY)	Metro	5.36	5.52	5.51	5.61	5.50
	Bus	5.71	6.21	6.01	6.34	6.20
	IPT	7.58	8.04	7.84	7.79	7.90
	ToE	71.32	77.74	76.93	77.05	78.28
Mean number of transfers	Metro	1.01	0.89	1.07	0.90	0.84
	Bus	1.09	0.98	1.02	1.03	1.25
	IPT	1.61	1.63	1.83	1.73	1.89

Note. For private car travellers, the mean travel distance to motorway entrances is 38.66km, and the mean travel time is 2,629.29s.

5. Model specifications

This section outlines the model specifications for assessing simultaneous and sequential estimation methods using the collected SP data, along with the computation methods for model performance indicators.

5.1 Model comparison framework

As suggested by Raveau et al. (2010), simultaneous and sequential estimation can lead to differences in forecasting and policy evaluation. Hence, this study introduces a comparative framework to assess the applicability of simultaneous and sequential estimation in the context of modelling intermodal behaviour. Specifically, two sets of model specifications are customised to conduct simultaneous and sequential estimation, as illustrated in Figure 7.

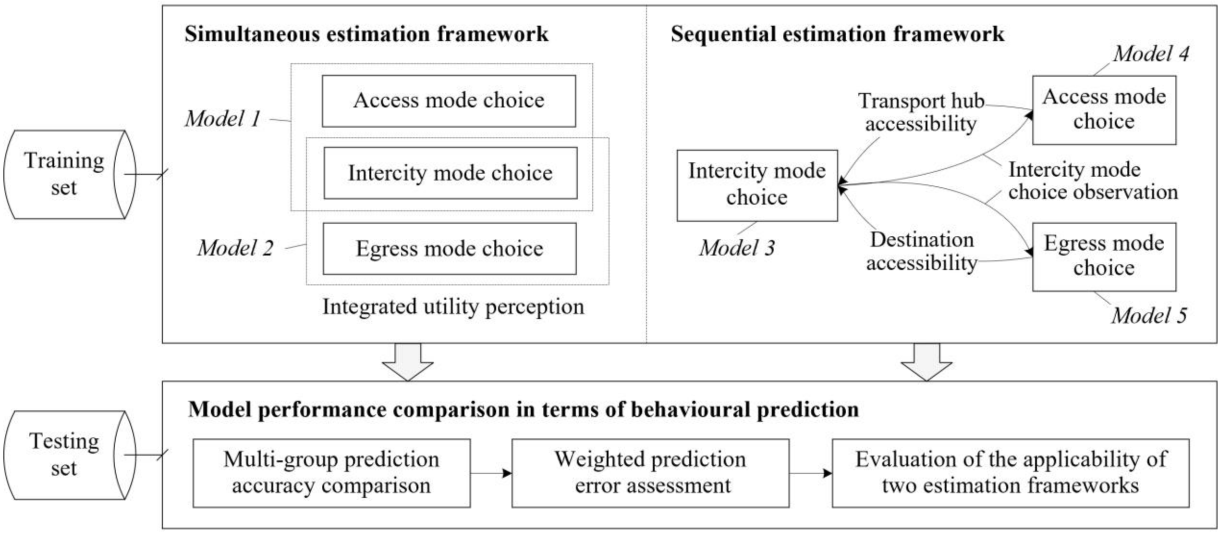


Figure 7 Illustration of model comparison framework

To ensure a fair comparison of model performance, the collected SP samples are randomly partitioned into training and testing sets in a 1:3 ratio. The training set is utilised for calibrating the models, while the testing set is employed to assess the models' behaviour prediction errors. Inspired by discussions on simultaneous and sequential choice problems (Donkers et al. 2020; Freidin et al. 2009), this study incorporated different decision-making sequences into the two estimation frameworks using the following hypotheses, with the primary distinction between the two sets of models outlined as follows.

In the simultaneous estimation framework, behavioural models are allowed to predict intercity mode choice in the testing dataset given known feeder mode choice outcomes, and vice versa. Namely, simulating the decision process when the multi-stage choices have been made simultaneously.

Under the sequential estimation framework, behavioural models predict feeder mode choice using the testing samples given the known intercity mode choice outcome. The difference lies in predicting intercity mode choice when feeder mode choice outcomes are unknown. Logically, intercity mode is considered a predecision before determining feeder modes, aligning with the essence of sequential decision-making.

5.2 Choice models under simultaneous estimation framework

The simultaneous estimation framework aims to simulate the joint decision-making of intermodal travellers across three stages of choices using a cross-nested model structure. As illustrated in Figure 7, the cross-nested logit (CNL) model operates under the assumption that intermodal travellers perceive the total utility of the intercity mode and its feeder modes. In comparison to the regular nested structure, CNL models enable a more flexible integration across multiple choices by allowing the alternatives allocated to more than one nest, as shown in Figure 8.

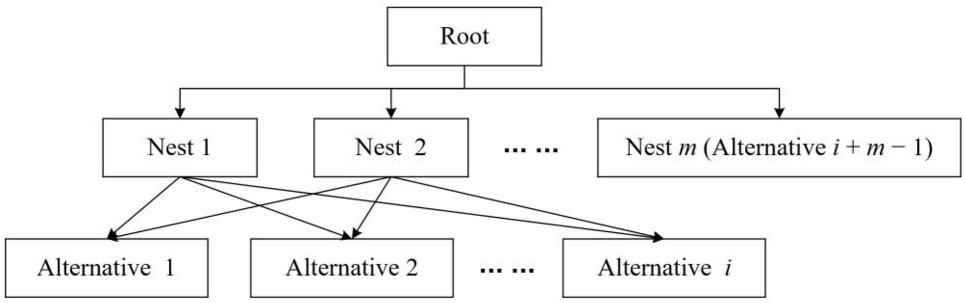


Figure 8 CNL model structure

Let $\chi_{r,m}$ denote the allocation of alternative *r* to nest *m*, which satisfies the following constraints.

$$0 \le \chi_{r,m} \le 1, \quad \forall r, m \tag{1}$$

$$\sum_{r=1}^{A_m} \chi_{r,m} = 1 \tag{2}$$

where A_m is the set of alternatives in nest *m*.

For traveller *i*, the probability of choosing alternative *r* is given by Eq. (3).

$$P_{i,r} = \sum_{m=1}^{N_r} P_{i,r|m} P_{i,m} \tag{3}$$

where N_r is the set of nests to which alternative *r* belongs. $P_{i,r|m}$ is the conditional probability of traveller *i* choosing alternative *r* in nest *m*, and $P_{i,m}$ is the marginal probability of traveller *i* choosing nest *m*.

The formulae for calculating $P_{i,r|m}$ and $P_{i,m}$ are given as

$$P_{i,r|m} = \frac{\chi_{r,m}{}^{\mu_m} e^{\mu_m V_{i,r|m}}}{\sum_{j=1}^{A_m} \chi_{j,m}{}^{\mu_m} e^{\mu_m V_{i,j|m}}} \tag{4}$$

$$P_{i,m} = \frac{\left(\sum_{j=1}^{A_m} \chi_{j,m}{}^{\mu_m} e^{\mu_m V_{i,j|m}}\right)^{1/\mu_m}}{\sum_{e=1}^{N}\left(\sum_{j=1}^{A_e} \chi_{j,e}{}^{\mu_e} e^{\mu_e V_{i,j|e}}\right)^{1/\mu_e}} \tag{5}$$

where N is the set of nests. μ_m is the scale parameters for the lower level of the CNL model, given that the normalisation of the model is performed at the top, $\mu_m > 1$ always holds.

To ensure realistic model calibration, particularly in reducing computational complexity, the three-stage mode decisions are divided into two choice modelling problems. Given that intercity travel constitutes the most central part of the entire journey, Model 1 is calibrated to account for the combination of intercity and access mode choice, while Model 2 is utilised to interpret the combination of intercity and egress mode choice. Figure 9 depicts the specific structure of the CNL models, which comprise seventeen alternatives (A1 to A17) contained in nine independent nests, including five nests of intercity modes (N1 to N5) and four nests of feeder modes (N6 to N9). All alternatives, except A17, are simultaneously included in an intercity mode nest and a feeder mode nest in the same proportion. For instance, A1 refers to intermodal travel consisting of HSRa as the intercity mode and the metro as the access mode in Model 1 or as the egress mode in Model 2.

5.3 Choice models under sequential estimation framework

The sequential estimation framework employs a multi-stage approach to estimate multiple choices sequentially, emphasising the implicit costs relevant to each decision. Following the depiction in Figure 7, the core component of intercity mode choice is considered as the predecision and estimated in the first order, i.e., Model 3, wherein the implicit costs of feeder trips are measured in the form of accessibility. Subsequently, the secondary components of feeder mode choices are estimated using Models 4 and 5, capturing the behavioural preferences influenced by intercity mode choice using dummy variables. In line with the alternative set depicted in Figure 9, five alternative intercity modes are included in Model 3 based on an MNL structure, as shown in Figure 10.

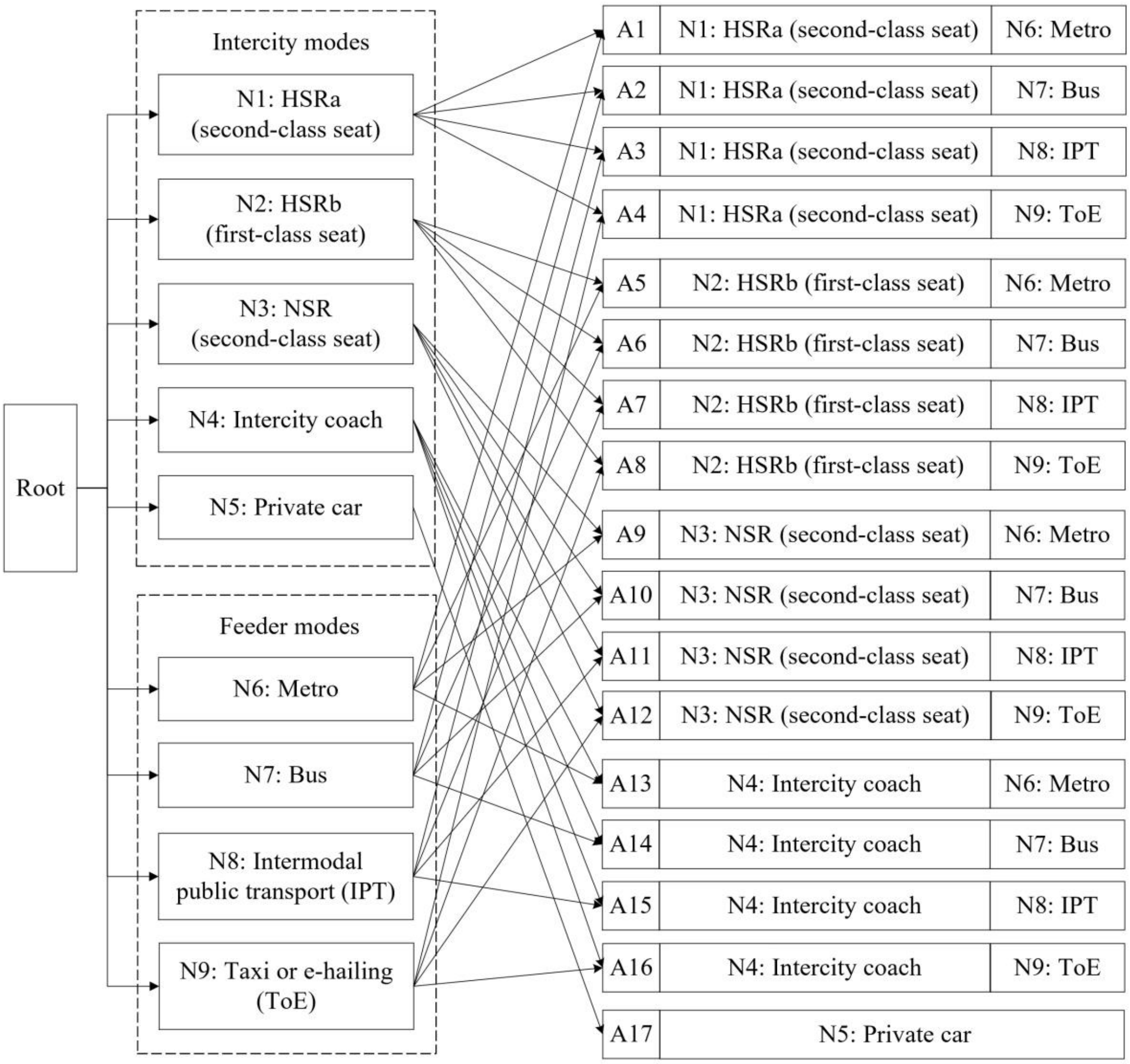


Figure 9 CNL model structure for intermodal travel behaviour

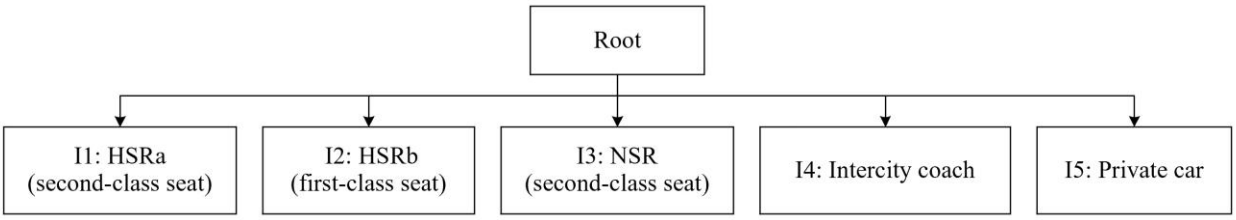


Figure 10 Model structure for intercity mode choice

For traveller i, the deterministic term of perceived utility of choosing alternative r can be expressed as follows.

$$V_{\mathrm{IM}}^{i,r} = \boldsymbol{\beta}_{\boldsymbol{\Phi}_{\mathrm{IM}}} \boldsymbol{\Phi}_{\mathrm{IM}}^{i,r} + \boldsymbol{\beta}_{\boldsymbol{\Gamma}_{\mathrm{IM}}} \boldsymbol{\Gamma}_{\mathrm{IM}}^{i,r} + \boldsymbol{\beta}_{\mathbf{T}_{\mathrm{IM}}} \mathbf{T}_{\mathrm{IM}}^{i} + \boldsymbol{\beta}_{\mathbf{S}_{\mathrm{IM}}} \mathbf{S}_{\mathrm{IM}}^{i} \tag{6}$$

$$\boldsymbol{\Gamma}_{\mathrm{IM}}^{i,r} = \left\{ X\left(HA_{i,r}\right),\ X\left(DA_{i,r}\right) \right\} \tag{7}$$

where $V_{\mathrm{IM}}^{i,r}$ is the utility of choosing alternative (intercity mode) r perceived by traveller i in

Model 3. $\mathbf{\Phi}_{\mathrm{IM}}^{i,r}$ and $\mathbf{\Gamma}_{\mathrm{IM}}^{i,r}$ are the vectors of explanatory variables regarding the explicit and implicit costs of alternative *r* perceived by traveller *i*. $\mathbf{\Phi}_{\mathrm{IM}}^{i,r}$ measures the LOS variables of intercity mode, and $\mathbf{\Gamma}_{\mathrm{IM}}^{i,r}$ reflects travellers' overall perception of the convenience of shuttle services in feeder trips, represented by the accessibility to the transport hubs. Note that $\mathbf{\Gamma}_{\mathrm{IM}}^{i,r}$ does not apply to the alternative of private cars. $\mathbf{T}_{\mathrm{IM}}^{i}$ and $\mathbf{S}_{\mathrm{IM}}^{i}$ are the vectors of explanatory variables regarding travel characteristics (e.g., travel purpose and fellow traveller) and socio-demographics of traveller *i*. $\boldsymbol{\beta}_{\mathbf{\Phi}_{\mathrm{IM}}}$, $\boldsymbol{\beta}_{\mathbf{\Gamma}_{\mathrm{IM}}}$, $\boldsymbol{\beta}_{\mathbf{T}_{\mathrm{IM}}}$, and $\boldsymbol{\beta}_{\mathbf{S}_{\mathrm{IM}}}$ are the coefficient vectors for $\mathbf{\Phi}_{\mathrm{IM}}^{i,r}$, $\mathbf{\Gamma}_{\mathrm{IM}}^{i,r}$, $\mathbf{T}_{\mathrm{IM}}^{i}$, and $\mathbf{S}_{\mathrm{IM}}^{i}$ to be estimated. $HA_{i,r}$ and $DA_{i,r}$ are the accessibility to the transport hub and destination when choosing alternative *r* perceived by traveller *i*. $X\left(HA_{i,r}\right)$ and $X\left(DA_{i,r}\right)$ are the functions of $HA_{i,r}$ and $DA_{i,r}$, respectively.

As the primary explanatory factors for the implicit cost arising from the remaining stages of travel, $HA_{i,r}$ and $DA_{i,r}$ aim to reflect the overall service level of shuttle transport in the access and egress trips. This can be quantified using logsum terms, as follows.

$$HA_{i,h} = \ln\left\{\sum_{a\in A_{\mathrm{AC}}} \exp\left(\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{AC}}} \mathbf{\Phi}_{\mathrm{AC}}^{i,a,h}\right)\right\},\ r \to h \tag{8}$$

$$DA_{i,d} = \ln\left\{\sum_{e\in A_{\mathrm{EG}}} \exp\left(\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{EG}}} \mathbf{\Phi}_{\mathrm{EG}}^{i,e,d}\right)\right\},\ r \to d \tag{9}$$

where $HA_{i,h}$ is the transport hub accessibility for the access trip from the residence of traveller *i* to the departure transport hub *h* (in line with using alternative intercity mode *r*). $DA_{i,d}$ is the destination accessibility for the egress trip from the arrival transport hub to the destination *d* (in line with using alternative intercity mode *r*). A_{AC} and A_{EG} are the sets of alternatives of access and egress modes. $\mathbf{\Phi}_{\mathrm{AC}}^{i,a,h}$ and $\mathbf{\Phi}_{\mathrm{EG}}^{i,e,d}$ are the vectors of explanatory variables reflecting explicit costs of access trip to transport hub *h* by access mode *a*, and egress trip to destination *d* by egress mode *e*. $\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{AC}}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{EG}}}$ are the coefficient vectors for $\mathbf{\Phi}_{\mathrm{AC}}^{i,a,h}$ and $\mathbf{\Phi}_{\mathrm{EG}}^{i,e,d}$. It should be noted that $\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{AC}}}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{\Phi}_{\mathrm{EG}}}$ are pre-calibrated using training samples, serving as prior knowledge and are not estimated in Model 3.

The formula for calculating the probability of traveller *i* choosing alternative *r* is given as:

$$P_{\mathrm{IM}}^{i,r} = \frac{\exp\left(V_{\mathrm{IM}}^{i,r}\right)}{\sum_{l\in A_{\mathrm{IM}}} \exp\left(V_{\mathrm{IM}}^{i,l}\right)} \tag{10}$$

where $P_{\mathrm{IM}}^{i,r}$ is the probability of traveller *i* choosing alternative *r*. A_{IM} is the set of alternative intercity modes.

Regarding the feeder mode decisions in Models 4 and 5, the alternative set consists of metro, bus, IPT, and ToE, denoted by alternatives 1 to 4 (F1 to F4) in Figure 11. F3 is unavailable in the egress trip, as illustrated in the SP scenario design.

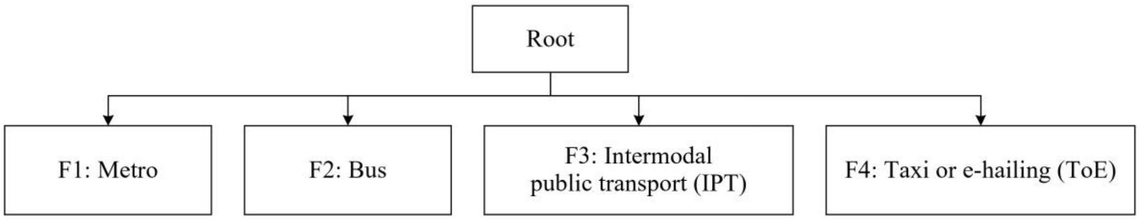


Figure 11 Model structure for feeder mode choice

Using the same composition as in Eq. (6), the perceived utility of feeder modes is as follows.

$$V_{\mathrm{FM}}^{i,f} = \boldsymbol{\beta}_{\boldsymbol{\Phi}_{\mathrm{FM}}} \boldsymbol{\Phi}_{\mathrm{FM}}^{i,f} + \boldsymbol{\beta}_{\boldsymbol{\Gamma}_{\mathrm{FM}}} \boldsymbol{\Gamma}_{\mathrm{FM}}^{i,f} + \boldsymbol{\beta}_{\mathbf{T}_{\mathrm{FM}}} \mathbf{T}_{\mathrm{FM}}^{i} + \boldsymbol{\beta}_{\mathbf{S}_{\mathrm{FM}}} \mathbf{S}_{\mathrm{FM}}^{i} \tag{11}$$

$$\boldsymbol{\Gamma}_{\mathrm{FM}}^{i,f} = \left\{ \mathrm{X}_{\mathrm{HSRa}}^{i,f},\ \mathrm{X}_{\mathrm{HSRb}}^{i,f},\ \mathrm{X}_{\mathrm{NSR}}^{i,f},\ \mathrm{X}_{\mathrm{COA}}^{i,f} \right\} \tag{12}$$

where $V_{\mathrm{FM}}^{i,f}$ is the utility of choosing alternative (access or egress mode) *f* perceived by traveller *i* in the subsequent decision Models 4 and 5. $\boldsymbol{\Phi}_{\mathrm{FM}}^{i,f}$ and $\boldsymbol{\Gamma}_{\mathrm{FM}}^{i,f}$ are the vectors of explanatory variables about the explicit and implicit costs of alternative *f* perceived by traveller *i*. $\boldsymbol{\Phi}_{\mathrm{FM}}^{i,f}$ measures the LOS variables of feeder mode *f*, which is written as $\boldsymbol{\Phi}_{\mathrm{AC}}^{i,a,h}$ in Model 4 and $\boldsymbol{\Phi}_{\mathrm{EG}}^{i,e,d}$ in Model 5. $\boldsymbol{\Gamma}_{\mathrm{FM}}^{i,f}$ describes the sequential effects of intercity mode decisions on feeder mode choice preferences. $\mathbf{T}_{\mathrm{FM}}^{i}$ and $\mathbf{S}_{\mathrm{FM}}^{i}$ are the vectors of travel characteristics and socio-demographics of traveller *i*. $\boldsymbol{\beta}_{\boldsymbol{\Phi}_{\mathrm{FM}}}$, $\boldsymbol{\beta}_{\boldsymbol{\Gamma}_{\mathrm{FM}}}$, $\boldsymbol{\beta}_{\mathbf{T}_{\mathrm{FM}}}$, and $\boldsymbol{\beta}_{\mathbf{S}_{\mathrm{FM}}}$ are the coefficient vectors for $\boldsymbol{\Phi}_{\mathrm{FM}}^{i,f}$, $\boldsymbol{\Gamma}_{\mathrm{FM}}^{i,f}$, $\mathbf{T}_{\mathrm{FM}}^{i}$, and $\mathbf{S}_{\mathrm{FM}}^{i}$ to be estimated. $\mathrm{X}_{\mathrm{HSRa}}^{i,f}$, $\mathrm{X}_{\mathrm{HSRb}}^{i,f}$, $\mathrm{X}_{\mathrm{NSR}}^{i,f}$, and $\mathrm{X}_{\mathrm{COA}}^{i,f}$ are dummy variables that represent if traveller *i* chooses HSRa, HSRb, NSR or coach as his/her intercity mode, forming an integral part of $V_{\mathrm{FM}}^{i,f}$.

Similarly, the probability of traveller *i* choosing alternative *f*, considering the sequential effects of intercity mode decision, can be expressed as follows.

$$P_{\mathrm{FM}}^{i,f} = \frac{\exp\left(V_{\mathrm{FM}}^{i,f}\right)}{\sum_{h \in A_{\mathrm{FM}}} \exp\left(V_{\mathrm{FM}}^{i,h}\right)} \tag{13}$$

where $P_{\mathrm{FM}}^{i,f}$ is the probability of traveller *i* choosing alternative *f* in Models 4 and 5. A_{FM} is the set of alternatives of Models 4 and 5.

5.4 Model performance indicator

Two dimensions of choice prediction errors are introduced to evaluate the performance of two sets of models. The prediction error of the probability for the chosen travel mode is used as the first indicator, calculated as follows.

$$PE_{r,h} = \left|\eta_r - \bar{\eta}_{r,h}\right|,\ r \in A \tag{14}$$

$$\bar{\eta}_{r,h} = \sum_{s \in S_{\text{test}}} p_{r,h}(s) \Big/ Q_{\text{test}} \tag{15}$$

where $PE_{r,h}$ is the prediction error of probability for alternative r chosen from A based on hypothesis h (either under simultaneous or sequential estimation framework). A is the set of alternatives, denoting the set of intercity or feeder modes. η_r is the percentage of choosing alternative r in the testing set. $\bar{\eta}_{r,h}$ is the predicted probability of choosing alternative r based on hypothesis h in the testing set. $p_{r,h}(s)$ is the probability of a testing sample s choosing alternative r derived from the hypothesis h-based model, where s belongs to the testing set S_{test}. Q_{test} is the size of S_{test}.

The weighted prediction error is further introduced as another performance indicator to assess the overall accuracy of the models. The formula for calculating weighted prediction errors is given as:

$$WPE_h = \frac{\sum_{r \in A} PE_{r,h} q_r}{Q_{\text{test}}} \tag{16}$$

where WPE_h is the weighted prediction error of choice models under hypothesis h. q_r is the frequency of alternative r chosen in the testing set.

6. Results and implications

6.1 Model estimation results

Based on the SP data specified in Section 4 and model specifications introduced in Section 5, this section presents model estimation results and draws behavioural implications through comparative analysis.

All the models in this study are calibrated using Biogeme, an open-source Python package for discrete choice models (Bierlaire 2020). For model calibration, 10,029 SP observations are randomly selected to form the training set, while the remaining 3,522 samples constituted the testing set. The variance inflation factor (VIF) is calculated to measure the degree of multicollinearity among explanatory variables. For details on the calculation methods of VIF, refer to Shrestha (2020). The literature typically considers 5 or 10 as the threshold value and suggests that VIF > 5 is a cause for concern, while VIF > 10 indicates a serious collinearity problem (Menard 2001). All the variables used to account for the alternatives' utility in the final models are found to be not significantly correlated, with the VIF values less than 5.

Utilising the simultaneous estimation approach, Models 1 and 2 are estimated as presented in

Table 5.

Table 5 Simultaneous estimation results

Explanatory variables	Units	Specific to	Model 1		Model 2	
			Est.	T-rat.	Est.	T-rat.
LOS attributes						
In-vehicle travel time	Hour	A1~A17	−0.364	−8.26	−0.617	−19.75
Non-reimbursable travel expense	CNY	A1~A17	−0.00545	−16.71	−0.0075	−13.36
Egress connection time	Hour	N6, N7, N8, N9	N/A	N/A	−1.25	−10.37
Transfer required in feeder trip	0-1	N6, N7	−0.402	−13.19	−0.289	−9.39
Interaction of walking distance and travelling with vulnerable groups	Km*0-1	N6, N7, N8	−0.0118	−4.31	−0.227	−2.89
Travel characteristics						
Travelling for business	0-1	A17	−0.424	−4.08	−0.663	−6.46
Travelling with companions	0-1	A17	0.456	6.56	0.415	6.19
Travelling with vulnerable groups	0-1	A17	0.764	7.66	0.867	9.29
		N8	−0.415	−4.04	N/A	N/A
Travelling with checked baggage	0-1	A17	1.05	15.55	0.872	13.00
		N9	1.12	22.18	0.220	7.49
Socio-demographics						
Male	0-1	A17	0.612	10.28	0.534	9.26
Aged over 50 years	0-1	N3	−0.379	−3.87	−0.382	−4.04
		N6	N/A	N/A	−0.0526	−3.20
Below bachelor’s degree	0-1	N3	0.427	5.67	0.369	5.13
		N7	N/A	N/A	0.155	3.48
Monthly income < ¥ 6,000	0-1	N3	0.560	7.62	0.477	7.29
		N6	N/A	N/A	0.0669	4.40
Monthly income > ¥ 15,000	0-1	N2	N/A	N/A	0.321	4.45
		N9	0.399	7.77	0.102	5.02
Car owner	0-1	A17	2.05	25.72	1.92	26.42
Others						
Alternative-specific constants	N/A	N1	1.18	22.45	1.23	24.71
		N3	−0.546	−6.10	N/A	N/A
		N4	−2.19	−7.28	N/A	N/A
		N6	1.45	26.10	0.488	13.64
		N8	0.690	13.53	N/A	N/A
		N9	0.186	2.80	0.333	10.51
Scale parameters for nests	N/A	N1	3.41	11.87	16.20	6.91
		N2	1.74	6.71	7.46	6.33
		N3	2.01	6.02	1.46	10.52
		N4	1.03	3.73	5.26	6.51
		N6	3.99	7.06	14.52	8.39
		N7	1.12	12.23	1.49	18.29
		N8	1.58	5.85	N/A	N/A
		N9	1.63	7.60	2.10	17.61
Model summary						
Number of parameters			30		30	
Sample size			10,029		10,029	
Initial log-likelihood			−27,872.50		−25,723.88	
Final log-likelihood			−17,681.67		−17,757.06	
Adjusted Rho-square			0.365		0.309	

Note. For the notations of nests and alternatives, see Figure 9. ‘N/A’ indicates insignificant or unapplicable variables that are not included in the final model.

Both models exhibit acceptable goodness-of-fit, as indicated by the values of the adjusted Rho-square. Regarding the scale parameters of nested structures, as illustrated in Eq. (5), $\mu_m > 1$ always holds. The increasing value of μ_m indicates an increased correlation across the alternatives in nest *m*. When μ_m collapses to a base value of 0, it indicates an absence of correlation, namely the nested model is equivalent to the MNL model. In Model 1, all the scale parameters adhere to the constraint, although N4 has a relatively low value of 1.03, indicating a relatively weak correlation within the intercity coach nest. In Model 2, eight estimated scale parameters validate that the alternatives within the nests correlate well.

Regarding the LOS variables, the signs of travel time and expense-related variables in both models are negative, aligning with expectations. As indicated in numerous previous studies (Capurso et al. 2019; Hess et al. 2018; Román et al. 2014; Zhou et al. 2020), the purpose of business travel significantly influences travellers' choice preferences. This influence is attributed not only to the urgency of business travel but also to the reimbursability of expenses. Therefore, in this study, respondents' self-reported travel purpose and the feasibility of reimbursement are utilised to calculate travel expenses they need to pay on their own, i.e., non-reimbursable travel expenses. Additionally, the transfer required in the feeder trip decreases the perceived utility. Concerning the walking distance in feeder trips, an interaction term is introduced to improve model fit. It indicates that travellers with vulnerable groups significantly perceive the negative impact of walking distance.

Travel characteristic-related dummy variables are primarily employed to interpret the preferences for private cars (A17) in the intercity stage, IPT (N8) and ToE (N9) in feeder trips. Specifically, travelling with companions, vulnerable groups, or carrying checked baggage are factors that facilitate the choice of driving in intercity travel, whereas travelling for business shows the opposite. Additionally, carrying large luggage is a significant predictor of ToE preference, while the presence of vulnerable groups decreases the likelihood of using IPT. In terms of socio-demographic factors, males and car owners prefer driving in intercity travel. NSR is less appealing to senior travellers but shows increased preferences among low-income and low-education groups. Relatively, high-income groups tend to choose HSRb. As for feeder modes, the metro is the preferred choice for low-income groups but is less preferred by senior travellers. The low-education dummy shows a positive influence on the alternative of the bus, while the high-income dummy adds utility to choosing ToE.

Furthermore, based on the sequential estimation framework, the intercity mode and feeder mode choices are estimated sequentially. The estimation results for the intercity mode choice model are presented in Table 6.

Table 6 Sequential estimation results (intercity mode choice)

Explanatory variables	Units	Specific to	Model 3	
			Est.	T-rat.
Explicit costs of intercity travel				
In-vehicle travel time	Hour	I1~I5	−0.664	−17.86
Non-reimbursable travel expense	CNY	I1~I5	−0.0102	−19.72
Implicit costs of feeder trips				
Transport hub accessibility	N/A	I1~I4	0.496	3.95
Maximum destination accessibility	0-1	I1~I4	0.184	4.26
Travel characteristics				
Unfamiliar with destination city	0-1	I5	−0.234	−2.87
Travelling for business	0-1	I5	−0.468	−5.47
Travelling with vulnerable groups	0-1	I3	−0.492	−3.11
	0-1	I5	0.550	7.18
Travelling with checked baggage	0-1	I2	0.372	4.28
	0-1	I5	0.647	10.76
Annual intercity travel frequency ≥ 9	0-1	I1	0.301	4.60
Socio-demographics				
Monthly income < ¥ 6,000	0-1	I3	0.802	10.35
Monthly income > ¥ 15,000	0-1	I4	−0.846	−2.42
Car owner	0-1	I5	1.37	22.88
Others				
Alternative-specific constants	N/A	I2	−1.64	−37.01
		I3	−1.95	−26.58
		I4	−3.06	−25.16
Model summary				
Number of parameters			17	
Sample size			10,029	
Initial log-likelihood			−16,141.05	
Final log-likelihood			−9,686.251	
Adjusted Rho-square			0.399	

Note. In-vehicle travel time for private car (I5) is composed of two parts: the duration from the traveller's residence to the motorway entrance (retrieved from the Baidu Map WEB-API) and the subsequent duration from there to the final destination (as specified in the SP scenario). Refer to Figure 10 for the notations of alternatives.

In Model 3, the signs for travel time and non-reimbursable travel expenses are consistent with those in Models 1 and 2. In addition to the explicit costs of intercity modes, further explanation is offered by the implicit costs arising from the feeder trips. Two accessibility-related variables that quantify the overall impression of the convenience of feeder trips are used to account for implicit costs in intercity mode decisions. By testing various specifications of $\boldsymbol{\Gamma}_{\mathrm{IM}}^{i,r}$ in Eq. (7), the final Model 3 utilises a numerical form of transport hub accessibility, and a dummy variable named maximum destination accessibility is defined to capture travellers' preferences for the most convenient onward transport from transport hubs to the destination. For example, given a traveller's destination, if the onward transport at hub x is of the highest level of service compared to that at other hubs, the maximum destination accessibility would be set to 1 for the alternative of intercity modes arriving at hub x. It would be set to 0 for the alternatives arriving at other hubs. The estimates for the two variables are 0.496 (t-value 3.95) and 0.184 (t-value 4.26), respectively, demonstrating a significant impact of relevant shuttle services on the present decision. Namely, when travellers decide on intercity modes, they will roughly consider the difficulty of accessing the transport hub and reaching the destination in a preliminary evaluation. Consequently, intercity modes offering better access and egress transport services turn out to be more attractive to intermodal travellers.

Regarding travel characteristics that show significant effects on perceived utility, travellers who are unfamiliar with the destination city (0 represents frequent or occasional visit; 1 represents rarely or never visited) or those travelling for business purposes do not favour using private cars. Conversely, travelling with vulnerable groups and checked baggage, as well as owning a private car, increases the probability of driving. With respect to the preference for rail transport, HSRa is the most preferred intercity mode for frequent travellers. Travelling with checked baggage triggers the need for more seating space and thus increases the utility of HSRb. As a less costly but time-consuming mode relative to HSR, NSR is more attractive to travellers on low incomes but is not preferred by vulnerable groups. Additionally, high-income travellers are found to be less inclined to use an intercity coach. The negative estimates of the three alternative-specific constants imply that, apart from the above factors, travellers have a potential preference for HSRa.

The estimation results for the feeder mode choice models are presented in Table 7.

Table 7 Sequential estimation results (feeder mode choice)

Explanatory variables	Units	Specific to	Model 4		Model 5	
			Est.	T-rat.	Est.	T-rat.
Explicit costs of feeder trips						
In-vehicle travel time	Hour	F1~F4	−0.297	−5.89	N/A	N/A
In-vehicle travel time > 30min	0-1	F1~F4	N/A	N/A	−0.671	−6.60
Egress connection time	Hour	F1~F4	N/A	N/A	−1.96	−2.74
Non-reimbursable travel expense	CNY	F1~F4	−0.00421	−6.28	−0.0407	−19.06
No transfer required	0-1	F1, F2	0.293	5.73	0.142	1.83
More than one transfer	0-1	F1, F2	−0.669	−10.74	N/A	N/A
Walking distance > 500m	0-1	F1, F2	N/A	N/A	−0.0695	−2.15
Walking distance > 1km	0-1	F1~F3	−0.494	−4.57	N/A	N/A
Access trip distance > 15km	0-1	F2	−0.474	−4.90	N/A	N/A
		F3	0.452	6.00	N/A	N/A
Implicit costs of intercity travel						
HSRa as intercity mode	0-1	F1	1.08	4.23	0.195	2.45
		F2	−0.473	−4.28	−0.582	−6.12
		F4	−0.328	−2.94	N/A	N/A
HSRb as intercity mode	0-1	F1	0.791	2.93	N/A	N/A
		F2	N/A	N/A	−0.616	−6.40
		F4	0.329	2.51	0.556	5.71
NSR as intercity mode	0-1	F1	0.457	1.72	0.186	3.29

		F2	N/A	N/A	0.184	2.93
		F4	−1.27	−8.32	−0.371	−4.84
Coach as intercity mode	0-1	F1	N/A	N/A	−0.332	−2.47
		F2	N/A	N/A	0.610	3.27
		F4	−0.521	−2.15	N/A	N/A
Travel characteristics						
Travelling for business	0-1	F2	−0.520	−4.39	N/A	N/A
		F4	0.938	9.85	N/A	N/A
Travelling with companions	0-1	F4	N/A	N/A	0.271	5.24
Travelling with vulnerable groups	0-1	F4	1.31	12.47	1.09	10.88
Travelling with checked baggage	0-1	F2	N/A	N/A	−0.179	−1.94
		F4	1.85	24.64	1.01	15.10
Annual intercity travel frequency ≥ 9	0-1	F1	0.223	3.02	N/A	N/A
Socio-demographics						
Master's degree or above	0-1	F4	0.337	4.56	N/A	N/A
Monthly income < ¥ 6,000	0-1	F2	N/A	N/A	0.276	4.03
		F3	0.169	2.45	N/A	N/A
Monthly income > ¥ 15,000	0-1	F2	N/A	N/A	−0.359	−3.39
		F4	0.399	4.92	0.403	5.99
Others						
Alternative-specific constants	N/A	F2	−0.230	−1.85	−0.192	−3.26
		F3	−0.320	−2.21	N/A	N/A
		F4	−1.43	−6.41	−0.377	−3.86
Model summary						
Number of parameters			26		23	
Sample size			8,182		8,182	
Initial log-likelihood			−10,836.56		−8,988.846	
Final log-likelihood			−7,740.075		−7,772.305	
Adjusted Rho-square			0.283		0.133	

Note. For the notations of alternatives, see Figure 11.

Given that samples choosing private cars as the main mode do not involve feeder mode choices, these samples are excluded from the dataset when calibrating Models 4 and 5. The results indicate significant predictors for explicit costs, including in-vehicle time, connection time between intercity mode and egress mode, travel expense, number of transfers, walking distance, and access distance. Notably, various specifications are tested for in-vehicle travel time and walking distance in the modelling process. As the continuous form of in-vehicle travel time proves insignificant in Model 5, it is then converted into a dummy variable, indicating a difference in travel time perception during egress trips compared to intercity and access trips. Furthermore, there is a distinction in the thresholds used to define dummy variables for walking distance in access and egress trips. It is found that intermodal travellers appear to be more sensitive to longer travel times and exhibit preferences for shorter walking distances in the egress stage compared to the access stage. The total access distance shows significant influences on travellers' preferences for F2 and F3. Specifically, the bus is not preferred for long-distance access (> 15km), while IPT is more desirable in this context.

In terms of travel characteristics and socio-demographics, both Models 4 and 5 suggest that travellers in vulnerable groups, those with oversized baggage, or high incomes are inclined toward ToE, which provides enhanced comfort and privacy. Differentiated behavioural preferences are also evident in the access and egress trips. Taking the alternative of ToE as an example, for access mode choice, business travellers and individuals with higher education exhibit additional preferences. Meanwhile, in the egress trip, travellers with companions are more inclined to use ToE. Additionally, for access mode choice, the results indicate that frequent travellers prefer the metro, low-income travellers opt for IPT, and business travellers dislike buses. While in the egress stage, travellers primarily show significant preferences for buses, including the positive effect of the low-income dummy variable and negative effects of carrying checked baggage and high-income dummy variables. The alternative-specific constants estimated in both models are negative, suggesting the presence of unobserved factors that generally favour travellers using the metro as a feeder mode.

Regarding the implicit costs associated with the intercity mode decision, the sequential effect is captured using four dummy variables. These variables represent four types of predecision regarding the intercity mode, as defined in Eq. (12). Each dummy variable is included in the utility function of each alternative and has independent parameters to be estimated, intending to thoroughly examine the existence of the sequential effect. Only variables with significant estimates are retained in the final models. Taking the estimates in Model 4 as examples, rail users, including HSRa, HSRb, and NSR, express a stronger inclination to use the metro as an access mode. HSRb users also exhibit a preference for ToE, while the other two categories of travellers, similar to intercity coach users, demonstrate opposing tendencies towards ToE. Furthermore, HSRa users tend to disfavour bus shuttles, presumably due to inconvenient rail-to-ground connections and unsatisfactory bus service punctuality.

As seen from the signs of all the dummy variables in Figure 12, feeder mode preferences related to intercity mode decisions remain largely consistent between the access and egress stages, except for IPT, which is unavailable in the egress trip. A notable difference is that HSRa users show a preference for ToE in the egress stage but a reluctance in the access stage. The discrepancy is speculated to be attributed to increased fatigue as travellers reach the end of their journey, prompting a natural inclination for a more comfortable egress mode. This also underscores the necessity of distinguishing travel behaviour in these two stages.

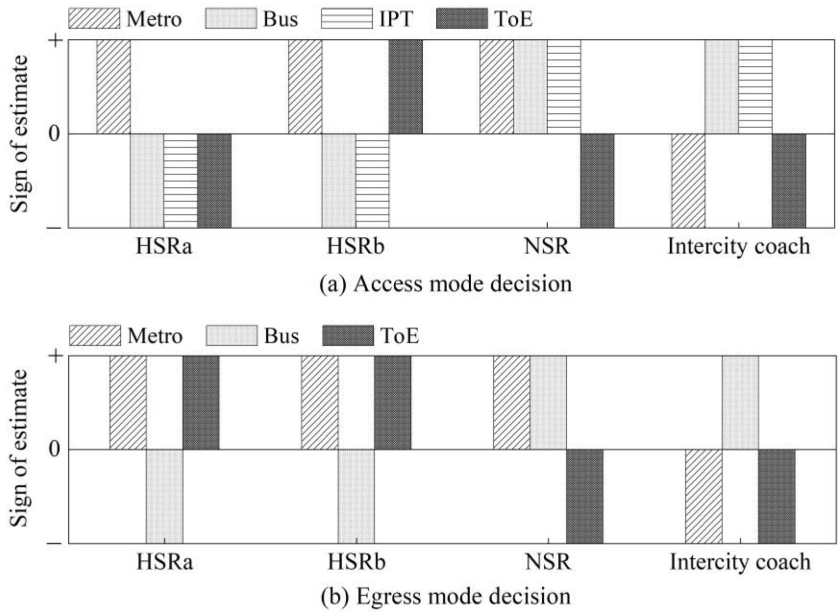


Figure 12 Impact of intercity mode decision on feeder mode preferences

6.2 Comparative analysis of behaviour prediction performance

Using the models obtained and the performance indicators defined in Eqs. (14)-(16), the behavioural prediction performance of two sets of models is compared on the testing sample set. Among all 3,522 sets of intermodal choice observations, 699 sets involve an intercity mode choice of private cars that does not involve feeder mode choices. Thus, only the remaining 2,823 samples are used for feeder mode choice prediction. Namely, Q_{test} in Eqs. (15) and (16) equals 3,522 or 2,823 depending on the tested models.

To comprehensively compare the models' predictive accuracy, travel characteristics are used as criteria for multigroup analysis. A total of nine groups are considered, classified by travel purpose, fellow traveller, baggage size, and intercity travel frequency. Note that each sample in the testing set can fall into more than one group. The statistical results for prediction errors under simultaneous and sequential estimation frameworks are reported in Tables 8, 9, and 10, for access mode choice, intercity mode choice, and egress mode choice, respectively.

It can be observed that the prediction accuracy largely depends on the alternatives. For instance, IPT has the highest error (2.24% and 1.88% for the total samples under the two estimation frameworks, see Table 8) in predicting feeder mode choice. Regarding intercity mode choice, sequential estimation shows the highest prediction error in the alternative of HSRa (0.71% for the total samples, see Table 9). In comparison, simultaneous estimation performs worst in predicting the probability of using private cars (1.68% for the total samples, see Table 9). Holistically, the prediction errors of models using simultaneous estimation are greater than those by sequential estimation at all three stages of travel.

Table 8 Comparison of prediction errors (access mode choice)

Groups		Sample sizes	Simultaneous estimation framework				Sequential estimation framework			
			Metro	Bus	IPT	ToE	Metro	Bus	IPT	ToE
Travel purpose	Leisure	1,877	1.244%	0.981%	1.427%	0.798%	0.729%	0.757%	1.800%	0.314%
	Business	946	0.780%	2.930%	3.999%	7.709%	0.236%	0.525%	2.045%	2.806%
Fellow traveller	Alone	1,594	0.381%	0.175%	2.436%	2.230%	1.330%	0.751%	1.851%	0.230%
	Accompanied	1,229	0.788%	0.492%	1.994%	1.697%	0.794%	0.222%	1.922%	2.937%
	With vulnerable groups	199	10.858%	5.077%	3.056%	18.991%	2.724%	2.238%	3.649%	3.162%
Baggage size	Carry-on baggage	2,164	1.765%	0.280%	2.207%	0.722%	0.937%	0.257%	2.077%	0.883%
	Checked baggage	659	0.798%	0.187%	1.406%	2.392%	1.339%	0.560%	1.242%	2.021%
Travel frequency	Non-frequent traveller	1,300	1.441%	0.264%	1.404%	0.301%	2.271%	0.794%	1.748%	1.316%
	Frequent traveller	722	1.496%	0.819%	3.238%	2.561%	0.975%	0.159%	1.572%	2.388%
Total testing samples		2,823	0.536%	0.318%	2.243%	2.025%	0.405%	0.328%	1.882%	1.149%

Note. The tested models estimated by simultaneous and sequential methods are Models 1 and 4, respectively.

Table 9 Comparison of prediction errors (intercity mode choice)

Groups		Sample sizes	Simultaneous estimation framework					Sequential estimation framework				
			HSRa	HSRb	NSR	Coach	Car	HSRa	HSRb	NSR	Coach	Car
Travel purpose	Leisure	2,476	2.095%	1.570%	0.459%	0.684%	1.636%	0.956%	1.473%	0.435%	0.011%	0.093%
	Business	1,046	0.307%	1.234%	0.895%	0.731%	1.783%	0.141%	1.720%	0.422%	0.381%	1.903%
Fellow traveller	Alone	1,969	2.663%	0.908%	0.743%	0.708%	2.802%	1.224%	0.720%	0.177%	0.469%	0.212%
	Accompanied	1,553	0.246%	0.880%	0.393%	0.987%	0.254%	0.068%	0.277%	0.753%	0.321%	0.865%
	With vulnerable groups	357	6.925%	4.467%	1.821%	1.175%	3.104%	8.853%	3.604%	0.244%	1.624%	3.381%
Baggage size	Carry-on baggage	2,544	1.369%	0.940%	0.148%	0.782%	2.078%	0.777%	0.298%	0.041%	0.025%	0.464%
	Checked baggage	978	1.407%	2.954%	2.073%	0.578%	0.634%	0.549%	1.114%	1.659%	0.500%	0.593%
Travel frequency	Non-frequent traveller	1,634	2.472%	0.920%	0.434%	0.705%	2.756%	1.210%	0.464%	0.064%	0.040%	0.723%
	Frequent traveller	892	0.534%	1.646%	0.141%	1.197%	0.121%	0.602%	1.222%	1.819%	0.549%	0.650%
Total testing samples		3,522	1.381%	0.885%	0.588%	0.698%	1.679%	0.714%	0.525%	0.431%	0.121%	0.500%

Note. The tested models estimated by the simultaneous method are Models 1 and 2, and the reported prediction errors are calculated based on the mean values of these two models. The tested model estimated by the sequential method is Model 3.

Table 10 Comparison of prediction errors (egress mode choice)

Groups		Sample sizes	Simultaneous estimation framework				Sequential estimation framework			
			Metro	Bus	IPT	ToE	Metro	Bus	IPT	ToE
Travel purpose	Leisure	1,877	0.551%	0.647%	N/A	1.198%	0.770%	1.568%	N/A	0.798%
	Business	946	2.048%	5.102%	N/A	7.150%	0.898%	0.643%	N/A	0.256%
Fellow traveller	Alone	1,594	0.134%	0.697%	N/A	0.563%	0.963%	0.889%	N/A	0.074%
	Accompanied	1,229	0.179%	3.767%	N/A	3.947%	1.733%	3.053%	N/A	1.320%
	With vulnerable groups	199	6.118%	10.197%	N/A	16.315%	1.792%	5.025%	N/A	6.816%
Baggage size	Carry-on baggage	2,164	1.009%	0.239%	N/A	1.247%	0.891%	0.946%	N/A	0.054%
	Checked baggage	659	1.776%	5.841%	N/A	7.616%	2.024%	0.438%	N/A	2.462%
Travel frequency	Non-frequent traveller	1,300	0.612%	0.027%	N/A	0.639%	1.261%	0.411%	N/A	0.851%
	Frequent traveller	722	3.749%	1.470%	N/A	5.219%	3.867%	0.085%	N/A	3.782%
Total testing samples		2,823	0.144%	1.246%	N/A	1.390%	0.211%	0.827%	N/A	0.616%

Note. The tested models estimated by simultaneous and sequential methods are Models 2 and 5, respectively. Given that IPT is unavailable in egress choice tasks (refer to Figure 4), the test follows the same assumption to account for the remaining three alternatives only.

As for the prediction performance across groups, a notable phenomenon is that prediction accuracy tends to be lower for groups with smaller sample sizes. For instance, the testing sample size of groups travelling with vulnerable companions is 199 out of 2,823 in forecasting feeder mode choice and 357 out of 3,522 in forecasting intercity mode choice. The highest errors for a single alternative in this group are up to 18.99% for ToE by simultaneous estimation in Table 8, 8.85% for HSRa by sequential estimation in Table 9, and 16.32% for ToE by simultaneous estimation in Table 10. Generally, the variations of errors across groups are indiscernible under the sequential estimation framework, demonstrating more robust performance in behaviour forecasting relative to simultaneous estimation.

The weighted prediction errors from the two estimation methods are further examined and reported in Table 11. In contrast to simultaneous estimation, the results reveal that sequential estimation exhibits lower weighted prediction errors across all three stages of choices. This underscores its superior suitability for modelling intermodal travel behaviour under the specific data conditions considered in this study. The findings emphasise the significance of investigating the decision-making process in multiple-choice scenarios, cautioning against the default assumption that simultaneous estimation is inherently superior to sequential estimation, particularly concerning demand forecasting outcomes.

Table 11 Statistical results for weighted prediction errors

Estimation methods	Access mode choice	Intercity mode choice	Egress mode choice
Simultaneous estimation framework	1.118%	1.330%	0.893%
Sequential estimation framework	0.806%	0.627%	0.497%

6.3 Implications

This study contributes research implications for relevant studies in two key aspects. Firstly, it identifies the behavioural determinants of intermodal travel across three travel stages within the context of mega-city regions. The findings suggest variability in the effects of explanatory variables across stages and validate the differences in preferences for access and egress mode choices. Secondly, the results of behaviour prediction highlight the importance of incorporating rational presumptions into choice modelling. The sequential estimation method confirms superior forecasting performance over the three stages of intermodal travel, questioning the default assumption of simultaneous estimation in existing models, and suggesting an outcome-oriented approach for relevant behavioural studies. Additionally, the proposed comparative model estimation framework shows transferability in addressing multiple decision problems, enabling a comprehensive exploration of the practical value of choice models.

Furthermore, this study offers practical implications for enhancing the accuracy of estimating intermodal travel demand for regional transport systems. There are additional application values for achieving on-demand and seamless scheduling between intercity and intracity transport. This plays an imperative role in advancing Mobility as a Service practice, particularly as its latest applications expand the focus from urban mobility to intercity mobility. The findings provide potential insights into tailoring incentive policies for intermodal mobility based on travellers' behavioural preferences obtained. Moreover, the proposed models lay the groundwork for predicting the dynamics of mobility patterns alongside the evolution of transport hubs, providing an assessment basis for future transport hub planning and integration.

7. Conclusions

This study aims to illuminate intermodal mobility in mega-city regions within the context of enhanced intercity accessibility. The research focuses on identifying the behavioural determinants of intermodal travellers at each stage of travel using stated preference survey data. Additionally, it seeks to validate the rationale behind simultaneous and sequential model estimation methods, with the criteria of achieving increased predictive accuracy of behavioural outcomes. The main conclusions drawn from this study are summarised as follows.

The choice models reveal a series of factors influencing individuals' decisions regarding intermodal travel, encompassing level-of-service attributes (e.g., in-vehicle travel time, non-reimbursable travel expense, and intermodal connection time), travel characteristics (e.g., travel purpose, fellow traveller, and intercity travel frequency), and socio-demographics. The results confirm differentiated choice preferences among travellers for access and egress travel modes. By employing different assumptions regarding the sequences of multiple decisions, the simultaneous estimation method validates the statistical soundness of the cross-nested structure. Meanwhile, the sequential estimation method indicates the existence of sequential effects across decisions, typically captured by the role of accessibility. From a more intuitive perspective on model prediction effectiveness, the weighted prediction errors for access, intercity, and egress mode choices are 1.12%, 1.33%, and 0.89% by simultaneous estimation, and 0.81%, 0.63%, and 0.50% by sequential estimation. Therefore, the latter is deemed statistically more suitable for interpreting and predicting intermodal travel behaviour than the former.

The findings underscore the importance of data-driven methods in behavioural studies, particularly for addressing multiple-choice problems, rather than relying on default assumptions. There are still limitations in data acquisition that require further research efforts. The implementation of questionnaire surveys and the collection of level-of-service attributes for customised travel scenarios unavoidably introduce a time lag. Additionally, the use of stated preference surveys limits the scope and quantity of choice observations. Exploring the utilisation of mobile phone signalling or trajectory data, coupled with advancements in behavioural modelling techniques, could represent a promising direction for future research to achieve a more comprehensive understanding of intermodal mobility within a broader spatial context.

Acknowledgement

This work was funded by the National Natural Science Foundation of China (52172312). Stephane Hess acknowledges the support of the European Research Council through the consolidator grant 615596-DECISIONS. The stated preference survey in this study was designed while Ning Huan stayed with Stephane Hess at the University of Leeds as a visiting PhD student.

Conflict of interest

On behalf of all authors, the corresponding author states that there are no conflicts of interest.

Reference

Bai, Y., Hu, Q., Ho, T.K., Guo, H., Mao, B.: Timetable optimization for metro lines connecting to intercity railway stations to minimize passenger waiting time. IEEE Transactions on Intelligent Transportation Systems. 22, 79–90 (2021). https://doi.org/10.1109/TITS.2019.2954895

Bergantino, A.S., Madio, L.: Intermodal competition and substitution. HSR versus air transport: Understanding the socio-economic determinants of modal choice. Research in Transportation Economics. 79, 100823 (2020). https://doi.org/10.1016/j.retrec.2020.100823

Bierlaire, M.: Estimating choice models with latent variables with PythonBiogeme. Report TRANSP-OR 160628. Series on Biogeme. Transport and Mobility Laboratory, ENAC, EPFL. (2016)

Bierlaire, M.: A short introduction to PandasBiogeme. Technical report TRANSP-OR 200605. Transport and Mobility Laboratory, ENAC, EPFL. (2020)

Bushell, J., Merkert, R., Beck, M.J.: Consumer preferences for operator collaboration in intra- and intercity transport ecosystems: Institutionalising platforms to facilitate MaaS 2.0. Transp Res Part A Policy Pract. 160, 160–178 (2022). https://doi.org/10.1016/j.tra.2022.04.013

Capurso, M., Hess, S., Dekker, T.: Modelling the role of consideration of alternatives in mode choice: An application on the Rome-Milan corridor. Transp Res Part A Policy Pract. 129, 170–184 (2019). https://doi.org/10.1016/j.tra.2019.07.011

Chaichannawatik, B., Kanitpong, K., Limanond, T.: Departure Time Choice (DTC) behavior for intercity travel during a long-holiday in Bangkok, Thailand. J Adv Transp. 1896261, (2019). https://doi.org/10.1155/2019/1896261

Cheng, Y.H., Tseng, W.C.: Exploring the effects of perceived values, free bus transfer, and penalties on intermodal metro-bus transfer users’ intention. Transp Policy (Oxf). 47, 127–138 (2016).

https://doi.org/10.1016/j.tranpol.2016.01.001

Diederich, A.: Sequential Decision Making. In: International Encyclopedia of the Social & Behavioral Sciences. pp. 13917–13922. Elsevier (2001)

Dobruszkes, F., Dehon, C., Givoni, M.: Does European high-speed rail affect the current level of air services? An EU-wide analysis. Transp Res Part A Policy Pract. 69, 461–475 (2014). https://doi.org/10.1016/j.tra.2014.09.004

Donkers, B., Dellaert, B.G.C., Waisman, R.M., Häubl, G.: Preference dynamics in sequential consumer choice with defaults. Journal of Marketing Research. 57, 1096–1112 (2020). https://doi.org/10.1177/0022243720956642

Fang, C., Yu, D.: Urban agglomeration: An evolving concept of an emerging phenomenon. Landsc Urban Plan. 162, 126–136 (2017). https://doi.org/10.1016/j.landurbplan.2017.02.014

Freidin, E., Aw, J., Kacelnik, A.: Sequential and simultaneous choices: Testing the diet selection and sequential choice models. Behavioural Processes. 80, 218–223 (2009). https://doi.org/10.1016/j.beproc.2008.12.001

Gokasar, I., Gunay, G.: Mode choice behavior modeling of ground access to airports: A case study in Istanbul, Turkey. J Air Transp Manag. 59, 1–7 (2017). https://doi.org/10.1016/j.jairtraman.2016.11.003

Gottmann, J.: Megalopolis: The urbanized Northeastern seaboard of the United States. MIT Press, Massachusetts (1961)

Hall, P.G., Pain, K.: The polycentric metropolis: Learning from mega-city regions in Europe. Routledge, London (2006)

Hensher, D.A., Rose, J.M.: Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure projects: A case study. Transp Res Part A Policy Pract. 41, 428–443 (2007). https://doi.org/10.1016/j.tra.2006.09.006

Hess, S., Spitz, G., Bradley, M., Coogan, M.: Analysis of mode choice for intercity travel: Application of a hybrid choice model to two distinct US corridors. Transp Res Part A Policy Pract. 116, 547–567 (2018). https://doi.org/10.1016/j.tra.2018.05.019

Huan, N., Yao, E., Xiao, Y.: Roles of accessibility and air-rail intermodality in shaping mobility patterns in mega-city regions: Behavioural insights from China. Cities. 143, 104591 (2023). https://doi.org/10.1016/j.cities.2023.104591

Huang, Y., Gan, H., Jing, P., Wang, X.: Analysis of park and ride mode choice behavior under multimodal travel information service. Transportation Letters. 14, 1080–1090 (2022). https://doi.org/10.1080/19427867.2021.1988438

Llorca, C., Molloy, J., Ji, J., Moeckel, R.: Estimation of a long-distance travel demand model using trip surveys, location-based big data, and trip planning services. Transportation Research Record: Journal of the Transportation Research Board. 2672, 103–113 (2018). https://doi.org/10.1177/0361198118777064

Lu, M., Zhu, H., Luo, X., Lei, L.: Intercity travel demand analysis model. Advances in Mechanical Engineering. 6, 108180 (2014). https://doi.org/10.1155/2014/10818

Luo, Q., Li, S., Hampshire, R.C.: Optimal design of intermodal mobility networks under uncertainty: Connecting micromobility with mobility-on-demand transit. EURO Journal on Transportation and Logistics. 10, 100045 (2021). https://doi.org/10.1016/j.ejtl.2021.100045

Menard, S.: Applied logistic regression analysis. SAGE Publications, Inc (2001)

Meyer de Freitas, L., Becker, H., Zimmermann, M., Axhausen, K.W.: Modelling intermodal travel in Switzerland: A recursive logit approach. Transp Res Part A Policy Pract. 119, 200–213 (2019). https://doi.org/10.1016/j.tra.2018.11.009

Miskeen M A, Alhodairi A M, Rahmat R A: Modeling a multinomial logit model of intercity travel mode choice behavior for all trips in Libya. International Journal of Civil and Environmental Engineering. 7, 636–645 (2013). https://doi.org/10.5281/zenodo.1087592

Rahman, M., Akther, M.S., Recker, W.: The first-and-last-mile of public transportation: A study of access and egress travel characteristics of Dhaka's suburban commuters. J Public Trans. 24, 100025 (2022). https://doi.org/10.1016/j.jpubtr.2022.100025

Ranjbari, A., Chiu, Y.C., Hickman, M.: Exploring factors affecting demand for possible future intercity transit options. Public Transport. 9, 463–481 (2017). https://doi.org/10.1007/s12469-017-0161-3

Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D., de Dios Ortúzar, J.: Sequential and Simultaneous Estimation of Hybrid Discrete Choice Models. Transportation Research Record: Journal of the Transportation Research Board. 2156, 131–139 (2010). https://doi.org/10.3141/2156-15

Román, C., Martín, J.C., Espino, R., Cherchi, E., Ortúzar, J. de D., Rizzi, L.I., González, R.M., Amador, F.J.: Valuation of travel time savings for intercity travel: The Madrid-Barcelona corridor. Transp Policy (Oxf). 36, 105–117 (2014). https://doi.org/10.1016/j.tranpol.2014.07.007

Shrestha, N.: Detecting multicollinearity in regression analysis. Am J Appl Math Stat. 8, 39–42 (2020). https://doi.org/10.12691/ajams-8-2-1

Simonson, I.: The effect of purchase quantity and timing on variety-seeking behavior. Source: Journal of Marketing Research. 27, 150–162 (1990). https://doi.org/10.2307/3172842

Tam, M.L., Lam, W.H.K., Lo, H.P.: The impact of travel time reliability and perceived service quality on airport ground access mode choice. Journal of Choice Modelling. 4, 49–69 (2011). https://doi.org/10.1016/S1755-5345(13)70057-5

van der Waerden, P., van der Waerden, J.: The relation between train access mode attributes and travelers' transport mode-choice decisions in the context of medium- and long-distance trips in the Netherlands. Transp Res Rec. 2672, 719–730 (2018). https://doi.org/10.1177/0361198118801346

Wang, K., Bhat, C.R., Ye, X.: A multinomial probit analysis of Shanghai commute mode choice. Transportation (Amst). 50, 1471–1495 (2023). https://doi.org/10.1007/s11116-022-10284-x

Wang, Y., Li, L., Wang, L., Moore, A., Staley, S., Li, Z.: Modeling traveler mode choice behavior of a new high-speed rail corridor in China. Transportation Planning and Technology. 37, 466–483 (2014). https://doi.org/10.1080/03081060.2014.912420

Wang, Y., Wu, B., Dong, Z., Ye, X.: A joint modeling analysis of passengers' intercity travel destination and mode choices in Yangtze River Delta megaregion of China. Math Probl Eng. 5293210, (2016). https://doi.org/10.1155/2016/5293210

Wen, C.H., Wang, W.C., Fu, C.: Latent class nested logit model for analyzing high-speed rail access mode choice. Transp Res E Logist Transp Rev. 48, 545–554 (2012). https://doi.org/10.1016/j.tre.2011.09.002

Wong, B., Habib, K.M.N.: Effects of accessibility to the transit stations on intercity travel mode choices in contexts of high speed rail in the Windsor–Quebec corridor in Canada. Canadian Journal of Civil Engineering. 42, 930–939 (2015). https://doi.org/10.1139/cjce-2014-0493

Yamamoto, T., Komori, R.: Mode choice analysis with imprecise location information. Transportation (Amst). 37, 491–503 (2010). https://doi.org/10.1007/s11116-009-9254-4

Yang, H., Feng, J., Dijst, M., Ettema, D.: Mode choice in access and egress stages of high-speed railway travelers in china. J Transp Land Use. 12, 701–721 (2019). https://doi.org/10.5198/jtlu.2019.1420

Yang, M., Wang, Z., Cheng, L., Chen, E.: Exploring satisfaction with air-HSR intermodal services: A Bayesian network analysis. Transp Res Part A Policy Pract. 156, 69–89 (2022). https://doi.org/10.1016/j.tra.2021.12.011

Yang, M., Zhao, J., Wang, W., Liu, Z., Li, Z.: Metro commuters' satisfaction in multi-type access and egress transferring groups. Transp Res D Transp Environ. 34, 179–194 (2015). https://doi.org/10.1016/j.trd.2014.11.004

Yao, E., Morikawa, T.: A study of on integrated intercity travel demand model. Transp Res Part A Policy Pract. 39, 367–381 (2005). https://doi.org/10.1016/j.tra.2004.12.003

Zhang, A., Wan, Y., Yang, H.: Impacts of high-speed rail on airlines, airports and regional economies: A survey of recent research. Transp Policy (Oxf). 81, A1–A19 (2019). https://doi.org/10.1016/j.tranpol.2019.06.010

Zhen, F., Cao, X., Tang, J.: The role of access and egress in passenger overall satisfaction with high speed rail. Transportation (Amst). 46, 2137–2150 (2019). https://doi.org/10.1007/s11116-018-9918-z

Zhou, H., Norman, R., Xia, J., Hughes, B., Kelobonye, K., Nikolova, G., Falkmer, T.: Analysing travel mode and airline choice using latent class modelling: A case study in Western Australia. Transp Res Part A Policy Pract. 137, 187–205 (2020). https://doi.org/10.1016/j.tra.2020.04.020