Transport Policy 176 (2026) 103909

Contents lists available at ScienceDirect

Transport Policy

journal homepage: www.elsevier.com/locate/tranpol

Estimating run choice models with innovative data collection: day-to-day tickets evolution in High Speed Rail

Francesco Russo, Giuseppe Musolino *, Domenico Sgro

DIIES – Dipartimento di ingegneria dell'Informazione, delle Infrastrutture e dell'Energia Sostenibile, Università Mediterranea di Reggio Calabria, Via Zehender, 89122, Reggio Calabria, Italy

ARTICLE INFO

Keywords:
High speed rail (HSR)
Travel demand
Run choice model
Ticket evolutions
Data collection

ABSTRACT

There are more than 110,000 km of planned and operative High-Speed Rail (HSR) lines in the world. A fundamental problem is to estimate the travel demand that uses HSR services. The paper considers models simulating the run choice of users among existing alternatives travelling between a given origin-destination pair. The research contribution concerns the proposal of a method for the identification of users from the observation of day-to-day tickets evolution. Ticket evolution constitutes a big and innovative dataset for national transportation system open to competition. The method is composed by two main parts. The former deals with the building of the choice set of alternatives, analysing the series of services tickets, and ends with the identification of user's choices. The latter phase deals with the specification and calibration of a run choice model. The proposed method has been tested on the relationship Rome-Milan (Italy), through the calibration of a disaggregated run choice model belonging to the class of random utility models. The obtained results can be important because give the possibility to update the model parameters from data obtained observing the ticket evolutions.

1. Introduction

High Speed Rail (HSR) lines generate effects on passengers' travel demand which are generally segmented into the three components: diverted, from other modes or other rail services; induced, in terms of trip frequency and destination generated by higher levels of service (e.g. reduction of travel time) due to HSR; economy-based, in terms of trip frequency and destination generated by the economy in the cities, or areas, served by HSR (Ben-Akiva et al., 2010; Givoni and Dobruszkes, 2013; Cascetta and Coppola, 2014; Russo et al., 2023).

Several studies have been carried out on HSR demand models in the scientific literature with the spread of HSR lines in the world. There was a diffusion of studies on models for the estimation of diverted demand, due to the attempt of capturing the demand diversion from the air mode and conventional rail services, as the opening of HSR lines caused competition in the inter-modal level (e.g., mode choice) and in intra-modal levels (e.g., service, company and run choice) (Cascetta and Coppola, 2012; Givoni and Dobruszkes, 2013).

The intercity travels for the HSR services (Di Gangi and Russo, 2023), thus the different competition levels among the diverted component of the demand are influenced by the fares' structures. Therefore, the fare structures may be classified into static and dynamic over time. On one hand, the static typology of fares is generally referred to the local rail services and to the bus services. On the other hand, the dynamic fare historically has characterized the air services and, in the last decades, it has been adopted also in the High Speed Rail (HSR) ones (see, among the others, Bergantino and Capozza, 2015; Malighetti et al., 2015; Russo et al., 2024).

According to the above considerations, some crucial themes emerge for the study of travel demand in the transport systems at national-scale in presence of inter-modal and intra-modal competition. Inter-modal competition has been the most studied in the literature and implies the possibility of using different attributes, while intra-modal competition, less studied, mainly refers to attributes connected to fares and departure times.

There are two important themes in intra-modal competition:

- The run choice model;
- The evolution of fares in the days before the trip.

The two themes are studied separately in the literature. The possibility of using innovative methods for collecting data concerning the

* Corresponding author.
E-mail address: giuseppe.musolino@unirc.it (G. Musolino).

https://doi.org/10.1016/j.tranpol.2025.103909
Received 29 May 2025; Received in revised form 18 October 2025; Accepted 16 November 2025
Available online 18 November 2025

evolution of fares allows the two themes to be combined, providing a decisive contribution to the development of sustainable transport policies in line with the goals and targets of Agenda 2030 (UN 2015, UN 2018). This, in order to allow the massive use of HSR systems which are the most sustainable for intercity trips at a national scale in terms of environmental (Dalkic et al., 2017; Tuchschmid, 2010; Baumeister, 2019; IlSole24Ore, 2024), economic (Campos et al., 2009) and social components (UIC, 2018).

The present work aims to put together the two above research lines that until now have been developed independently: the development of run choice models to estimate users' behaviour in the choice of run in the schedule-based intercity transport services (Cascetta et al., 1996; Nuzzolo and Russo, 1998) and the analysis and estimation of dynamic fare in the HSR and airline competition (Y.B. Xiao et al., 2008; Yao et al., 2013; Bergantino and Capozza, 2015; Malighetti et al., 2015). Specifically, the paper deals with models simulating choice behaviour of a HSR run of users driven by the dynamic patterns of fares.

The research contribution concerns the definition of a framework for the identification of users' choice of the HSR run, as well as the development of choice model in the above-mentioned dimension. The procedure is composed by two main phases. The former phase deals with the building of the choice set of alternatives, analysing the patterns of dynamic fares of HSR services, thus with the identification of user's choices. The latter phase deals with the specification, calibration and validation of a run choice model. The proposed method has been tested on the Rome-Milan city-pair (Italy).

The proposed method is innovative because it allows the run choice models to be calibrated by means of data that are easy to obtain, inexpensive and continuously updateable; such as the traffic counts generally used for aggregate calibration.

The subsequent sections of the paper are organized in the following way. Section 2 is composed by different paragraphs: a description of discrete choice models and transport supply models with a literature review about run choice models and an insight about the dynamic structure of fare. Section 3 regards the specification of the proposed procedure for the building of the choice set of alternatives and for the identification of the user's choice in the run dimension, as well as the specification of the choice model. Section 4 contains the results of the application of the method to the Rome-Milan section city-pair (Italy), as test case. Specifically, it is described the process of identifying user choices within the choice set through the proposed procedure. Moreover, the paper shows the results of the calibration of the specified run choice model. Final section reports the conclusions of the research and the future developments.

This work is aimed to support transport planners and decision-makers to define sustainable transport policies in the evaluation of investment in HSR lines and services, by means of methodological and modelling tools to assess the potential HSR travel demand.

2. Literature review

This section is organized into specific paragraphs. Paragraph 2.1 contains a description of the travel demand and transport supply models. Paragraphs 2.2 concerns a literature review about the run choice models. Paragraphs 2.3 present an insight about the characteristics of fares in the intercity transport services. Paragraph 2.4 presents a literature analysis of studies containing the results of models' calibrations from data not obtained from a direct survey.

2.1. Transport supply and travel demand models

The reference theoretical framework upon the present study relies on the Transportation System Models (TSMs), based on the topological-behavioural paradigm (Ben-Akiva et al., 1984; Ortúzar and Willumsen, 2001; Cascetta, 2013).

Supply models estimate the performance of the transport infrastructures and services usually defined by means of (congested) cost functions. They are classified according to different criteria. The first criterion relies on the level of aggregation, identifying two approaches: topological (or disaggregated) and aggregated. Topological approach relies on network models, composed by links, nodes, and cost functions (e.g. time-flow relationship). Aggregate approach estimates aggregate performances (e.g. average travel time, average speed) as the effect of the technological and functional dimensions of transport infrastructures.

Network models may be classified into (Nuzzolo et al., 2002; Cascetta, 2013): synchronic, or static, models, and diachronic, or dynamic models. Synchronic models have nodes with no specific temporal coordinates, and the same node may be representative of events that happen in different instants; this approach is generally used to represent continuous services, where the services are available at every instant of the time and at every point of the space (e.g. road private services). Diachronic, or dynamic, models, have nodes with explicit temporal coordinates, and each node is representative of an event that happen in a precise instant; this approach is generally used to represent discontinuous services, where the services are available at some instants of the time and in some points of the space (e.g. transit services).

As far as concerns the context of discontinuous services, two different modelling approaches are present in the literature. Line-based supply models are composed by two sub-models: the first is relative to the representation of the services; the second is relative to the access/egress network to/from the public transport system. Run-based supply models represent the services in terms of runs. Each run can be described by means of a sub-graph whose nodes represent the arrival and departure times of the vehicles at stops and whose links represent the trips from one stop to another or the dwelling at a given stop.

Travel demand models estimate the users' behaviour, in terms of trip choices, according to (congested) transport costs. Travel demand models may be differently classified in the literature.

A first classification outlines the models in relation to the existing variables (attributes), according to whether they relate to an aggregate of users, aggregated models, or to a single user, disaggregated models (Ben-Akiva et al., 1984; Ortúzar and Willumsen, 2001; Cascetta, 2013).

- Aggregated models rely on traffic flow data on some selected links (and nodes) of the network and they are commonly used due to the availability of the data to calibrate and validate the models.
- Disaggregated models rely on data about revealed/stated preferences of (potential) users and have greater theoretical and operational complexities, also linked to the less availability about users' choice behaviour.

A second classification leads to the segmentation of demand models into three classes.

- Statistical-descriptive models estimate the levels of demand through relationships with attributes belonging to the level-of-service and socio-economic category;
- Time series models utilize the historical data recorded in order to predict the demand flows, with fixed specifications;
- Partial-share models simulate the user choice process through a mechanism of partial sequential choices, or stages. The most common case is represented by multi-stage models, including generation, distribution, time choice (arrival/departure), service choice, route choice.

Partial-share disaggregated models are commonly used in the field of transportation systems engineering. They can be based on the theory of Random Utility (RU) (Ben-Akiva et al., 1984; Train, 2009; Ben-Akiva et al., 2019), Fuzzy Utility (FU) (Russo, 1997; Quattrone and Vitetta, 2011), or Quantum Utility (QU) (Vitetta, 2016, 2025; Di Gangi and Vitetta, 2021).

Commonly, the models vary in relation to the perceived utility function, by means of the different specifications among the various levels of choice (hierarchical or factorial), as well as the random residuals conduct towards two principal models types:

- models by choice probabilities specified in a closed form (e.g., Multinomial Logit, Nested Logit, …)
- models with simulated choice probabilities (e.g. Probit, Mixed Logit).

In the context of route choice, models rely on two main approaches: the frequency-based approach and the schedule-based approach. The run choice is discussed in the next section.

2.2. Run choice model

Two main approaches can be used in route choice modelling (Nuzzolo et al., 2000; Cascetta, 2013; Russo and Vitetta, 2003): the frequency-based approach and the schedule-based one. The frequency-based approach considers services in terms of sets of runs (lines). In this case line headways, or their inverse (the service frequencies), are considered explicitly. Therefore, the calculation of attributes users in relation to single runs is not possible. The schedule-based approach refers to services in terms of runs using the vehicle arrival-departure time. This approach allows us to take into account the evolution in time of both supply and demand, as well as run loads and level of service attributes.

The paragraph presents a brief literature review relative to run choice models in the context of High Speed Rail (HSR) and air transport. The research was carried out on the main scientific databases (e.g. Scopus, ScienceDirect) and the selected publications, in some cases, result to be more general respect to the issue proposed in this paper, thus dealing with other choice dimensions than the run choice one. The papers described below are clustered into two main groups concerning: air transport and High Speed Rail (HSR) (see Table 1).

On the HSR side, Nuzzolo et al. (2000) analysed the impacts of the railway service characteristics, such as travel time, fares, timetable, on the users' choices in the dimensions of mode, service (fast or slow), class (first or second) and run. The authors specified a choice model to simulate users' behaviour, in particular a nested-logit model, which takes into account the access/egress from/to the terminals and the desired departure/arrival times. The model has been calibrated with data obtained from a survey performed on an Italian railway lines. Yao and Morikawa (2005) modelled the intercity travel demand in Japan. Users' behaviour was analysed with discrete choice models, the nested-logit model. The data were supplied by a survey and by aggregate data. The model included the trip generation, the destination choice, the mode and run choices. Cascetta and Coppola (2012) forecasted the passenger demand with a schedule-based assignment model considering Italian HSR scenarios. The simulation was developed after the specification and calibration of a nested logit model among the mode-service-company-run levels. The model has been calibrated on the base of the data provided by traffic count and RP/SP survey carried out between the years 2009 and 2011, which take into account fares, travel times, access/egress times, runs departure times, HSR company as main attributes. The model was applied in an Italian case study on the Rome-Milan and Rome-Naples relationships. Cascetta and Coppola (2016) proposed a framework with the objective to predict passenger flows on individual rail run by developing a system of models composed by a national demand growth model, a multi-step demand model (generation/distribution/mode/service choice), and a stochastic assignment model. A random utility model has been specified, in particular a nested-logit model, for the choice dimensions of mode/service/class run. The model was estimated on an Italian case study, in which there is company competition in the context of High Speed Rail services.

On the air transport side, Anderson and Wilson (2003) explored the context of users' ticket purchase behaviour. The paper proposes a model to simulate users' choice probability in the case of reopening of a closed fare class. The authors objective is to explore the travellers' choices between waiting before buying a ticket with the objective of a cheapest fare occasion or purchasing a different type of ticket. Espino et al. (2008) discuss and analyse the users' willingness to pay for change ticket class in the Spanish airline services, thus the travellers' preferences in the case of different service offered by companies. The authors calibrated a multinomial logit and mixed-logit models with different configurations of the systematic utility, which includes level of service variables, price, penalty for ticket changes, service frequency, comfort, reliability and socioeconomic variables. Jung and Yoo (2014) examined the mode-service choice behaviour of travellers among low cost airlines, full service airlines and High Speed Rail (HSR) services in the context of domestic flights in South Korea. The companies' competition was simulated by specifying a multinomial logit and nested logit models, where the utility was composed of socioeconomic and level of service (fare, access time, frequency, travel time) variables.

Table 1
Literature review on run choice models in air and HSR mode-services.

Paper	Year	HSR	Air	Country	Choice dimension	System of models [Y/N]
Nuzzolo et al.	2000	x		Italy	Mode-service-class-run	Y
Anderson and Wilson	2003		x	–	Ticket	N
Yao and Morikawa	2005	x		Japan	Mode-service-run	Y
Espino et al.	2008		x	Spain	Ticket class	N
Cascetta and Coppola	2012	x		Italy	Mode-service-company-run	Y
Jung and Yoo	2014		x	South Korea	Mode-Service	Y
Cascetta and Coppola	2016	x		Italy	Mode/service	Y

2.3. Day-to-day ticket evolution in the intercity scheduled transport services

Ticket may be characterized by a static or dynamic structure. The static structure is typical associated to the regional and local rail transport, whereas historically the dynamic structure of fares is associated to the airline services.

However, nowadays the High Speed Rail (HSR) services are also characterised by the dynamic structure of fares. Hence, the dynamics implies that the fares value change over time, caused by several factors which could be classified into two groups respect to the travellers' behaviour of buying tickets: exogenous factors and endogenous factors. For instance, some endogenous factors are the time between the day of ticket purchase and the trip day, the level of seats occupancy of the train, while exogenous factors depend on HSR companies (pricing policy). In the case of Italian HSR companies, exogenous factors are mainly connected to the runs, the hour of the day and the day of the week, the availability of the fare in the last days before the trip day (NTV, 2013; Trenitalia, 2024).

More specifically, the travellers' behaviour of buying tickets could be different according to the typology of fare, which is typically associated to different complementary services, and according to the typology of user, which could be a business traveller ("time-based" behaviour) and a non-business traveller ("price-based" behaviour) (Xiao et al., 2008). On one hand, the time-based behaviour is supposed to choose the run and/or the ticket on the basis of temporal attributes, such as the minimum travel time, desired arrival/departure time. On the other hand, the price-based behaviour is supposed to choose the run and/or the ticket relying on the minimum ticket price available, thus on the minimum monetary cost.

As far as concern the above exogenous and endogenous factors, two groups of fare patterns may be identified: the supply-based fare pattern and the demand-based fare pattern.

As far as concerns the supply-based pattern, pricing strategies of high-speed rail and airlines companies are based on revenue management systems (see, among the others, McAfee and Te Velde, 2006; Talluri and Van Ryzin, 2006). These systems categorize fares into different booking classes, or fare buckets, and the availability of these classes is dynamically modified according to algorithmic predictions of demand, competitor pricing, overall seat occupancy reaching specific thresholds, and the time remaining until departure.

An important element to remark in the demand-based pattern, is that fare levels are on average lower where is present the competition among mode-services. A recent study showed that, in Italian city-pairs connections where there is no inter-modal competition (captive connections), an increase of 10 % in the airline modal share allowed companies to fix fares 7.1 % higher than less captive city-pairs connections. Moreover, the authors observed a J-shaped temporal patters of airline fares for city-pairs connections where inter-modal competition is present, while the patterns seemed monotonically increasing for more captive city-pairs connections (Bergantino and Capozza, 2015; Malighetti et al., 2015).

The demand-based group regards the services characterized by fares' changes due to company (pricing policy) and to the traveller's choice behaviour. In other terms, the impact of the demand, for instance the ticket purchasing behaviour contributes to modify the regular evolution of the fare curves. Typically, a demand-based fare pattern may show an evolution similar to the ones depicted in Figs. 1 and 2. The two figures show the non-regular changes over time of the ticket fares. In this cases it is possible to suppose that there is the decisive contribution of the travel demand. Specifically, Fig. 1 shows the case of evolution of HSR fares for one run and for different trip days; while Fig. 2 shows the case of evolution of HSR fares over time considering three HSR runs for a given trip day. The irregular variations of the value of the fares over time excludes that the variations are due to a mere intervention of companies. (Anderson and Wilson, 2003; Bergantino and Capozza, 2015; Malighetti et al., 2015).

2.4. Data collection methods

An in-depth literature review has been carried out on studies that consider an aggregate database given by traffic counts on the links of the transport network. The technique of traffic counts is currently one of the advanced alternative data collection method compared to direct interviews. The analysis of the literature aims to verify how the traffic counts traditionally used for private modes at an urban scale, have been spread in public transport systems at an extra-urban and national scales.

The analysis was developed on the basis of the technique called Snowball Sampling (Goodman, 1961), hence starting from the scientific databases: Scopus, Web of Science, Scholar. A total of 354 works were identified and considering the simultaneous presence of works in different databases the number of works was reduced to 163.

It was verified whether quantitative results of calibrations were reported in each work, thus obtaining 60 works. Two classifications were made on the 60 works.

1) Quantitative calibrations: 54 works aggregated and 6 works disaggregated.
2) Case studies: 36 works in urban context and 24 works in extra-urban context.

The combination of the above two classes leads to the following results:

- 31 works of case studies in urban context and aggregated calibrations;
- 23 works of case studies in extra-urban context and disaggregated calibrations;
- 1 work of case studies in extra-urban context and disaggregated calibrations;
- 4 works of case studies in urban context and disaggregated calibrations.

A summary of the 4 works of interest concerning case studies in urban contexts (Table 2) is reported below. The last right column of Table 2 reports the publication that is the source of each publication examined.

De Luca (2005) proposes a methodology which concerns a mixed disaggregate/aggregate specification and calibration system of models, with the aim of the simulation of the travel demand in relation to different scenarios, in which several data types are combined with results in terms of demand models considered by the author cheap and successful in the context of the simulation of the phenomenon considered. The proposed method has been tested in an Italian case study in an urban context. Nuzzolo et al. (2013) concerns the evaluation of the passengers on board for each surveyed vehicle, which has used the transport system, in addition with the investigation of the using of

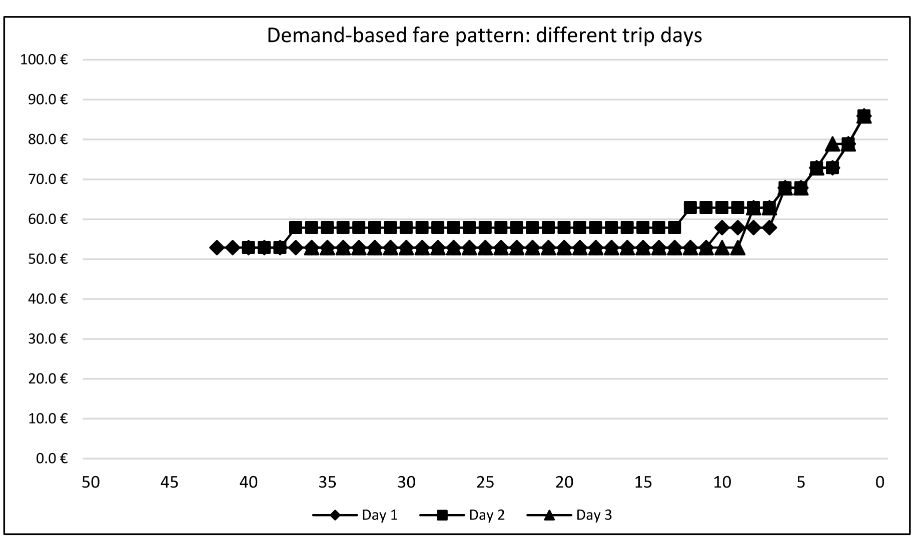

Fig. 1. Pattern of demand-based fares: case of one run and different days of trip.

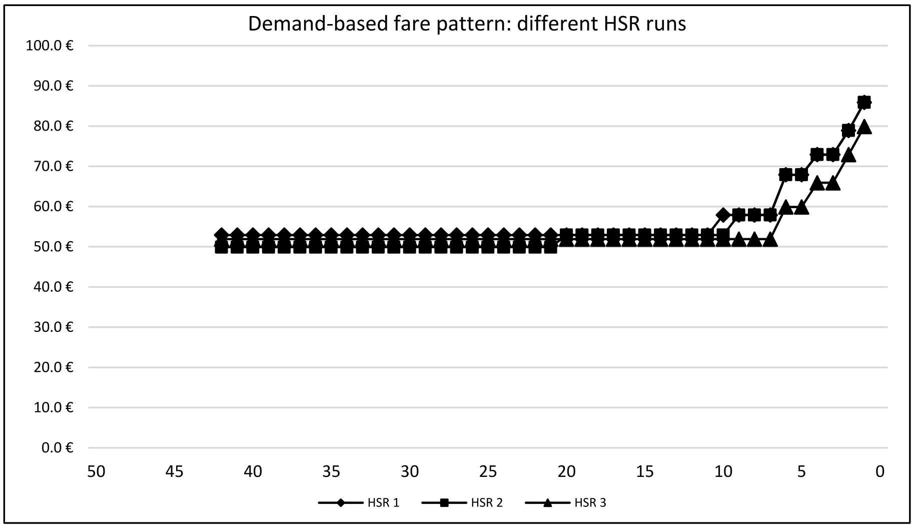

Fig. 2. Pattern of demand-based fares: case of different runs for the same day of trip.

Table 2
Results of the Snowball Sampling procedure.

Title of the paper	Author(s)	Year	Source
Demand modelling by combining disaggregate and aggregate data: an application to a displacement scenario for the Mount Vesuvius area	De Luca S.	2005	Cascetta and Russo, 1997
STOP: A Short term Transit Occupancy Prediction tool for APTIS and real time transit management systems	Nuzzolo A, Crisalli U, Rosati L, Ibeas A.	2013	Russo and Vitetta (2011)
Integrated travel demand models for evacuations: a bridge between social science and engineering	Russo F, Chilà G	2014	Russo and Vitetta (2011)
Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning	Siripirote T, Sumale A, Watling DP, Shao H.	2014	Cascetta and Russo, 1997

real-time information on users counts. The authors apply a specified procedure, called Short-term Transit Occupancy Predictor tool (STOP) in order to achieve the objective, as well as the method includes the updating of the choice model parameters. Moreover, the method was tested in a pilot case study in the Spanish country for an urban context. Russo and Chilà (2014) deals with the building of travel demand models for risk scenarios, hence in particular for evacuations. It is proposed a classification and a description of the different types of hazardous scenarios, in terms of effects, hence for the temporal evolution of the event. It is analysed the behavior of the users by means of specification and calibration of behavioral models, based on data provided by Stated Preference (SP) surveys. The proposed method was tested in an urban context in a municipality of the southern Italy. Siripirote et al. (2013) proposes a work concerning a procedure for updating the parameters of the behavioral model, characterized by several dimensions in terms of choice levels, by means of a Maximum Likelihood (ML) method, as well as the estimation of the trip chain of the vehicle, with data provided by sensors embedded in the roads. The adopted solution algorithm is the Expectation-Maximization (EM). The proposed solution has been tested in a modified Sioux Falls network, thus an urban context.

2.5. Gaps in the literature and research contribution

The literature review shows that the two research lines considered in this work until now have been developed independently. Few papers were presented in the field of run choice models to estimate users' behaviour in the choice of run in the schedule-based intercity transport services (air and HSR), while a more extended research is present in the field of the analysis and estimation of fare evolution in the HSR system.

The present work aims to put together the two research lines using an innovative method for collecting data about tickets evolution:

- the analysis and estimation of dynamic fares of HSR services;
- the development of run choice models to estimate users' behaviour in the choice of run in the schedule-based intercity transport service.

The paper deals with models simulating choice behaviour of a HSR run of users according to dynamic patterns of fares. The research contribution concerns the definition of a method for the identification of users' choice of the HSR run. The specification-calibration of a choice model in the run dimension supports the validation of the proposed method.

3. Formulation of the proposed method

3.1. Definitions and general notation

Let's:

- i, user
- t, day of trip of user i
- K, interval of days of ticket purchasing before t;
- $k \in K$, generic day of ticket purchasing; where $k = 0$ coincide with the day t of travel (it is worth noting that k increases as the interval between day of ticket purchasing k and day of trip t increases);
- R_k, set of runs available at day k for travelling at day t;
- r, r', r''$\in R_k$, available runs at day k (belonging to the set R_k)
- G, maximum number of fare types (classes) for all runs and all companies at day of trip t;
- g, generic type (class) of fare $g \in G$;
- $w_k(r, g)$, fare of run r at day k for fare type (class) g;
- $\mathbf{W}_k[R, G]$, matrix of fares, composed of a number of rows equal to the number of runs, R, of all companies at day k, and a number of columns equal to the total number of fares (classes) offered by all companies.
- $\Delta\mathbf{W}_k[R, G]$, matrix of fare differences, which has the same structure of $\mathbf{W}_k[R, G]$ and where each element $\Delta w_k(r, g)$ is obtained as difference between the value of the ticket at day k, $w_k(r, g)$, and the value of the ticket at day k+1, $w_{k+1}(r, g)$.

- $N_k = \sum_r N_k(r)$, number of fare changes (mutations) associated to the runs of choice set R_k in two consecutive days k+1 and k.

It is worth noting that complete formalization should be expressed as:

- R_k should be expressed in complete form as R_k^t (set of runs), highlighting that day k refers to day of trip t; furthermore, considering that for every day k the set of runs is always fixed in relation to day t, the set can finally be represented as R;
- K should be expressed as K^t highlighting that the set of runs refers to day t
- k should be expressed as k^t, highlighting that the generic day k refers to the day of trip t.

The superscript t is not reported below because in the formalizations t is assumed to be fixed. The superscript t is inserted in the cases when it is necessary to avoid doubts between the two different days of trip.

The following two paragraphs present the steps of the proposed framework, which is composed of two parts (as depicted in Fig. 3a and 3b):

- identification of the chosen run (par. 3.2);
- run choice (par. 3.3).

3.2. *Identification of the chosen run*

This section presents the part of the proposed framework for identifying the run chosen by a user. The hypothesis considered is that the users' choice of run can be identified in relation to the day-to-day ticket evolutions observed between two consecutive days, considering when the user buys the ticket before the day of trip.

The ticket evolution is considered as the innovative data collection because it has a huge quantity of information that needs specific collection but gives the possibility to update the model in a cheap, repeatable and easy way.

The first part of the framework is composed of the three following steps (Fig. 3a):

- step 1. Ticket coding; where every fare structure of every company is described by means of a vector;
- step 2. Day-to-day ticket costs' evolution, where the day-to-day fare mutations are compared;
- step 3. Identification of number of mutations for each run in each day and association to each user's choice.

Step 1. Ticket coding

In general, each company offers a fare structure which, in the simplest (historical) form, is given by classes of progressive quality where the price increases with the class improvement. The evolution of transport mode-services allows to offer fare structures that depend on multiple factors, Therefore, the traditional vector of classes may become a two-dimensional, or a multi-dimensional, matrix. Every multi-dimensional structure can however be reduced to a one-dimensional vector for the generic run r that represents the fare for a set G of different classes.

The output of the step 1 is a matrix fares $\mathbf{W}_k$ [R, G] with $k \in K$ (see Table 3).

Step 2. Day-to-day ticket costs' evolution

The day-to-day ticket costs' evolution consists in building a new set of matrices differences, evidencing all the modified data referring to the day-to-day ticket series. The single matrix is given by:

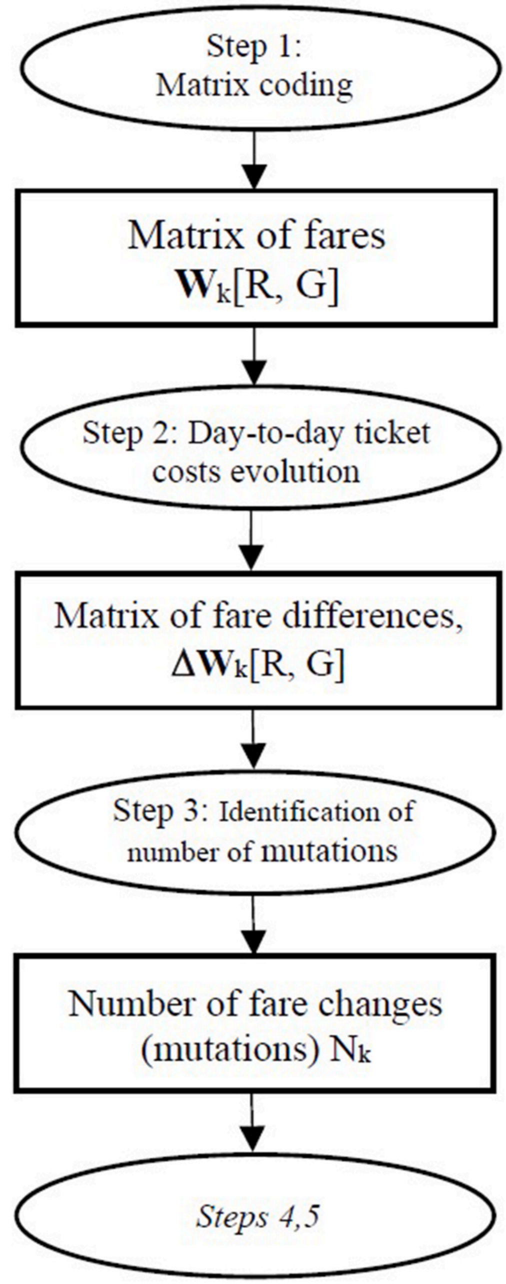

Fig. 3a. Procedure for the identification of chosen run.

$$\Delta\mathbf{W}_k[R, G] = \mathbf{W}_k[R, G] - \mathbf{W}_{k+1}[R, G] \tag{1}$$

Obtained as difference between the matrix of ticket costs, $\mathbf{W}_k$[R, G] at day k and the matrix of ticket costs $\mathbf{W}_{k+1}$[R, G] at previous day k+1.

The generic element of the above matrix $\Delta\mathbf{W}_k$[R, G] is obtained as difference between the value of the ticket at day k, $w_k(r, g)$, and the value of the ticket at day k+1, $w_{k+1}(r, g)$.

$$\Delta w_k(r, g) = w_k(r, g) - w_{k+1}(r, g) \tag{2}$$

The output of the step 2 is a matrix fares $\Delta\mathbf{W}_k$ [R, G].

Step 3: Identification of number of mutations

The step regards the identification of the number of mutations, that consists of counting the mutations, for each run r, in each day k. A ticket mutation is associated to a ticket purchase of (at least) a single user.

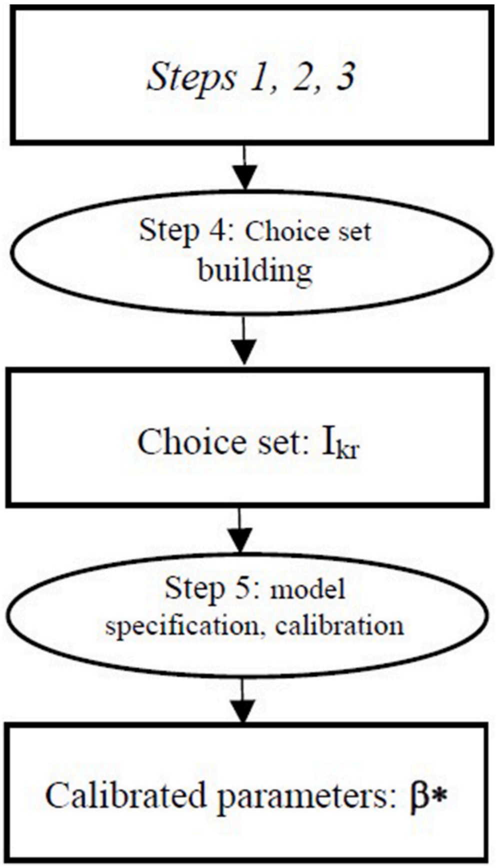

Fig. 3b. Procedure for the run choice.

Table 3
General matrix of fares: $\mathbf{W}_k$[R, G].

	g = 1	g = 2	g = 3	g = 4	g = G
r=1	$w_k(1,1)$	$w_k(1, 2)$	$w_k(1,3)$	$w_k(1,4)$	$w_k(1,G)$
r	$w_k(r,1)$	$w_k(r,2)$	$w_k(r,3)$	$w_k(r,4)$	$w_k(r,G)$
r=R	$w_k(R,1)$	$w_k(R,2)$	$w_k(R,3)$	$w_k(R,4)$	$w_k(R,G)$

Once the matrix $\Delta\mathbf{W}_k$[R, G] is calculated for ∀ k ∈ K, the identification consists in the definition of $\mathbf{W'}_k$[R, G], which is a matrix of coded changes obtained from the matrix $\Delta\mathbf{W}_k$[R, G].

Each element $w'_k(r, g)$ may assume the following values:

$$w'_k(r,g) = \begin{cases} 0, & \text{if } \Delta w_k(r,g) = 0 \\ 1, & \text{if } \Delta w_k(r,g) \neq 0 \end{cases} \tag{3}$$

Assuming that fares evolve as a result of demand, the modification of the fare level indicates that at least one traveller has bought that ticket.

According to the above hypothesis and given the matrices $\mathbf{W'}_k$(R, G) ∀ k ∈ K, it is possible to evaluate the number of choices for every run, which is equal to:

$$N_k(r) = \sum_g w'_k(r,g) \tag{4}$$

with $N_k(r)$, number of ticket changes associated to run r (∈R) in two consecutive days k+1 and k.

The sum of $N_k(r)$ on all the runs provides the number of changes associated to the runs of choice set R in two consecutive days k+1 and k, N_k, which is the output of the step 3:

$$N_k = \sum_r N_k(r) = \sum_r \sum_g w'_k(r,g) \tag{5}$$

3.3. Run choice

This section presents the second part of the proposed framework related to run choice, which is composed of the two following steps (Fig. 3b):

Step 4: Choice set building;
Step 5: Run choice.

3.3.1. Choice set building

The simulation of choice set generation has been widely studied for the path choice dimension. In this case, a choice set perception model was proposed to build perceived path choice sets (formation level) and to estimate the probability associated to each perceived path choice set (extraction level). Paths belonging to the path choice set may be generated with a mono-criterion approach, if users are supposed to generate paths according to the same criterion, or multi-criteria approach, if users are supposed to generate (one or more) paths for each criterion considered. In general, according to Manski (1977), a probability can be associated to each path choice set. The main papers dealing with the above models are the following: Manski (1977); Ben-Akiva et al. (1984); Russo and Vitetta (1995); Morikawa (1996); Ben-Akiva et al. (2002); Cascetta et al. (2002).

The paper considers, as first case with the objective to validate the proposed framework, the application of a mono-criterion approach based on the minimum "distance" of a couple of early and late runs from the Desired Arrival Time (DAT) of user at destination. Other single criteria in the mono-criterion approach may be considered for the formation of path choice set like the one consisting in the minimum price of tickets. Moreover, a combination of the several criteria may be considered in a multi-criteria approach.

Given the generic run choice $w'_k(r, g) = 1$, the choice set of the user that chooses the run r in the day k is I_{kr}.

The DAT is identified by assuming that the user has chosen the run that minimizes the delay/advance respect to the DAT itself.

Considering a diachronic approach, if the time period considered is sub-divided into intervals of the same width and the average point of the time period is indicated with δ_z, the DAT of the user who has chosen the run r is given by (see Fig. 4):

$$DAT = \delta_z : \min_z |\tau_r - \delta_z| \tag{6}$$

where τ_r is the expected arrival time of the run r that had the fare mutations.

It is assumed that the choice set is composed by alternative runs operated at day t, considering the price of tickets present at day k.

It is possible to build the choice set with different criteria. Among them, the most interesting are the following:

a) the run r chosen by the user and the alternative one closest to DAT are considered:
b) the runs inside the time period: $I_{kr} \equiv R_k$.

A binomial choice model may be used by assuming the hypothesis a) introduced above for the choice set. Given the expected arrival time of the run r, τ_r, the two runs closest to δ_z are the following:

I. if the run r is in advance respect to the DAT, the alternative run, r_{alt}, is the one delayed closest to the DAT;

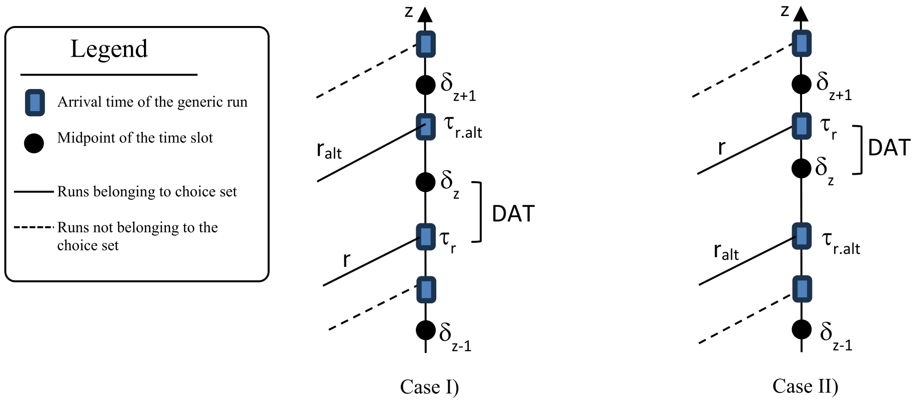

Fig. 4. Examples of runs that compose the choice set.

II. if the run r is delayed respect to the DAT, the alternative run, r_{alt}, is the one in advance closest to the DAT.

3.3.2. Choice model

The choice models rely on random utility theory (Ben-Akiva and Lerman, 1985; McFadden, 2001; Cascetta, 2013). The main assumption is that a user i chooses a run r among the set of perceived alternatives I_{kr}, maximizing her/his associated perceived utility, U_r^i.

As the utility, U_r, is a random variable, is it not possible to establish with certainty which run is chosen by the user i. However, it is possible to express the probability that the run $r \in I_{kr}$, will be chosen by user i as the probability, p_r^i, that the perceived utility of run r, U_r^i, is greater than the perceived utility of every other run r', $U_{r'}$, with $r, r' \in I_{kr}$. Indicating with $\boldsymbol{\beta}$ the vector of parameters it is possible to write:

$$p_r^i(\boldsymbol{\beta}) = \text{Prob}\left(U_r^i(\boldsymbol{\beta}) > U_{r'}^i(\boldsymbol{\beta})\right)\ \forall r' \neq r, r, r' \in I_{kr} \tag{7}$$

Model (7) may be differently specified. The specification considered here is the binomial Logit (descending from the generation of $I_{r,k}$ according to hypothesis a):

$$p_r^i(\boldsymbol{\beta}) = \exp\left(V_r^i(\boldsymbol{\beta})\right) \Big/ \left(1 + \exp\left(V_r^i(\boldsymbol{\beta}) - V_{r'}^i(\boldsymbol{\beta})\right)\right) r', r \in I_{kr} \tag{8}$$

where:

$V_r^i(\boldsymbol{\beta}) = E[U_r^i]/\theta = \boldsymbol{\beta}^T \bullet \mathbf{x}$
$E[U_r^i]$, expected value of utility associated to the run r by user i;
θ, parameter of Logit model;
$\boldsymbol{\beta}$, vector of parameters to be calibrated (or updated);
$\mathbf{x}$, vector of attributes.

3.3.3. Likelihood function specification

The calibration allows to obtaining estimates of vector of parameters, $\boldsymbol{\beta}$, according to users' choices. The method of parameter estimation used is the Maximum Likelihood (ML), which provides the values of the unknown parameters that maximize the probability of observing the users' choices. The probability of observing the choices of a users' sample, or likelihood of the sample, depends on the choice model and on the sampling strategy (Cascetta, 2013).

In the case of simple random sampling, the observations are statistically independent and the probability, or likelihood, of observing the set of choices of the users' sample is given by the product of the probabilities that each user i chooses the run r.

Since the probabilities $p_r^i(\boldsymbol{\beta})$ depend on the vector $\boldsymbol{\beta}$, also the probability L of observing the entire sample is a function of the unknown parameters:

$$L(\boldsymbol{\beta}) = \prod_{i=1\ldots n} p_r^i(\boldsymbol{\beta}) \tag{9}$$

The estimate of maximum likelihood, $\boldsymbol{\beta}^*$, of vector of parameters, $\boldsymbol{\beta}$, is obtained by means of eq. (7):

$$\boldsymbol{\beta}^* = \max L(\boldsymbol{\beta}) = \max \prod_{i=1\ldots n} p_r^i(\boldsymbol{\beta}) \tag{10}$$

Different likelihood functions may be specified considering different cases, as reported in the following.

- <u>Likelihood function L</u> for all days $k \in \underline{K}$

For the generic day k, recalling eq. (4) that defines the number of mutations for each run r, the value of L_k is equal to:

$$L_k = \prod_{r \in R_k} p_{kr}^{N_k(r)} \tag{11}$$

where: p_{kr} is the probability of choosing the run r at day k.

In the general case of multiple runs chosen at day k, each run has been chosen $N_k(r)$ times, the L_k function for the day k is given by eq. (11) and the overall L function, for all days $k \in K$, is given by:

$$L = \prod_{k \in K} L_k = \prod_{k \in K} \prod_{r \in R_k} p_{kr}{}^{N_k(r)} \tag{12}$$

- <u>Likelihood function L</u> for each class of fare

Given a class of fare, g, chosen in a day k for travelling on the run r and considering the set of alternatives composed of the different runs, the specification of the likelihood function is:

$$L_{kr}(g) = \prod_{g \in G} p_{kr}(g)\ \forall g : w'(r, g) = 1 \tag{13}$$

The condition: $w'(r,g) = 1$ ensures that the considered probabilities, $p_{kr}(g)$, are the ones associated to the observed runs chosen by the travelers.

By aggregating for all runs r, the likelihood function becomes:

$$L_k = \prod_{r \in R_k} L_{kr}(g) = \prod_{r \in R_k} \prod_{g \in G} p_{kr}(g) \tag{14}$$

Finally by considering all days k, the likelihood function is:

$$L = \prod_{k \in K} L_k = \prod_{k \in K} \prod_{r \in R_k} \prod_{g \in G} p_{kr}(g) \tag{15}$$

It is possible to consider other specifications of the likelihood function considering further cases. The first could be the case where the run is given, $r = r^*$, and the class of fare is variable. The second could be the case where both run and class of the fare are variable. In a more general way is possible to consider also the variation of day of trip t.

Fig. 5 summarizes the research contributions of the paper in the process of choice set building and of the likelihood specification. The

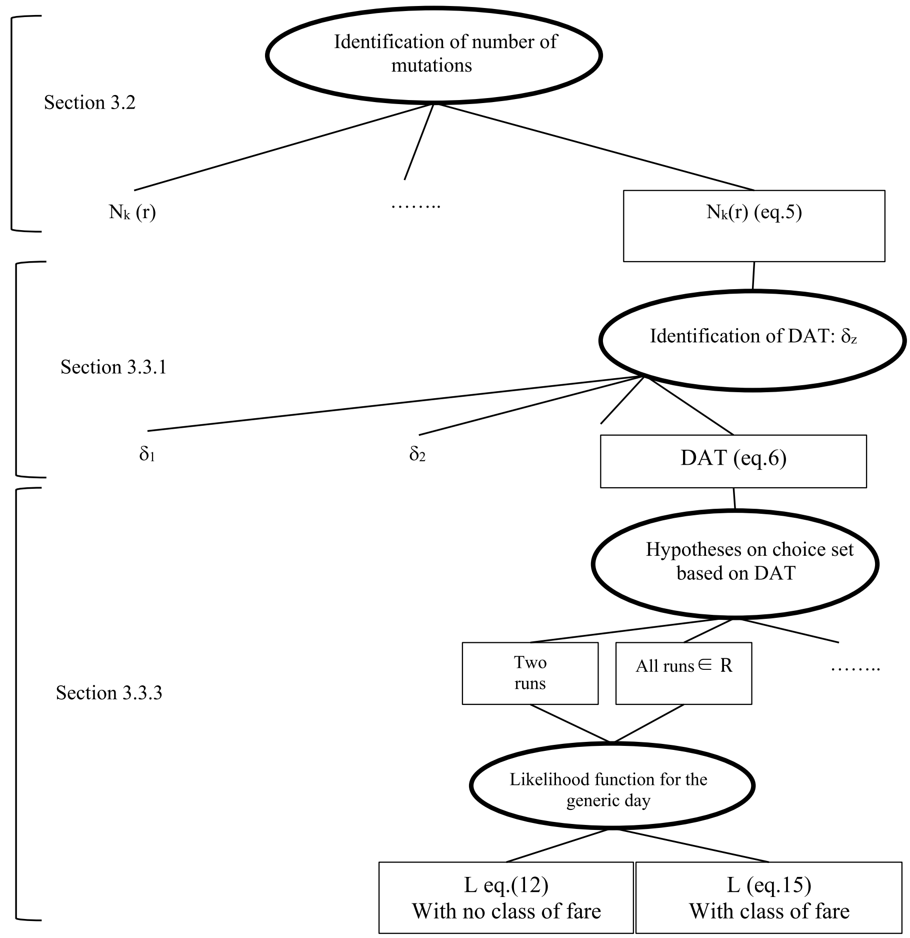

Fig. 5. Building process of the choice set and of the likelihood function.

first contribution concerns the identification of number of mutations described in section 3.2. The second contribution concerns the choice set building based on DAT described in section 3.3.1. The third contribution concerns the specification of the likelihood function described in section 3.3.3.

4. Results of proposed method

The method proposed in section 3 have been validated considering a real experimental test site of HSR line existing along the Rome-Milan pair.

The section is sub-divided into several paragraphs. Paragraph 4.1 summarizes the characteristics of the Rome-Milan HSR line. Paragraph 4.2 describes the application of the individual steps of the proposed procedure. Paragraph 4.3 presents the results of the calibration of the run choice model.

4.1. HSR line between the Rome-Milan pair (Italy) and survey on tickets evolution

The HSR line between the Rome-Milan pair is the backbone of the Italian HSR network. As far as concerns the supply, the line belongs to traditional Basic Network for a short infrastructural stretch and belongs to the High Speed/High Capacity Basic Network for the Florence-Milan infrastructural stretch equipped with ERMTS system. The line has 447 km of length, and is completely electrified with double track (www.rfi.it). The frequency of HSR services is currently 33 runs/day for Trenitalia company and 32 runs/day for Nuovo Trasporto Viaggiatori - NTV company.

As far as concerns the demand, it is estimated that during the decade 2012–2022 about 7 million passengers diverted from conventional railway services to HSR services and that 19 million diverted from other transport modes (private car, buses, and air) to HSR services, with 17 million generating this demand (Russo, 2021).

The survey was carried out by examining the tickets every day available on the websites of Trenitalia (www.trenitalia.com), and of Nuovo Trasporto Viaggiatori - NTV (www.italotreno.it), with the exclusion of seasonal tickets. The survey period ranges from 2nd of August 2023 to 8th of September 2023.

The survey was conducted every day k (e.g. August 2nd 2023, etc.) on ticket costs assuming two days of the trip t: t_1 = September 1st' 2023 and t_2 = September 8th' 2023 (Fig. 6).

4.2. Identification of the chosen run

Step1: Ticket coding

Let's consider:

$t = t_1$ = September 1st' 2023
$K=K_1=30$ days
$k+1 = 10$ (August 23rd' 2023) and $k = 9$ (August 24th' 2023) belonging to K.

The matrix coding is operated on an example of HSR fares and the output is a vector for the two days of purchasing $k = 10$ and $k = 9$, as shown in Table 4a and 4b.

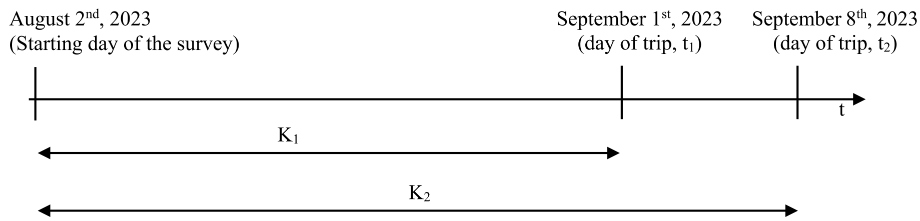

Fig. 6. Survey periods (K) and days of trip (t).

Tab. 4a
Matrix $\mathbf{W}_{k=10}[r = r_5, G = 15]$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r_5	95 €	112 €	129 €	129 €	295 €	72,9 €	79,9 €	91,9 €	91,9 €	–	54,9 €	57,9 €	–	–	–

Tab. 4b
Matrix $\mathbf{W}_{k=9}[r = r_5, G = 15$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r_5	95 €	112 €	129 €	129 €	295 €	72,9 €	79,9 €	91,9 €	91,9 €	–	54,9 €	61,9 €	–	–	–

Step 2. Day-to-day ticket costs' evolution

The matrix $\Delta\mathbf{W}_{k=9}[r_5, 15]$ is obtained from eq. (2): $\Delta\mathbf{W}_{k=9}[R, 15] = \mathbf{W}_9[r_5, 15]$ - $\mathbf{W}_{10}[r_5, 15]$.

The analysed element that differ from day k+1 to day k is the element positioned in g = 12.

Table 5 presents the value $\Delta w_9(r_5, 12) = w_{10}(r_5, 12) - w_9(r_5, 12) = 61{,}9$ € $-57{,}9$ € $= 4{,}0$ €.

Step 3: Identification of number of mutations

The matrix $\mathbf{W'}_{k=9}$ is obtained from eq. (3) (see Table 5), where each element $w_9(r_5, 15)$ may assume the values '0' or '1'. Table 6 presents the matrix where $w'_9[r_5, 12] = 1$.

The individual mutation, reported in Table 6 and is clustered and summed in relation to each day k and run r, obtaining the number of mutations, $N_k(r)$. Then, the mutations are summed for all runs r, obtaining N_k, which is the output of the step 3.

Given five days considered {k = 12, k = 11, k = 10, k = 9, k = 8} as a subset of K, the number of mutations for each day k is calculated for four days: $\{N_{12} = 0, N_{11} = 5, N_{10} = 11, N_9 = 14\}$. The total number of travelers identified in this example is N = 30. Table 7 contains the above operations on N = 30.

4.3. Run choice model

4.3.1. Choice set building

The choice set building refers to the case I described in section 3.3.1 (Fig. 4), concerning the identification of the DAT, the calculation of the penalty and the identification of the two runs that compose the choice set.

The runs belonging to R are extracted among the runs present in the timetable of runs supplied by the company Trenitalia with the following characteristics:

- runs operating along a given origin-destination pair: Rome-Milan and vice versa;
- runs arriving at the station within a defined time window: 10:30 a.m. - 1:00 p.m.;
- runs having one intermediate stop at most.

The first run composing the choice set is the one chosen by the uses, r, which is the run where at least one fare change has been observed (see Fig. 7). The time axis z is discretized into time intervals of 30 min (11:00; 11:30; 12:00; 12:30) and the δ_z are identified as the midpoints of each time interval ($\delta_{z-2} = 10{:}45$; $\delta_{z-1} = 11{:}15$; $\delta_z = 11{:}45$; $\delta_{z+1} = 11{:}15$; $\delta_{z+2} = 12{:}45$). Therefore, the DAT is identified for case I, as the point that results from the calculation of the minimum distance between each δ_z and the arrival time τ_r of each run (eq. (6)). In the example presented, the chosen run is r = 9618, with a corresponding $\tau_{r4} = 11{:}58$, and the resulting DAT is $\delta_z = 11{:}45$. The value of penalty is $\Delta = |\delta_z - \tau_{r4}| = 11{:}45$–$11{:}58 = 13$ min. The second run belonging to the choice set is the first one arriving at the station in advance with respect to the DAT and, in this case, it is the run $r = r_6$ with $\tau_{r6} = 10{:}58$. The choice set is $I_{rk} = \{r_4, r_6\}$.

4.3.2. Specification of run choice model and calculation of likelihood

The functional form of the model for the estimation of probability that run r is chosen by a user is a binomial logit (eq. (8)).

Several specifications have been tested. The specification considered is:

$$\text{model 1}: V_r = \beta_w\, w_r + \beta_\Delta\, \Delta_r + \beta_t\, T_r \quad \forall r \in R \tag{16}$$

$$\text{model 2}: V_r = \beta_w w_r + \beta_t T_r + \beta_s\, s_r \quad \forall\, r \in R \tag{17}$$

The attributes specified in the run choice model are described below:

Table 5
Matrix $\Delta\mathbf{W}_{k=9}[r_5, 15]$.

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r_5	0	0	0	0	0	0	0	0	0	0	0	4,0€	0	0	0

Table 6
Matrix $\mathbf{W'}_{k=-9}$[9620, 15].

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
r_5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

Table 7
Example of calculation of total number of identified travellers (N = 30).

k	12	11	10	9	
r	$N_{12}(r)$	$N_{11}(r)$	$N_{10}(r)$	$N_9(r)$	
r_1	0	2	4	4	
r_2	0	0	3	4	
r_3	0	1	2	2	
r_4	0	0	0	0	
r_5	0	2	2	1	
r_6	0	0	0	0	
r_7	0	0	0	0	
r_8	0	0	1	2	
N_k	N_{12}=0	N_{11}=5	N_{10}=11	N_9=14	30

- w_r [€], ticket cost associated to run r at day k and belonging to class g;
- $\Delta_r = \tau_r - \delta_z$ [h] penalty defined as the time interval between the DAT (or in general δ_z for the alternative run), and the arrival time of the run r at the station τ_r;
- $T_r = \tau_{arr,r} - \tau_{dep,r}$ [h], on-board travel time from the origin to the destination of the run r, calculated as the difference between arrival time and departure time of run r at the departure/arrival railway stations;
- s_r, number of intermediate stops of run r between the departure and arrival railway stations.

The descriptive analysis has been executed on data related to a sample of N = 141 identified travellers extracted during a survey executed in 2023. The results are presented in Table 8. The average on-board travel time is $T_r = 3.33$ [h], with a standard deviation of 0.26 [h]. The penalty, Δ_r oscillates from −0.47 [h] to 0.25 [h]. The average value of the fare, $w_r = 87.77$ [€] with a standard deviation of 26.12 [€]. The average value of intermediate stops $s_r = 0.69$, with a standard deviation of 0.46.

The framework has been validated by means of the parameters' calibration of the systematic utilities specified in eqs. (16) and (17)

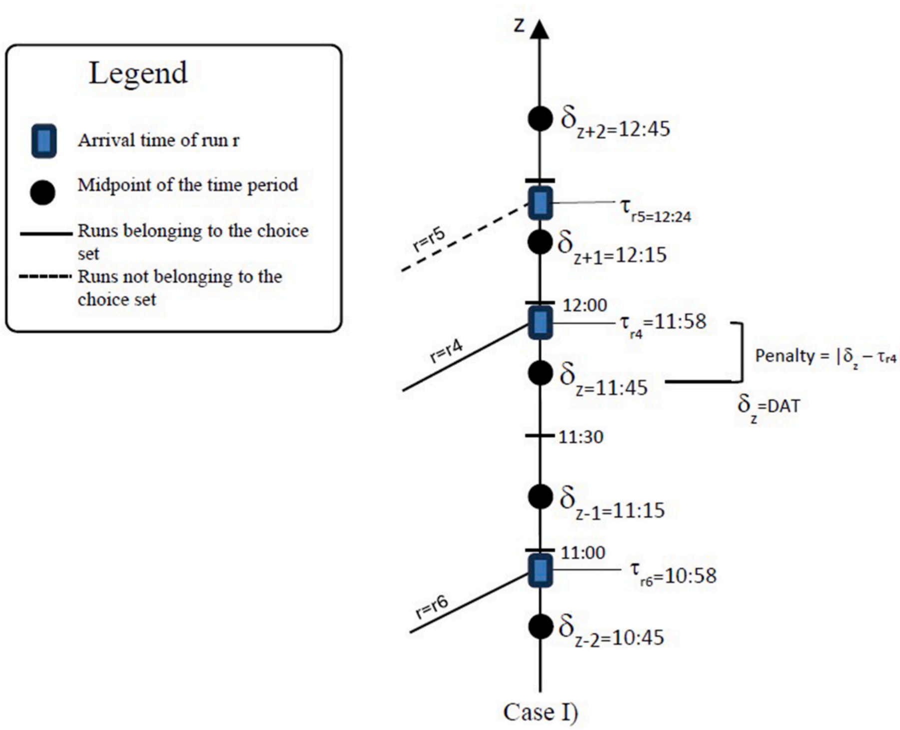

Fig. 7. Example of runs composing the choice set (Case I).

Table 8
Descriptive analysis of attributes considered (total number of identified travellers: N = 141).

Attribute	Symbol	U.o.M.	Range	Min	Max	Avg	S.D.
On-board travel time	T	h	[0; +∞]	2.98	3.98	3.33	0.36
Penalty	Δ	h	[-∞; +∞]	−0.47	0.25	−0.18	0.29
Ticket cost	w	€	[0; + ∞]	44.90 €	249.00 €	87.77 €	26.12 €
Intermediate stops	s		[0; 1]	0	1	0.69	0.46

U.o.M. = Unit of Measure; Avg = Average; S.D. = Standard Deviation.

Table 9
Results of the calibrated parameters.

Model		β_w	β_T	β_Δ	β_s
1	Parameter	−0.0242	−0.7188	−0.9304	
	(t-stud)	(-1.0132)	(-1.9632)	(-1.0885)	
2	Parameter	−0.0327	−1.1391		−0.7743
	(t-stud)	(-1.1662)	(-2.8773)		(-2.5915)

considering a binomial logit (eq. (8)). The specification of the likelihood function considered is reported in eq. (15).

Two models have been calibrated and the results are reported in Table 9 in terms of average value of parameter and value of t-student, as statistics to verify the significance of the value of parameter.

As far as concerns the calibrated parameters of the model 1, the ticket cost parameter is $\beta_w = -0.0242$ (t-value $= -1.0132$); the parameter associated to the on-board travel time $\beta_T = -0.7188$ (t-value $= -1.9632$); whereas the penalty parameter is β_Δ = -0.9304 (t-value $= -1.0885$). The calibrated parameters of the model 2 present a parameter of ticket cost $\beta_w = -0.0327$ (t-value $= -1.1662$), a parameter associated to the on-board travel time: $\beta_T = -1.1391$ (t-value $= -2.8773$); whereas the intermediate stops parameter is $\beta_s = -0.7743$ (t-value $= -2.5915$). It is relevant to underline that all the calibrated parameters of the two models, which are associated to level-of-service attributes, have a correct sign and the t-value in the parameter β_T of the model 1 results to be significant, as well as the parameters β_T and β_s of the model 2.

5. Conclusions

After the diffusion of HSR lines in the world, many studies on diverted demand were developed in the attempt of capturing the diversion of passengers from the air mode and conventional rail services towards HSR. In the context of diverted demand, fare structures of HSR services play a key role in the competition among modes (inter-modal) and services (intra-modal) in the intercity context.

There are two important research lines in the field of competition among services: the development of run choice models to estimate users' behaviour in the dimension of run choice in schedule-based intercity transport services, and the analysis and estimation of dynamic fare inside the HSR and airline services.

The innovative method for collecting data through the observation of fares allows to put together the two above research lines that until now have been developed independently.

According to the above lacks in the literature, the research proposes an original method for the identification of users' behaviour in the choice dimension of run from the observation of the day-to-day evolution of ticket costs. The method is composed by two main phases. The former deals with the identification of user's choices analysing the evolution of the tickets costs. The latter phase deals with building of the choice set of alternatives and with the specification, calibration and validation of a run choice model.

The proposed method has been validated for the HSR services along the relationship Rome-Milan (Italy), by means of the specification-calibration-validation of a run choice model belonging to the class of RUMs. The aim of the validation has been to verify the possibility to ground the proposed method into a specified and calibrated choice model.

The method presents the following advantages and limitations.

The main advantages are reported below.

- The method ensures the identification of users' choices without the execution of direct surveys that are expensive in terms of time and monetary cost.
- The method supports the use of reverse assignment calibration according to assigned distributions of available seats.
- The innovative data collection used allows to obtain a huge amount of information that supports the method for the specification-calibration-validation of choice models in the dimensions of run, fare class, period of the day, origin-destination pair, and so on.

On the side of the limitations, the method does not provide indications about the number of users that choose the run r at day k. This information could be estimated on the basis of the allocation of available seats on each run r for each type of ticket. In general, it can be assumed that $P_{r,w} \leq P_r$ is the total number of seats reserved for ticket cost w on a run r, with P_r seats capacity on a run r. The hypothesis on the identification of the chosen run at day k derives from the analysis of the mutation of ticket costs between k+1 and k.

The conclusion is that the method has some advantages that could allow to obtain information about users' behaviour by means of an innovative data collection because of the huge amount of information that gives the possibility to build and update a choice model in a cheap, repeatable and easy way; and some limitations that need to be overcame in the future; however, the first results reported in this study seem encouraging.

As example, the method could be extended to other mode-services, like the airline services, where companies need to evaluate ex-ante the levels and classes of fares that will be purchased by potential demand (e.g. before introducing a service along an origin-destination pair) in order to compete with other companies (intra-modal competition) and with other modes (inter-modal competition).

CRediT authorship contribution statement

Francesco Russo: Validation, Supervision, Methodology, Conceptualization. **Giuseppe Musolino:** Methodology, Investigation, Formal analysis, Conceptualization. **Domenico Sgro:** Investigation, Formal analysis, Data curation.

Acknowledgements

This study was carried out within the MOST – Sustainable Mobility National Research Center (CUP C33C22000240001) and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) –MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033 17/06/2022, CN00000023). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Data availability

Data will be made available on request.

References

Anderson, C.K., Wilson, J.G., 2003. Wait or buy? The strategic consumer: pricing and profit implications. J. Oper. Res. Soc. 54, 299–306.

Baumeister, S., 2019. Replacing short-haul flights with land-based transportation modes to reduce greenhouse gas emissions: the case of Finland. J. Clean. Prod. 225, 262–269, 2019.

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, MA.

Ben-Akiva, M., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modelling interurban route choice behaviour. In: Proceedings of the 9th International Symposium on Transportation and Traffic Theory. VNU Science Press, pp. 299–330.

Ben-Akiva, M., Cascetta, E., Coppola, P., Papola, A., Velardi, V., 2010. High speed rail demand forecasting in a competitive market: the Italian case study. In: Proceedings of the World Conference of Transportation Research (WCTR). Lisbon, Portugal.

Ben-Akiva, M., McFadden, D., Train, K., 2019. Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis. FNT in Econometrics 10 (1–2), 1–144. https://doi.org/10.1561/0800000036.

Ben-Akiva, M., Walker, J., Bernardino, A.T., Gopinath, D.A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In: Perpetual Motion: Travel Behavior Research Opportunities and Application Challenges. Elsevier, pp. 431–470.

Bergantino, A.S., Capozza, C., 2015. One price for all? Price discrimination and market captivity: evidence from the Italian city-pair markets. Transp. Res. Part A 75, 231–244.

Campos, P., De Rus, G., Barron, J., 2009. A review of HSR experiences around the world. In: Rus, De (Ed.), The Economic Effects of high-speed Rail Investment. BBVA Foundation, Bilbao, pp. 19–32. https://econpapers.repec.org/bookchap/fbbreport/2012126.htm. (Accessed 21 February 2025).

Cascetta, E., 2013. Transportation Systems Engineering: Theory and Methods. Springer Science & Business Media.

Cascetta, E., Coppola, P., 2016. Assessment of schedule-based and frequency-based assignment models for strategic and operational planning of high-speed rail services. Transp. Res. Part A 84, 93–108 (2016).

Cascetta, E., Coppola, P., 2012. An elastic demand schedule-based multimodal assignment model for the simulation of high-speed rai(HSR) systems. Assoc. Eur. Oper. Res. Soc. 1, 3–27, 2012.

Cascetta, E., Coppola, P., 2014. High speed rail (HSR) induced demand models. Proced. Soc. Behav. Sci. 111, 147–156.

Cascetta, E., Russo, F., 1997. Calibrating aggregate travel demand models with traffic counts: estimators and statistical performance. Transportation 24 (3), 271–293.

Cascetta, E., Biggiero, L., Nuzzolo, A., Russo, F., 1996. A system of within-day dynamic demand and assignment models for scheduled intercity services. In: The Schedule-based Approach in Dynamic Transit Modelling. Springer, pp. 1–15.

Cascetta, E., Russo, F., Viola, F.A., Vitetta, A., 2002. A model of route perception in urban road networks. Transp. Res. Part B Methodol. 36 (7), 577–592.

Dalkic, G., Balaban, O., Tuydes-Yaman, H., Celikkol-Kocak, T., 2017. An assessment of the CO2 emissions reduction in high speed rail lines: two case studies from Turkey. J. Clean. Prod. 165, 746–761 (2017).

De Luca, S., 2005. Demand modelling by combining disaggregate and aggregate data: an application to a displacement scenario for the Mount Vesuvius area. WIT Trans. Built Environ. 85, 859–868.

Di Gangi, M., Russo, F., 2023. Design of hybrid rail services on conventional and high-speed lines. Int. J. Trans. Dev. Integr. 7 (2), 113–121. https://doi.org/10.18280/ijtdi.070206.

Di Gangi, M., Vitetta, 2021. A. Quantum utility and random utility model for path choice modelling: specification and aggregate calibration from traffic counts. J. Choice Model 40, 100290.

Espino, R., Martin, J.C., Roman, C., 2008. Analyzing the effect of preference heterogeneity on willingness to pay for improving service quality in an airline choice context. Transp. Res. Part E 44, 593–606 (2008).

Givoni, M., Dobruszkes, F., 2013. A review of Ex-Post evidence for mode substitution and induced demand following the introduction of high-speed rail. Transp. Rev. 33, 720–742. https://doi.org/10.1080/01441647.2013.853707.

Goodman, L.A., 1961. Snowball sampling. Ann. Math. Stat. 32 (1), 148–170.

IlSole24Ore webpage. https://lab24.ilsole24ore.com/cop27-dati-CO2-mondo/. (Accessed 26 February 2024).

Jung, S.Y., Yoo, K.E., 2014. Passenger airline choice behavior for domestic short-haul travel in South Korea. J. Air Transport. Manag. 38, 43–47 (2014).

Malighetti, P., Paleari, S., Redondi, R., 2015. EasyJet pricing strategy: determinants and developments. Transportmetrica: Transp. Sci. 8 (11), 686–701.

Manski, C.F., 1977. The structure of random utility models. Theor. Decis. 8, 229–254.

McAfee, R.P., Te Velde, V., 2006. Dynamic pricing in the airline industry. Handb. Econ. Inform. Syst. 1, 527–567.

McFadden, D., 2001. Disaggregate Behavioral Travel Demand's RUM Side - a 30-Year Retrospective.

Morikawa, T., 1996. A hybrid probabilistic choice set model with compensatory and non-compensatory rules. In: Proceedings of the 7th WCTR. Sydney, Australia.

NTV, 2013. Contratto Di Trasporto. Available online at: https://www.italotreno.it/it. (Accessed 29 January 2024).

Nuzzolo, A., Crisalli, U., Gangemi, F., 2000. A behavioural choice model for the evaluation of railway supply and pricing policies. Transp. Res. Part A 34, 395–404 (2000).

Nuzzolo, A., Russo, F., Crisalli, U., 2002. Transit Network Modelling. the Schedule-based Dynamic Approach. FrancoAngeli, Milano.

Nuzzolo, A., Russo, F., 1998. Departure time and path choice models for intercity transit assignment. In: Travel Behaviour Research: Updating the State of Play, Juan De Dios Ortuzar, David Hensher, Sergio Jara Diaz. Elsevier, Amsterdam isbn 0 08 043360X.

Nuzzolo, A., Crisalli, U., Rosati, L., Ibeas, A., 2013. STOP: a short term transit occupancy prediction tool for APTIS and real-time transit management systems. Proceedings of the 2013 IEEE 16th International Conference on Intelligent Transportation Systems (ITSC), pp. 1372–1377.

Ortúzar, J.D.D., Willumsen, L.G., 2001. Modelling Transport, third ed. John Wiley, Chichester, NY, USA. ISBN 978-0-471-86110-2.

Quattrone, A., Vitetta, A., 2011. Random and fuzzy utility models for road route choice. Transp. Res. Part E Logist. Transp. Rev. 47, 1126–1139.

Russo, F., 2021. Which high-speed rail? LARG approach between plan and design. Fut. Trans. 1, 202–226.

Russo, F., Vitetta, A., 1995. Networks and assignment models for the Italian national transportation system. In: Proceedings of the 7th WCTR. Sydney, Australia.

Russo, F., 1997. Fuzzy theory in transportation field: fuzzy sets for simulating path choice behaviour. In: Advances in Intelligent Systems, pp. 279–283.

Russo, F., Chilà, G., 2014. Integrated travel demand models for evacuations: a bridge between social science and engineering. Transport. Res. Pol. Pract. 67, 69–84.

Russo, F., Vitetta, A., 2011. Reverse assignment: calibrating link cost functions and updating demand from traffic counts and time measurements. Inverse Problems Sci. Eng. 19 (7), 921–950.

Russo, F., Sgro, D., Musolino, G., 2023. Sustainable development of railway corridors: methods and models for high speed rail (HSR) demand analysis. In: Gervasi, O., et al. (Eds.), Computational Science and its Applications – ICCSA 2023 Workshops. ICCSA 2023. Lecture Notes in Computer Science, 14110. Springer, Cham. https://doi.org/10.1007/978-3-031-37123-3_36.

Russo, F., Sgro, D., Musolino, G., 2024. Dynamic structure of fares for high speed rail services. In: Gervasi, O., Murgante, B., Garau, C., Taniar, D., Rocha, A.M.A.C., Faginas Lago, M.N. (Eds.), Computational Science and its Applications – ICCSA 2024 Workshops. ICCSA 2024. Lecture Notes in Computer Science, 14822. Springer, Cham. https://doi.org/10.1007/978-3-031-65318-6_13.

Russo, F., Vitetta, A., 2003. An assignment model with modified logit, which obviates enumeration and overlapping problems. Transportation 30 (2), 177–201.

Siripirote, T., Sumale, A., Watling, D.P., Shao, H., 2013. Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning. J. Pub. Transport. 16 (3), 47–66.

Talluri, K.T., Van Ryzin, G.J., 2006. The Theory and Practice of Revenue Management, 68. Springer Science & Business Media.

Train, K., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, Cambridge; NY, USA. ISBN 978-0-521-76655-5.

Trenitalia, 2024. Condizioni generali di trasporto. Available online at: https://www.trenitalia.com/it/informazioni/condizioni_generaliditrasporto.html. (Accessed 29 January 2024).

Tuchschmid, M., 2010. Eco calcolatore FFS. Centro svizzero per gli inventari ecologici c/o EMPA. Available online: http://docplayer.it/amp/15743061-Ecocalcolatore-ffs-relazione-di-base-versione-1-0-settembre-2010.html. (Accessed 24 July 2023).

UIC High Speed advisor. Leboeuf M. (2018). High Speed Rail fast track to sustainable mobility. UIC Passenger Department. 1-76.

UN-United Nations, 2018. Global indicator framework for the sustainable development goals and targets of the2030 agenda for sustainable development. Available online: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202023%20refinement_Eng.pdf. (Accessed 24 July 2023).

UN-United Nations, 2015. Transforming our world: the 2030 agenda for sustainable development. Available online: https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf. (Accessed 24 July 2023).

Vitetta, A., 2025. Path choice in transport systems: comparing random, quantum, and fuzzy utility models in a small network. Front. Fut. Trans. 6, 1544947. https://doi.org/10.3389/ffutr.2025.1544947.

Vitetta, A., 2016. Quantum utility model for route choice in transport systems. Travel Behav. Soc. 3, 29–37.

Xiao, Y.B., Chen, J., Liu, X.L., 2008. Joint dynamic pricing for two parallel flights based on passenger choice behavior. Syst. Eng. Theory Prac. 28 (1), 46–55.

Xiao, Y.-B., Chen, J., Liu, X.-L., 2008. Joint dynamic pricing for two parallel flights based on passenger choice behavior. Xitong Gongcheng Lilun yu Shijian/Syst. Eng. Theory Pract. 28 (1), 46–55.

Yao, E., Morikawa, T., 2005. A study of an integrated intercity travel demand model. Transp. Res. Part A 39, 367–381.

Yao, E., Yang, Q., Zhang, Y., Sun, X., 2013. A study on high-speed rail pricing strategy in the context of modes competition. Discrete Dynam Nat. Soc., 715256