DOCTORAL SCHOOL
Università Mediterranea di Reggio Calabria

DEPARTMENT OF INFORMATION ENGINEERING, INFRASTRUCTURE
AND SUSTAINALBLE ENERGY
(DIIES)

PHD IN
INFORMATION ENGINEERING

S.S.D. CEAR/03
XXXVIII CYCLE

Modeling run choices and fare structures in High-Speed Rail by means of innovative data collection

Candidate
DOMENICO SGRO
[redacted]

Tutor
prof. GIUSEPPE MUSOLINO
[redacted]

Coordinator
prof. TOMMASO ISERNIA
[redacted]

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

MOST
CENTRO NAZIONALE PER LA MOBILITÀ SOSTENIBILE

Università degli Studi Mediterranea di Reggio Calabria
Via dell'Università, 25 Reggio Calabria (Italy)

DOMENICO SGRO

Modeling run choices and fare structures in High-Speed Rail by means of innovative data collection

Domenico Sgro is a PhD candidate in Information Engineering at the Università degli Studi Mediterranea di Reggio Calabria. His research focuses on demand and supply in high-speed rail transportation, with particular emphasis on run choice behavior and dynamic fare structures based on innovative data collection methods. He has served as presenting author at international scientific conferences, and he is the author of several international scientific publications.

This work was supported by the European Union - Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.4, CUP C33C22000240001, National Center for Sustainable Mobility (CN00000023 - program 'MOST').

The research activity carried out by Università Mediterranea degli Studi di Reggio Calabria focused on the following areas.

- "Comparison under sustainability criteria of the existing HSRs in Italy with EU experiences, identification of solutions for Southern Italy and LARG paradigm". The research activity focused on the definition of Lean, Agile, Resilient and Green (LARG) paradigm to support the development of programming of HSR lines by considering Agenda 2030.
- "Framework for analysis of High-Speed Rail (HSR) demand to pursue sustainable development in Southern Italy, consistently with national transport planning, the EU TEN-T/RFC corridors and core port nodes". The research activity focused on the analysis of passenger and freight rail transport supply through the development of centrality and alignment indicators to assess the structural and functional role of nodes and links in the existing and future railway network connecting Italian cities.
- "Framework for analysis of capacity of EU TEN-T5/RFC corridors, in the context of EU legislative proposals of the revision of the TEN-T/RFC, considering core port nodes". The research activity focused on the analysis of intercity passenger mobility by developing diverted and induced travel demand models between Rome and Reggio Calabria (Italy), as case study of Southern Italy. The potential modal shifts and new generated trips in response to the introduction of HSR mode-service have been estimated.
- "Methods and models for the analysis of sustainability of alternative solutions of maintenance and materials. Development of case-study for Southern Italy railway network" and "Methods and models for the assessment of suitable track solutions based on internal and external constraint: Development of case-studies in Southern Italy railway network". The research activity focused on the analysis based on drone surveys, aimed at monitoring track conditions for preventive maintenance, combined with environmental sustainability assessments using the Life Cycle Assessment (LCA) methodology to evaluate the environmental impact of the track and its components—including the use of sustainable recycled materials —over their entire life cycle.

The references of the main publications regarding the above research areas are reported in paragraph 7.3.

"Commit your works to the Lord, and your plans will succeed"
(Proverbs 16:3)

Acknowledgements

First and foremost, I wish to express my sincere gratitude to my tutor, Professor Giuseppe Musolino, whose support, guidance, and mentorship have been invaluable throughout my PhD. From my earliest days as a master's student, he has been a constant source of inspiration and knowledge, inspiring in me a deep passion for the transportation sector, which has shaped the course of my academic path thus far. I am profoundly grateful for the opportunities for growth and the experiences that have enriched my learning throughout this journey.

My heartfelt appreciation also extends to the entire research team of the LAST Laboratory: Prof. Francesco Russo, whose wisdom and guidance have been a steady compass; Prof. Corrado Rindone, for his precious advice and for all that he has generously shared with me; Prof. Antonino Vitetta, an inexhaustible source of knowledge from whom I have learned extensively; Prof. Paola Panuccio, for her valuable contributions and her constant kindness and support; and my colleague Marialuisa Moschella, with whom I shared both the achievements and the challenges that marked these years. Working alongside them has been an honor and a privilege, their collaboration has deeply enriched my doctoral experience.

I am deeply grateful to my family. To my grandmother, Pina, for her enduring presence and love throughout my life; to my mother, Palmina, for her sacrifices and her unwavering faith in my potential. My sincere thanks also go to my friends, whose support and encouragement have accompanied me at every step.

With deep gratitude to Alessandra, my beloved, for her immense love, patience, enthusiasm, and steadfast encouragement, especially during the most challenging moments of this PhD.

With heartfelt appreciation to my mentor, Dr. Tony Moscato, whose guidance and wisdom inspire me every day.

Last but not least, my deepest gratitude belongs to my Lord and Savior, Jesus Christ, who is my everything and the source of my life.

I encourage every reader of this thesis who is embarking on this path, or on any journey in life, to pursue excellence in all they do.

To my dad, Giovanni

Summary

Abstract

There are more than 110,000 kilometers of High-Speed Rail (HSR) lines in operation, in construction and planned in the world. The entering of HSR mode-services generate competition with other rail services and other transport modes. A relevant problem is the estimation of the travel demand and of the supply of fares connected to HSR.
HSR travel demand models at the levels of mode, service, company, run are mainly focused on the estimation of the diverted demand from the air mode and conventional rail services. HSR lines compete at the inter-modal level (e.g., mode choice) and at the intra-modal levels (e.g., service, company and run choices).
HSR supply of fares, similarly to the air services, are characterized by evolutions and changes over time. The fares variations are caused by several factors connected both to travelers' behavior and to strategies of the HSR companies.
The two above themes are studied separately in the scientific literature. The thesis proposes an innovative method for collecting data relating to the ticket's evolution allowing the two lines of research to be combined and providing a research contribution in the development of travel demand and transport supply models. The thesis aim is to put together the two above research lines that have been historically developed independently.
The innovative method for collecting data allows, on the demand side, the identification of user's choices and the specification, calibration and validation of run choice models; and, on the supply side, the specification, calibration and validation of fare functions for the representation of the dynamic variation over time of HSR fares.
The research activity is composed by three main phases. The first phase concerns the analysis of the state of the art of the two above lines of research aimed at the individuation of lacks of literature. The second phase regards the definition of a framework for the identification of user's choices and the specification, calibration and validation of run choice models, as well as for the estimation of fare structures, which are based on the innovative data collection method. The method allowed the identification of users' choice in terms of run on the basis of the fares mutation between consecutive days The third phase concerns the experimentation of the proposed method and it is subdivided into two parts. The former part concerns the design and execution of the fares' survey in order to collect data and build the database. The survey has been executed on a set of HSR lines operating along the relationship Rome-Milan (Italy) during several years. The latter part concerns the development of the demand and supply models in terms of specification, calibration and validation of disaggregated run choice models belonging to the class of random utility models, and of aggregated fare structure models.
The obtained results are important because give the possibility to calibrate and update the demand and supply model parameters from big data, such as the daily tickets supplied by HSR companies. The method ensures the identification of users' choices without the execution of traditional surveys that are expensive in terms of time and monetary cost and allows to obtain a huge amount of information for the specification-calibration-validation of choice models in the dimensions of run. The results of this work could support transport planners and decision-makers to develop sustainable transport policies in the evaluation of investment in HSR lines and services, by means of methodological and modelling tools to assess current and potential HSR travel demand.

Introduction

High Speed Rail (HSR) is a transport mode-service that allows to achieve sustainable mobility goals due to the environmental, economic and social benefits (United Nations, 1987; United Nations, 2015; United Nations, 2018). Among the seventeen Sustainable Development Goals (SDGs) defined, HSR generates direct and indirect impacts on the following ones.

On one hand, HSR has direct impacts on the Goals 9 and 11.

Goal 9 concerns "Industry, Innovation and Infrastructure", hence building and improving of innovative, sustainable and resilient infrastructures and industrialization; by means of investments in green energy, information and communication technologies.

Goal 11 regards "Sustainable Cities and Communities", thus it generates inclusivity, safety, resiliency and sustainability of the cities; for the reason that half of the worldwide population is living in cities, and it is expected to grow in the future. The latter considerations cause the necessity of improving green spaces and urban planning in order to be inclusive and accessible.

On the other hand, HSR has indirect impacts on Goals 7, 8, 10 and 13.

Goal 7 "Affordable and Clean Energy" is oriented to guarantee sustainable, affordable, reliable and modern energy, because of the huge amount of energy consumption results to be relevant the impact on climate change in terms of greenhouse gas emission.

Goal 8 "Decent work and Economic Growth" is aimed at promoting an inclusive and sustainable economic growth; moreover, it is focused on full and productive employment, decent work for all, and improved labor rights as well as working conditions worldwide.

Goal 10 "Reduce Inequalities" is aimed at the reduction of the inequalities within and among countries. It is focused on promoting social, economic, and political inclusion for everyone, regardless of age, gender, disability, ethnicity, or income.

Goal 13 "Climate Action" is a call for urgent action to cope with climate changes and its impacts. It pursues on the strengthening of resilience, improving climate awareness, as well as integrating climate measures into national policies.

HSR has experienced rapid growth both in supply, in terms of services and infrastructures, and demand sides, in terms of passengers transported. Concerning the supply side (Fig. I.1), there are 64,698 kilometers of HSR lines in operation, the 19,927 kilometers under construction and 53,201 kilometers planned (UIC, 2024). Specifically, China results to have the largest HSR network with about 45,390 km, whereas the whole Europe continent results to have 12,454 km of lines, the Middle East instead 1,681 km, North America with 735 km of lines and Africa content with 186 km.

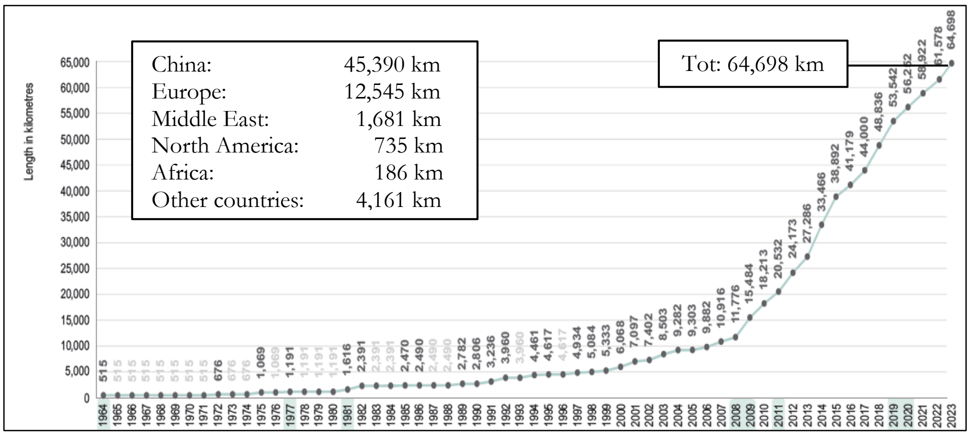


Fig. I.1. Worldwide growth of HSR lines (Source: UIC, 2024)

Regarding the demand side there are almost three billion and half of passengers that have travelled worldwide in 2022 (UIC, 2025). Fig. I.2 shows the evolution of the annual traffic trend in the countries where HSR has a relevant presence. Regarding year 2022 the highest values of traffic flows, with almost 450,000 billion of passenger-km, is stated for China country; the other countries as a whole present values of traffic flows of about 240,000 billion of passenger-km. The global value of annual average growth in 2022 is 3.19; with a value of 2.60 in China.

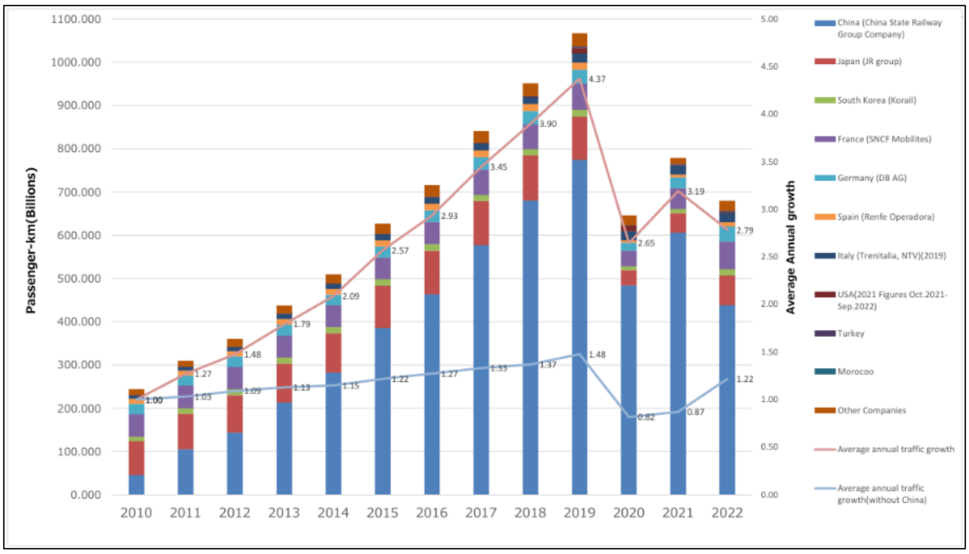


Fig. I.2. Demand trend among countries and years (Source: UIC, 2025)

The growth of HSR infrastructures and services caused competition at inter-modal level (e.g., mode choice) and at intra-modal level (e.g., service, company and run choices), hence on different travelers' choice levels in the context of intercity trips (Nuzzolo and Russo, 1998; Cascetta and Coppola, 2012; Givoni and Dobruszkes, 2013; Russo et al., 2023). Consequently, it is important to investigate the phenomenon of travelers' mobility in order to better understand the users' behavior in relation to the supply of HSR services.

As far as concerns the intra-modal competition, two relevant lines of research are present in the current scientific literature: the run choice modelling and the analysis of the structure of dynamic fares.

Leaving from the fundamental work of McFadden (1972, 1974), the run choice modelling in scientific literature relies on the Manski (1977) assumption, according to which two steps may be identified: the choice set generation and the run choice.

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 after the work of Mansky are the following: Ben-Akiva & Lerman (1985), Cascetta et al. (1995, 2002a), Ben-Akiva et al. (2002), Yao & Morikawa (2005). The choice model generally relies on random utility theory, whose main assumption is that a user chooses a run among the set of perceived alternatives, maximizing their associated perceived utility.

Two main approaches may be used in run choice modelling (Nuzzolo and Russo, 1993, 1997, 2003; Nuzzolo et al., 2000; Russo & Vitetta, 2003; Ottomanelli et al., 2007; Cascetta, 2013): 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 structures of fares in literature are basically clustered into static and dynamic. On one hand, the static structure typically concerns the local transit services or those not on the market; on the other hand, the dynamic structure of fares regards the conventional and HSR services and airline services (Zheng & Liu, 2016; Russo et al., 2024). The dynamic structure of fares is characterized by variations of the ticket value over the period of availability, thus during the days on which it is possible to purchase a ticket. The variations of the tickets rely on variables connected to commercial purposes, such as pricing policy and strategies of the transport company; as well as to travelers such as choice behavior in terms of modes, services, run/flight, ticket categories. Hence, transport companies define the level of fares mainly in order to maximize their revenues without demand variation.

Despite travelers may be time-based, who choose the trip in relation to level-of-service characteristics of the trip (e.g. travel time, arrival time at destination, departure time from the origin) and price-based, who generally choose the fare that minimize the monetary cost of their trip (Russo et al., 2024; Russo et al., 2025).

Both run choice and fares models are commonly calibrated by means of data collected by traditional surveys of passenger mobility, operated through direct interviews to a sample of travelers in case of run choice models; and by way of direct information about a sample of fares supplied by companies in case of fares models.

The traditional surveys used to obtain data for models' calibration have been recently accompanied by new forms of surveys that use emerging Information and Communication Technologies (e-ICT) applied to transport. Some examples are presented in city logistics (Russo & Comi, 2022; Campisi et al., 2023; Cirianni et al., 2023;), in mobility-as-a-service (Musolino, 2022; Rindone, 2022; Russo, 2022; Vitetta, 2022) and in ports and maritime transport (Carlan et al., 2016; Musolino et al., 2022) applications.

The e-ICT tools are commonly incorporated into stand-alone digital platforms of transport operators and companies (e.g. freight couriers, transit companies, port terminals). However, their transversal and shared use among the different transport actors is not common. The existing e-ICT tools for mobility applications may be classified into five categories: Internet of things (see, among the others, Atzori et al., 2017); big data (Chen et al., 2016; Anda et al., 2017); digital twin (Fuller et al., 2020; Jones et al., 2020); artificial intelligence (Cirianni et al., 2023; Russo et al., 2025); blockchain (see, among the others, Astarita et al., 2025).

The literature shows that the two research lines, considered in this thesis, until now have been developed independently. There are few papers in the field of run choice models to estimate users' behavior in the choice of run in the schedule-based intercity transport services (air and HSR), whereas more extended research is present in the field of the analysis and estimation of fare evolution in the HSR system.

According to the above lacks in the literature the thesis' aim is to respond to two main Research Questions (RQs).

RQ1. Is it possible to obtain reliable information about travelers' behavior (e.g., run choice) from big data related to the evolution of dynamic fares?

RQ2. Does the information extracted from big data allow to unify the two lines of research concerning the run choice models and the dynamic fare functions into the single research line commonly called in the literature "reverse assignment"?

In order to respond to the two above RQs the thesis proposes an innovative method for data collection; the latter is based on the acquisition of data from big storage of big repository, containing a great volume and variety of historical information about the characteristics of the transport supply (fares). The research contribution concerns the definition of a framework for the identification of users' choice of HSR runs from big data stored, structured and selected. The proposed method has been validated by means of the specification-calibration-validation of a choice model in the run dimension, on the demand side; and the specification, calibration and validation of fare functions for the representation of the dynamic variation over time of HSR fares, on the supply side.

According to two RQs specified above, the thesis is organized into several chapters.

Chapter 1 presents and classifies the existing literature on the two above research lines centered on HSR. The same provides the necessary elements for the identification of lacks in literature, therefore with the aim of providing a contribution to the advancement of literature.

The chapter is organized into four paragraphs. Paragraph 1.1 regards the HSR definitions from several points of view, such as technical definitions from infrastructural, operational and economic points of view. Paragraph 1.2 concerns the run choice models, which is organized into sub-paragraphs containing the HSR demand components and models, the specification of models related to the different choice dimensions, hence the classification of the literature in relation to different criteria and finally are described the methods of data collection. Paragraph 1.3 contains state-of-the-art regards the fare structures, in particular with some definitions and the description of the fare models. Paragraph 1.4 regards the description of the existing literature gaps.

Chapter 2 presents general definitions and notations of variables used in data-bases and models. Specifically, the main reference variables are defined: the day of trip and the day of ticket purchasing; thus the statement of the problem has been schematized. Therefore the variables relating respectively to the run choice model and to the model of fare structures are defined.

Chapter 3 gives a description of the fare survey carried out. The survey was designed and conducted over three different years: 2022, 2023 e 2025, in order to extract the information related to the fare supplied of HSR runs, belonging to different companies. Therefore, the data-base was designed in order to store the observed data.

Chapter 4 presents the steps regarding the proposed framework, which is composed of three main parts, it allows the identification of the chosen run by the user. Therefore, the run choice model, the likelihood and the estimator are specified. It is important to underline that the calibrations are carried out on the basis of the extracted samples described in chapter 3. Moreover, several hypotheses concerning the choice set building are specified and used for the calibration of the parameters. The chapter is organized into three paragraphs: identification of the hypothesized chosen run (par. 4.1), specification of run choice model (par. 4.2), calibration of run choice models (par. 4.3).

The chapter presents the specification, calibration and validation of the framework aimed at the fare modelling building process. Are analyzed the fare structures, thus the evolution of the fares, by means of considering a single run and set of runs, clustered by time slots.

Chapter 5 concerns the fare supply modelling and it is organized into four paragraphs. Paragraph 5.1 contains the specification of the applied framework for the fares analysis. Paragraph 5.2 contains the examples of the fares observations, by means of single runs and sets of runs, analyzed as k varies and t fixed. Paragraph 5.3 regards the examples of the fares observations, by means of single runs and sets of runs, analyzed as t varies and k fixed. Paragraph 5.4 regards the examples of the fares observations, by means of single runs and sets of runs, analyzed as k varies and t fixed. Paragraph 5.5 regards the pilot analysis on a single run, analyzed as t varies and k fixed and as k varies and t fixed.

Paragraph the calibrated parameters of the specified fare models, as k varies and k fixed, and regarding single runs and set of runs. The chapter contains Appendix B, where there are some additional model specifications.

Chapter 6 regards the conclusions, providing the main achievements, the limitations of the research and the future research developments.

1.
State of the art

The chapter concerns the state of the art and contains information about the work carried out on the existing literature on high-speed rail. The same provides the necessary elements for the identification of lacks in literature, therefore with the aim of providing a contribution to the advancement of literature and therefore covering the identified lacks.

The chapter is organized into four paragraphs. Paragraph 1.1 regards the HSR definitions from several points of view, such as technical definitions from infrastructural, operational and economic points of view. Paragraph 1.2 concerns the run choice models, which are organized into sub-paragraphs containing the HSR demand components and models, the specification of models related to the different choice dimensions, hence the classification in relation to the typology of model and finally are described the methods of data collection. Paragraph 1.3 contains the state-of-the-art and regards the fare structure, in particular with some definitions and the description of the fare models. Paragraph 1.4 regards the description of the literature gaps individuated among the three paragraphs which concern the sectors of literature research.

1.1. HSR definitions

The paragraph reports the main definitions of HSR existing in the technical literature, which may be classified as definitions based on the characteristics of the railway infrastructure, such as the maximum speed; and definitions based on the characteristics of the railway services, such as the use of the railway infrastructure by the different services, as well as from the transport demand side.

On one hand, the HSR from the infrastructure point of view results to be defined by different documents recalled in the following.

EU introduced in 1996 (EU, 1996) some definitions of HSR associated to the maximum speed that trains could reach while travelling on a specific track segment. According to EU, the term "High Speed Rail" indicates a system "composed of the railway infrastructures comprising lines and fixed installations, of the trans-European transport network, constructed or upgraded to be travelled on at high speeds, and rolling stock designed for travelling on those infrastructures.". The EU Directive 48/1996 established that high speed infrastructure comprises three types of lines:

- "specially built high-speed lines equipped for speeds generally equal to or greater than 250 km/h";
- "specially upgraded high-speed lines equipped for speeds of the order of 200 km/h";
- "specially upgraded high-speed lines which have special features as a result of topographical, relief or town-planning constraints, on which the speed must be adapted to each case.".

European Rail Infrastructure Managers in 2008 (EIM, 2008) classified railway lines according to the different demand segments, and infrastructure performances (axle load). Among the identified classes, the characteristics of the following railway networks are reported below (see Table 1.1):

- High-Speed (HS) passenger network with speeds higher than 250 km/h;
- Conventional High-Speed (CHS) passenger network with speed up to 250 km/h;
- Heavy Freight (HF) network with 100 km/h and up to 35 tons for axle load;
- High-Speed/Logistical Freight (HS/LF) network with speeds up to 250 km/h.

A strategic vision for the railway networks in the year 2035 has been also proposed; in particular, a maximum speed of 360 km/h was foreseen for the HS network, taking into account the business needs of potential travelers (Tab. 1.1).

Tab.1.1. Vision of railway networks at year 2035

	Railway network	2006/2007 Speed (km/h)	Axle load (tons)	2035 Speed (km/h)	Axle load (tons)
Pax	H.S.(*)	300	18	360	18
	C.H.S.(**)	200	20	250	18
Freight	H.F.(-)	100	25	100	Up to 35
	H.S./H.F.(°)	160	20	Up to 250	18

(*)High Speed; (**)Conventional High Speed; (-)HF: Heavy Freight; (°)HS-HF: High Speed-Heavy Freight. (Source: EIM, 2008)

EU directive of European Parliament and of the Council of 11 May 2016 (EU, 2016) integrated the criteria for identifying HSR lines introduced in EU (1996):

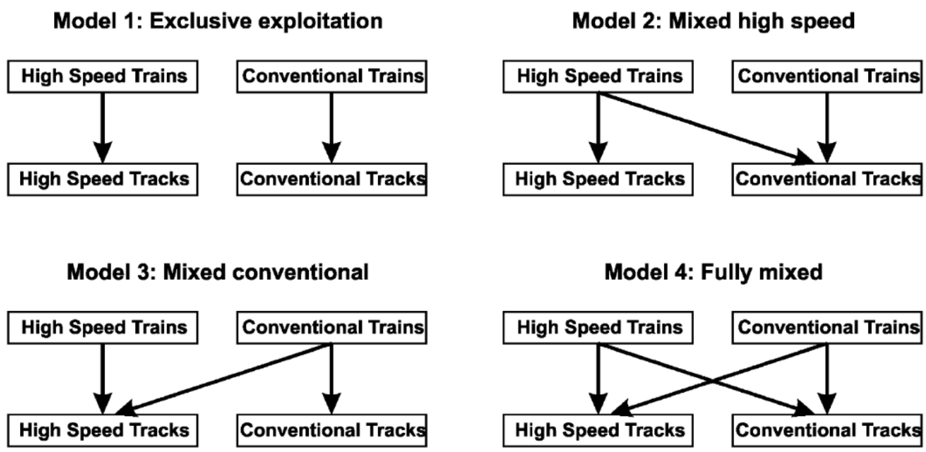
- specially built high-speed lines equipped for speeds generally equal to or greater than 250 km/h;
- specially upgraded high-speed lines equipped for speeds of the order of 200 km/h;
- specially upgraded high-speed lines which have special features as a result of topographical, relief or town planning constraints, to which the speed must be adapted in each case. This category includes interconnecting lines between high-speed and conventional networks, lines through stations, accesses to terminals, depots, etc. travelled at conventional speed by 'high-speed' rolling stock.

Finally, UIC released a directive in 2018 (UIC, 2018), consistent with the EU standards, where "High speed rail (HSR) encompasses a complex reality involving many technical aspects, such as infrastructure, rolling stock and operations, as well as strategic and cross-sector issues including human, financial, commercial and managerial factors". Moreover, UIC states that "HSR is still a grounded, guided and low grip transport system: it could be considered to be a railway subsystem". The most important change comes from the speed: "HSR means a jump in commercial speed and this is why UIC considers a commercial speed of 250 km/h to be the principal criterion for the definition of HSR.". A secondary criterion is admitted on "average distances without air competition, where it may not be relevant to run at 250 km/h, since a lower speed of 230 or 220 km/h or at least above 200 km/h is enough to catch as many market shares as a collective mode of transport can do."

As far as concerns railway services, Campos & De Rus (2009) introduced an economic definition of the HSR service, which is defined as "the relationship of HSR with existing conventional services, and the way in which the use of infrastructure is used". They identified four exploitation models:

1. "Exclusive exploitation" model, when a complete separation exists between high speed and conventional services, each one with its own infrastructure (model adopted in Shinkansen technology in Japan);
2. "Mixed high-speed" model, when high speed trains run either on specifically built new lines or un upgraded segments of conventional lines (model adopted in TGV technology in France);
3. "Mixed conventional" model, where some conventional trains run on high speed lines (model adopted in AVE technology in Spain);
4. "Fully mixed" model, where both high speed and conventional services can run (at their corresponding speeds) on each type of infrastructure.

Fig. 1.1. Economic definition of the HSR service (source: Campos & De Rus, 2009)



According to Campos & De Rus (2009) the Italian railway network is categorized as "Fully Mixed", just for the Rome-Florence line. Particular attention has been given to the study of the Southern Italy railway system, with a focus on the experimental studies about the introduction of services using the conventional and HSR networks. A model has recently been proposed in order to analyze the services, defined as "hybrid", which operate on two different networks (Di Gangi & Russo, 2022), proposing the formalization of the optimal timetable project model with unchanged infrastructural resources (Ben-Akiva & Lerman, 1985).

The European Parliament and the Council of the European Union regulation, adopted on 13 June 2024 (EU & Council of the European Union, 2024), is an important update of the guidelines for the development of the Trans-European Transport Network (TEN-T). The objective is to enable a high-quality, interconnected, efficient, with multimodal transport infrastructure across the EU Countries, encouraging the internal market, territorial cohesion, and sustainable mobility.

It is underlined the dual-layer (or three-layer) structure of TEN-T, thus the core network, the extended core network, and the comprehensive network. The core network concerns the most strategic corridors and nodes, is expected to be concluded within 2030. The extended core network, introduced sections which results essential, however more complex or toughest to be implemented, it is estimated a deadline for the yar 2040. Therefore, the comprehensive network, which would cover the entire EU territory and ensuring territorial cohesion, is expected to be completed by 2050. This approach allows the EU to focus investments where it is generated the highest impact in a short term, while keeping a long-term vision aimed at the fully integrated, multimodal and interoperable transport system.

The regulation highlights the crucial role of intermodality (rail, road, inland waterways, maritime and air transport), interoperability and infrastructure upgrading. The annexes additionally contain maps and the lists of nodes (e.g., urban nodes, ports, airports, inland terminals) that have to be integrated with TEN-T. Moreover, the regulation individuates corridors across the network to support long-distance freight and passenger flows, including cross-border connections.

In addition, the guideline strengthens the connections between infrastructure development and strategic goals such as climate action, digitalization, and military mobility (in accordance with TEN-T framework), hence give a multi-dimension perspective to the transport policy study.

The European Commission in November 2025 (EU, 2025) declared the objective of improving the economy, the quality of jobs, the cohesion, to connect cities with the important milestone of climate goal, thus decarbonization and pollution decrease. It results the necessity to plan the implementation and harmonization of the High Speed Rail (HSR) network in the EU, with the elimination of the TEN-T infrastructural bottlenecks. A set of indicators are introduced to monitor the evolution of the entire project (e.g., HSR lines extension, average speed, flows, …), as well as setting up an annual survey to monitor the progress.

Another objective specified in EU (2025) in medium-long term (year 2040) concerns the development of European HSR network allowing the fast connection between the principal urban areas, with speeds above 200 km/h in order to reduce the travel time (Fig. 1.2). HSR lines have grown only by 17% compared to 2015.

The EU Commission promotes actions aimed at sustainable HSR connections by suggesting the Member States to plan the updating or new construction of HSR links with speeds above 250 km/h (coherently to the minimum TEN-T requirements). This project (EU, 2025) includes a financial strategy to take into account and coordinate all the aspects of a new HSR relationship. Moreover, another crucial aspect concerns the cross-border HSR networks, it is important the cooperation of rail track owners and High Speed Train (HST) operators in order to provide a service with matching standards.

The expected results regard the HSR context, thus the conventional network in terms of capacity upgrading (passenger, freight, military purposes). On the other hand, regard the attended development of the European Rail Traffic Management System (ERTMS), for a higher level of interoperability and safety.

Fig. 1.2. HSR fast connections between major cities (source: EU, 2025)

1.2. State of the art on run choice models

Within the High Speed Rail (HSR) context, defined in paragraph 1.1, the paragraph contains the analysis of the state of the art related to run choice models. The paragraph is subdivided into four subparagraphs. Paragraph 1.2.1 concerns HSR demand components and models. Paragraph 1.2.2 contains the specification of models relating to the different choice dimensions. Paragraph 1.2.3 concerns the literature review of run choice models. Lastly, paragraph 1.2.4 concerns the description of data collection methods.

1.2.1. HSR travel demand components and models

Travel demand models have their theoretical background in Transport Systems Models (TSMs) framework (McFadden, 1972, 1974, 1981; Ben-Akiva & Lerman, 1985; Cascetta et al., 1995; Ortúzar & Willumsen, 2001; Cascetta, 2009). TSMs simulate a transport system through a process, in which transport supply and travel demand interact generating the flows and the performances of the transport system. TSMs are composed of three main elements. The transport supply model simulates the utilities of users deriving from the use of transport infrastructures and services. The travel demand model simulates user choices based on the performance of transport infrastructures and services.

The supply-demand interaction model simulates the interaction between the user's choices and the performance of the infrastructures and the services, giving flow for each link.

Travel demand models may be broadly classified into two main approaches: aggregate models and disaggregated models.

Aggregated models may be segmented into three categories:

- statistical-descriptive models which estimate the levels of demand throughout relationships with attributes belonging to the level-of-service and socio-economic class;
- time series models which use historical data to forecast demand flows with given specifics (e.g. origin-destination relationship);
- partial-share models, which simulates the user choice method through a procedure of partial sequential choices, or steps; the most common case is constituted by multi-stage models, including trip generation, trip distribution, time choice (arrival/departure), service choice, route/run choice.

Disaggregated models are theoretically and operationally more complex in relation to the difficulty of finding data on user choice behavior. They can be based on the theory of discrete choice model (Manski & McFadden, 1981; Ben-Akiva & Lerman, 1985; Train, 2009; Ben-Akiva, McFadden, & Train, 2019). The discrete choice model has been specified with different formalizations of alternatives and linked perceived utility random utility (Manski & McFadden, 1981; Ben-Akiva & Lerman, 1985; Train, 2009; Ben Akiva, McFadden & Train, 2019), fuzzy utility (Russo, 1997; Quattrone & Vitetta, 2011), or quantum utility (Vitetta, 2016; Di Gangi & Vitetta, 2021).

Generally, the random utility models differ according to the perceived utility function, which can be specified by considering different functional relationships between the levels of choice (hierarchical or factorial), different hypotheses on structure of choice set of alternatives, and different hypotheses about the distribution of random residuals. The different hypotheses about the distribution of the random residuals lead to two main categories of models:

- models with choice probabilities expressed in closed form (e.g. multimodal logit, nested logit);

- models with simulated choice probabilities (e.g. probit, mixed logit).

As far as concerns the partial share models, the models associated to the main trip choice dimensions of travelers are reported in the following, according to Ben-Akiva & Lerman (1985), Ortúzar & Willumsen (2001), Cascetta (2009).

The travel demand generated by HSR mode-services is sub-divided in the scientific literature into three components (Ben-Akiva et al., 2010; Russo & Musolino, 2012; Cascetta & Coppola, 2014; UIC, 2018; Russo et al., 2023):

- Diverted demand:
  - from other modes (e.g., car, air, and bus) to HSR, and
  - from other rail services to HSR;
- Induced demand:
  - changes in trip frequency;
  - changes in trip destination;
- Economy-based demand:
  - changes in production-consumption generation;
  - changes in production-consumption location.

Diverted demand is a shift of demand, or diversion, towards HSR services. The diversion may occur either from other modes, for instance in the case of a shift from airplane/car to HSR services, or from other rail services, for instance in the case of shift from intercity services to HSR services (Dobruszkes, 2011; Givoni & Dobruszkes, 2013; D'Alfonso et al., 2016a). The diverted demand could depend on several attributes, which are mainly endogenous to the transport system, involving the level of service of the transport supply (i.e. on-board travel time, access-egress time, fare).

Existing studies show that entity of diverted demand towards HSR from other modes could be different in relation to HSR in-vehicle travel times. The revealed user behaviors concerning the diversion towards HSR mode-service from air mode, show the following elements (Givoni & Dobruszkes, 2013; UIC, 2018; Russo et al., 2023):

- when HSR travel times are less than 2.5 hours, travelers' choices are oriented almost totally to HSR alternative (train is the dominant mode);

- when HSR travel time is about 3.5 hours, travelers' choices are equally distributed between train and air;
- when HSR travel times are higher than 5 hours, travelers prefer air mode.

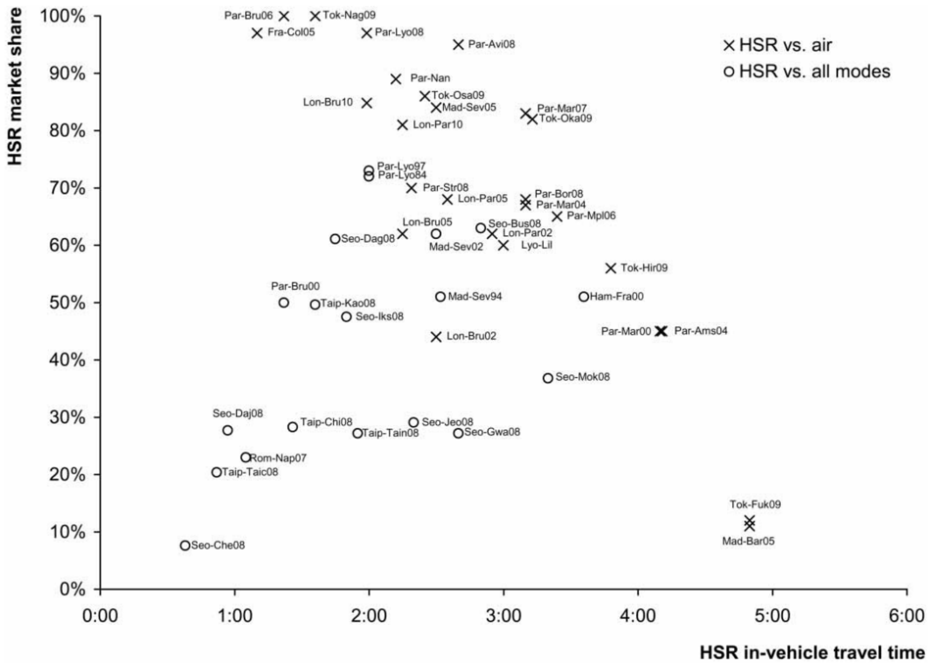


Fig. 1.3. HSR market share as a factor of in-vehicle travel time (source: Givoni & Dobruszkes, 2013)

Induced demand depends on (Cascetta & Coppola, 2014; Russo et al., 2024):

- travel behavior, in terms of frequency and destination and it is characterized by increasing accessibility of transport users; defined as utility in terms of reduced cost and new opportunities offered (Givoni & Dobruszkes, 2013; Wu & Martín, 2022; Musolino, 2024; Russo et al., 2024;).

Economy-based demand is characterized by national and international structure of the economy and concerns the component of new mobility indirectly determined by the modification of the economic conditions of the territories closer to the HSR lines. These changes consist in production-consumption, generation and location patterns. The dimension mainly involved with this segment is the one of new generation (Russo & Musolino, 2012, Cascetta & Coppola, 2014).

It has been conducted a bibliometric analysis on HSR travel demand on the mail scientific database: Scopus, Web of Science, Google Scholar. There were selected 72 papers, which were classified on the basis of the following criteria:

a) Demand component;
b) Country of analysis;
c) Choice dimension;
d) Approach of analysis.

Concerning the criterion of demand component (Fig. 1.4), the selection of articles mainly highlights the presence of papers concerning the description or the analysis of the diverted demand component (D). The diverted demand is a shift of traffic, or diversion, towards HSR mode-services due to the increasing utilities of users associated to the choice of HSR services rather than other transport modes (i.e. airplane, car), and other rail services (i.e. Intercity) (Dobruszkes, 2011; Givoni & Dobruszkes, 2013; D'Alfonso et al., 2016a).

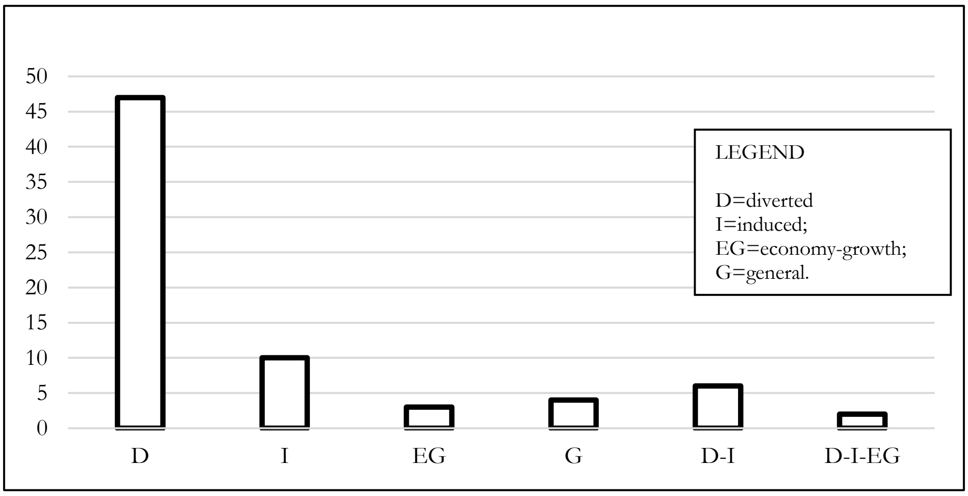


Fig. 1.4. Classification according to demand component

The diverted demand is the component of the demand that is, generally, the first macroscopic observed effect in the short-term after the opening of HSR lines, and for this reason it is the most studied in literature. It may be estimated by means of partial-share demand models, which simulate the choices of mode, service and run of travellers (Cascetta, 2013; Givoni & Dobruszkes, 2013; Nurhidayat et al., 2018; Wu & Martín, 2022).

The set of articles presents a number of works concerning the induced demand component (I) in quantity less than ten. The induced demand is the new traffic generated by the increasing accessibility of transport users (defined as utility in terms of reduced cost and new opportunities offered), due to the existence of HSR services (Givoni & Dobruszkes, 2013; Wu & Martín, 2022; Russo et al., 2024). The high level of service of HSR lines in terms of reduced travel time (as the sum of on-board travel time and access-egress time to city center), and high frequency connections between couple of cities reached by HSR determines two effects in the mid-term (Cascetta & Coppola, 2014; Russo et al., 2023). The first effect is a shift of traffic towards destinations served by HSR lines. As an example, users extend their area of opportunities for doing business, or for recreational activities, including destinations which were more difficult to reach before the opening of HSR lines. The second effect is the increment of frequency of trips along origin-destination pairs served by HSR mode-services. The literature dealing with induced demand is limited, as this segment of demand is more complex to identify, and it is connected with new travel needs of users that arise in the mid-term after the opening of HSR lines. This segment may be estimated with partial-share demand models, simulate the choices of making a trip and of destination (Russo et al., 2023; Musolino 2024; Russo et al., 2024).

The demand component connected to the economy-growth (EG), called economy-based demand, is indirectly determined by the HSR lines in the sense that the opening of an HSR line could cause an increase of value-added, employment and productivity in the cities, or areas, served by HSR, which in their turn, generate new travel demand. As an example, it could be recalled the case of the opening of new firms or offices close to a city served by HSR, which generates new travel demand of workers to/from these new destinations. This component of demand is generated by means of the so-called SETI models (Russo & Musolino, 2012; Russo et al., 2023). The set of articles contains less than five works.

In addition, works that concern the analysis of more than one demand components were selected, thus diverted and induced (D-I), and works that include all three demand components (D-I-EG): in both cases, the amount of studies is equal to and less than five.

The set of selected papers was also classified in relation to the country of study. The majority of the works concerns Chinese case studies (sixteen works), then nine works are identified dealing with Italian case studies, that are followed by seven works dealing with Spanish case studies. Finally, other studies deal with case studies in further countries in smaller numbers than the previous ones. Eight works concern meta-analyses concerning several countries. The above specified data are sketched in Fig. 1.5.

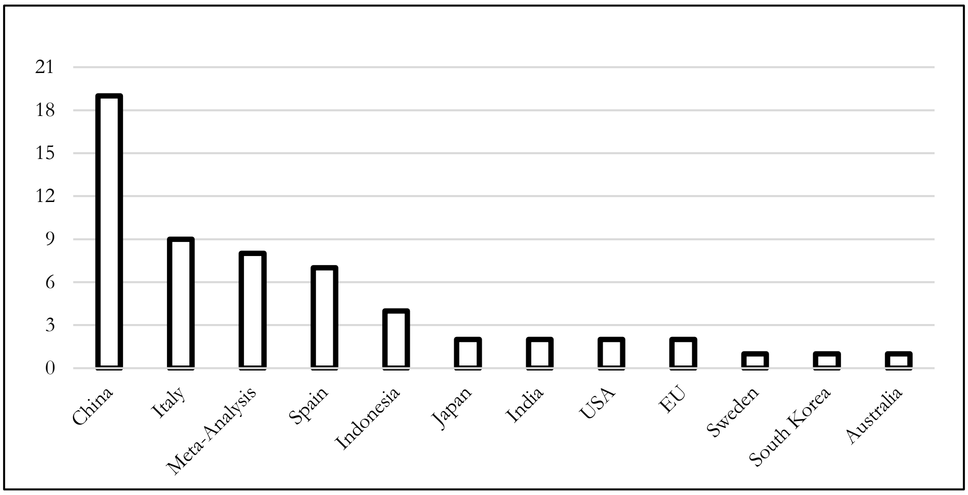


Fig. 1.5. Classification according to the country of case study

Some studies extracted from the selection of papers are briefly described below.

In Italy, Cascetta et al. (2011) presented a study of the impacts of the HSR service operating along the Rome-Naples relationship (Italy). The authors analyzed the diverted demand towards HSR services from other transport modes and from other rail services (e.g., Intercity, Eurostar services). They built a discrete choice model on the dimensions of mode-service-run for "home-based trips" and "non-home-based trips" purposes. Cascetta & Coppola (2012) calibrated a discrete choice model on the dimensions of mode-service-run for different travel purposes (e.g., business, and other purposes). The specified model was a multi-level nested logit model, where the different levels were transport modes, rail services, HSR companies and service classes. Cascetta & Coppola (2014) analyzed the effects of the entering of a single public HSR operator in the Italian railway market from 2005 to 2012; and the effects of the entering of a second private HSR operator from 2012. The authors developed an integrated modelling system to forecast the effects of competing timetables-services-prices between the HSR and air companies, car, conventional railway mode-services. Borsati & Albalate (2020) empirically studied the effects of the opening of HSR services in Italy on the total distances travelled by light vehicles along motorways during the period 2001–2017 and the entity of these effects after the opening of on-track competition between two HSR Italian companies.

As far as concerns France, Zembri (2010) provided an empiric study on the rail-air competition between HSR services, called TGV, and air services in France, and on the conditions resulting in the success of HSR service in terms of demand diverted. Behrens & Pels (2012) calibrated a nested (and mixed) multinomial logit models of HSR-air passenger demand traveling along the London-Paris relationship and estimated the direct elasticity of passenger demand with respect to frequency for business and leisure purposes. They considered the competition between a combination of four airports (Heathrow, Gatwick, Luton, London City) and four airline companies (Air France, British Airways, British Midland Airways, EasyJet) and the HSR service operated by Eurostar. Direct and cross elasticities are estimated with respect to travel time, frequency, and fare per each alternative, year, and trip purpose by means of a multinomial logit model.

The Spanish country was studied in some publications. Román et al., (2010) analyzed the extra-urban rail demand, considering the potential competitors to the HSR service, diverted by other modes of transport (air, private car and bus). A nested logit model is specified and calibrated, and the direct and cross elasticities are estimated. Román & Martín, (2014) simulate passenger choices between two transport alternatives: HSR rail and air services. Two models have been specified and calibrated: multinomial logit and mixed logit. Cross elasticity is introduced but not calibrated.

As far as concern China, Cheng (2010) examined the impact of HSR on the intercity transportation market in Taiwan. When HSR entered in operation the traffic was mainly diverted from air mode, conventional railway, and buses. It is worth noting that air transportation almost exited the market. Li & Sheng (2016) studied the diverted demand between two choice alternatives, air and

HSR, along the Beijing-Guangzhou corridor (China). They conducted a stated preference survey to estimate the parameters of multinomial logit-based discrete choice models. Zhang et al., (2017) used panel data of air passenger demand from 2010 to 2013 to analyze the effects of HSR on the main airlines in China. According to the authors, HSR services had relevant negative impacts on air demand, as they became much more elastic after the introduction of competing HSR services. Ren et al., (2019) analyzed the impact of HSR services on intercity travel behavior along the Chengdu-Chongqing corridor (China), with a focus on the diverted demand towards HSR services.

In Japan, Yao & Morikawa (2005) specified a multi-level discrete choice model on the dimensions of mode/service/run. The alternatives considered were bus, car, airplane. Clever & Hansen (2008) focused on competition between air and HSR modes in Japan, analyzing the trade-offs between accessibility, frequency, and speed of the two services.

Hensher (1997) specified and calibrated a discrete choice model in the dimensions of mode, service and run estimating the diverted demand to HSR from other modes in Australia. The direct and cross elasticities have been introduced and calculated.

As far as concern the criterion of choice dimension (Fig. 1.6), it is evident that the highest number of papers is on the level of mode (M) choice with sixteen works, therefore the mode and service (MS) level is studied in eight works; followed by the mode, service and run (MSR) levels analyzed in six works. The service choice is studied in two works; while there are two works that examine all the levels of choice considered, namely mode, service, company and run (MSCR).

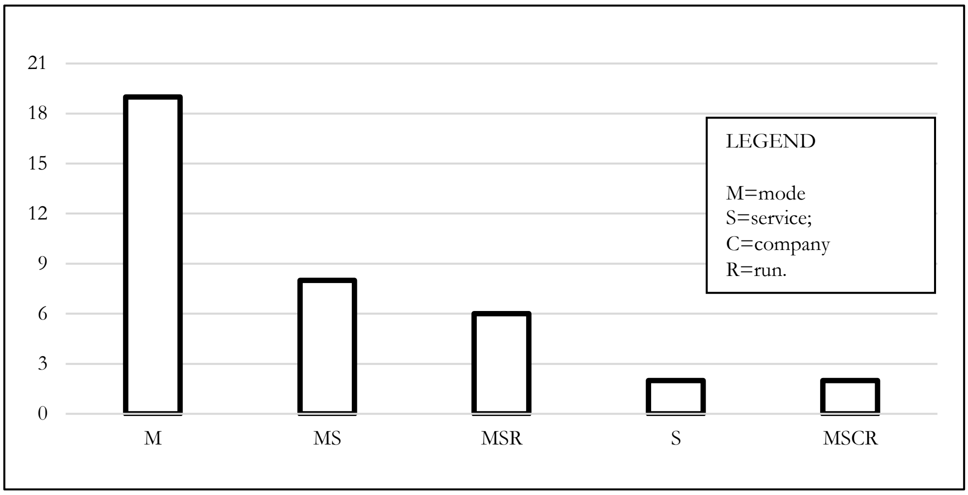


Fig. 1.6. Classification according to choice dimension

The selection of articles was classified according to the criterion of approach of analysis (Tab. 1.7).

The results show that two main methodological approaches may be identified for the assessment of the HSR travel demand: "ex-ante" and "before-after". The "ex-ante" group includes state-of-the-art and reviews of demand models developed for HSR lines planned in an examined country; whereas the "before-after" one regards the analysis and the comparison of demand observed before the construction and after the construction of HSR lines.

Some papers review the "ex-ante" approach in the existing literature. The paper of Wu & Martín (2022) presents a literature review of studies on passengers' preference regarding HSR. Nurhidayat et al. (2018) present a systematic analysis concerning HSR travel demand models at the mode choice level. The paper of D'Alfonso et al. (2016) provides a study on HSR-air transport competition in Europe on different classes of distance; thus, it analyzes diverted demand focusing on environmental implications.

The "before-after" approach is adopted in some literature reviews. Givoni & Dobruszkes (2013) analyze HSR diverted and induced demand based on before-and-after traffic data provided by a bibliometric analysis carried out on the main scientific databases. The comparison was carried out between years 1993 and 2010 and among several European couples of cities. The competition between HSR-air services is also treated in Dobruszkes (2011), who empirically compared the supply of air and HSR services along five relationships between European cities. The paper of Dobruszkes et al. (2014) analyzes the European air services and the impact of the entering of HSR on the mode-service levels, thus the modal competition between the two modes in terms of diverted demand. The study of Wan et al. (2016) investigates the HSR-air transport competition in Asian market.

The highest number of works are carried out considering possible future scenarios by applying the "before-after" approach (37 works), whereas the "ex-ante" approach was adopted in 24 works.

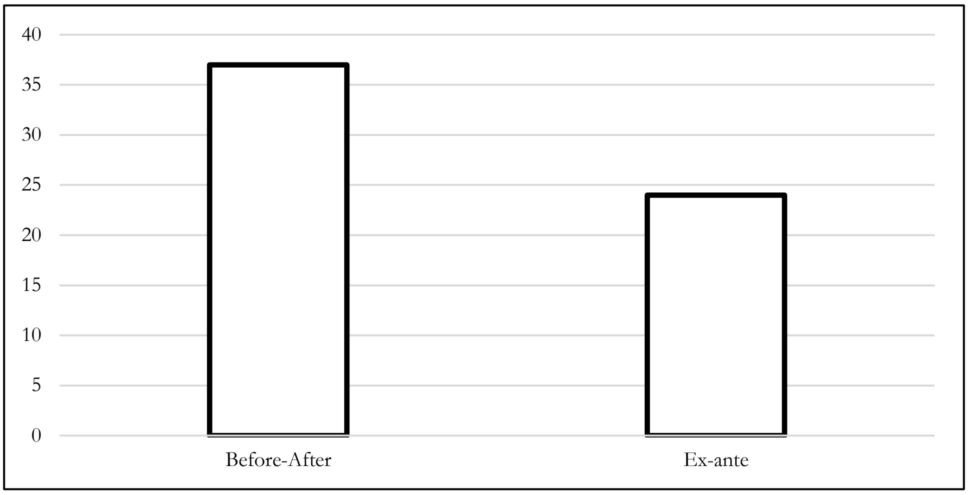


Fig. 1.7. Classification according to the approach of analysis

Some before-after studies present in the literature concern mainly the observation of the diverted demand component. There are few cases where the induced component has been explicitly quantified, while the demand generated from the economic growth was never quantified. The studies show, in general, positive effects in terms of diverted demand from other modes and from conventional rail services towards HSR services both (Givoni, 2006; Givoni & Dobruszkes, 2013; Beria et al., 2018;). In Italy the opening of HSR line from Turin to Salerno in 2009 generated an increment of railway traffic from 15 million of pax/year in 2009 to 43 million of pax/year in 2018 (Cascetta et al., 2020) of which 7 million have been diverted from conventional railway services, 19 million have been diverted from private car, buses, and air modes, and 17 million constitute the component of induced demand.

The selected publications, in some cases, are more general dealing with the presentation of a general framework for estimation of the HSR demand components described in the previous paragraph. As matter of fact, it is worth to recall the study of Ben-Akiva et al. (2010), that defined a general modelling framework for the evaluation of the three components of the HSR travel demand and their reciprocal interactions. The authors further analyzed the case of the entering of a new private HSR company in the Italian railway market, that generated a competition between private and public railway companies, and airline companies, on long-distance inter-city trips.

1.2.2. HSR demand components vs. choice dimensions modelling

According to the definitions of diverted demand and induced demand of the previous paragraph, it possible to associate the two HSR demand components and the dimension of choice models according to Table 1.2.

The diverted demand may be associated to the dimensions of mode-service and run. The mode-service choice model estimates the transport mode-service m chosen by the travelers given the origin o, the purpose s, the temporal interval h and the destination d (oshd):

$$d(m/oshd) = d(\mathbf{x}_m, \beta_m) \tag{1.1}$$

where: $\mathbf{x}_m$ is the vector of attributes and β_m is the vector of parameters to be calibrated.

The route, or run, choice model estimates the route, or run, r chosen by travellers given the origin, the purpose, the temporal interval, the destination and the mode-choice (oshdm):

$$d(r/oshdm) = d(\mathbf{x}_r, \beta_r) \tag{1.2}$$

where $\mathbf{x}_r$ is the vector of attributes and β_r is the vector of parameters to be calibrated.

The induced demand may be associated to the dimensions of trip generation and distribution. The trip generation model estimates the number of trips from an origin o, given the purpose and the temporal interval (sh):

$$d(o/sh) = d(\mathbf{x}_g, \beta_g) \tag{1.3}$$

where: $\mathbf{x}_g$ is the vector of attributes and β_g is the vector of parameters to be calibrated.

The trip distribution model estimates the percentage/probability of trips undertaken by travelers to a destination d, given the origin, the purpose and the temporal interval (osh):

$$p(d|osh) = p(\mathbf{x}_d, \beta_d) \tag{1.4}$$

where: $\mathbf{x}_d$ is the matrix of attributes and β_d is the vector of parameters to be calibrated.

The economy-based demand depends of factors that are exogenous to the transport system and may be associated to processes connected to the production-consumption generation and location (see Russo & Musolino, 2012).

Tab.1.2. HSR demand components and trip choice models

HSR demand component	Model	Type
Diverted	Mode-service/route(run)	Travel demand
Induced	Trip generation/distribution	
Economy-based	Production-consumption generation	SETI*
	Production-consumption location	

(*) Spatial Economic Transport Interaction model (source: Russo et al., 2023)

1.2.3. Run choice models

Two main approaches can be used in route choice modelling (Nuzzolo et al., 2000; Cascetta, 2013): 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 in relation to single runs is not possible, the method is useful for strategic planning.

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 method is useful for tactical planning.

The paragraph presents a literature review related 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 thesis, 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) (Tab. 1.3).

Tab.1.3. Literature review on run choice models in airline and HSR mode-services

Paper	Year	HSR	Air	Case study	Model
Nuzzolo et al.	2000	x		Italy	NL
Anderson & Wilson	2003		x	-	Purchase timing model
Yao & Morikawa	2005	x		Japan	NL
Espino et al.	2008	x		Italy	NL, schedule-based assignment
Jung & Yoo	2014	x	x	South Korea	MNL, NL
Cascetta & Coppola	2016	x		Italy	NL, stochastic assignment

With: NL=nested logit; MNL=multinomial logit.

On the HSR side, Nuzzolo et al., (2000) analyzed 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 line. Yao & Morikawa, (2005) modelled the intercity travel demand in Japan. Users' behaviour was analyzed with discrete choice models, the nested-logit model. The data was supplied by a survey and by aggregate data. The model included the trip generation, the destination choice, the mode and run choices. Cascetta & 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 basis 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 & Coppola (2017) 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 & Wilson, (2003) explored the context of users' ticket purchase behavior. 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 travelers' choices between waiting before buying a ticket with the objective of a cheapest fare occasion or purchasing a different type of ticket. Martín et al. (2008) discuss and analyze the users' willingness to pay for change ticket class in the Spanish airline services, thus the travelers' 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 & Yoo (2014) examined the mode-service choice behavior of travelers 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.

1.2.4. 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 alternative data collection methods 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 (WoS), Scholar.

The above mentioned procedure was applied to the papers listed below:

- Cascetta & Russo (1997) Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance. Transportation 24, 271–293;
- Russo & Vitetta (2011). Reverse assignment: calibrating link cost functions and updating demand from traffic counts and time measurements. Inverse Problems in Science and Engineering, 19, 7, 921-950.

Fig. 1.8 reports the synthetic diagram of the procedure that led to the final results described below. The procedure consists of 13 steps, applied to each article analyzed and 4 steps applied to compare the two analyzed papers. The steps are subdivided into two macro groups of analysis:

1) Analysis of the individual paper (11 steps);
2) Analysis and comparison between pairs of papers (4 steps).

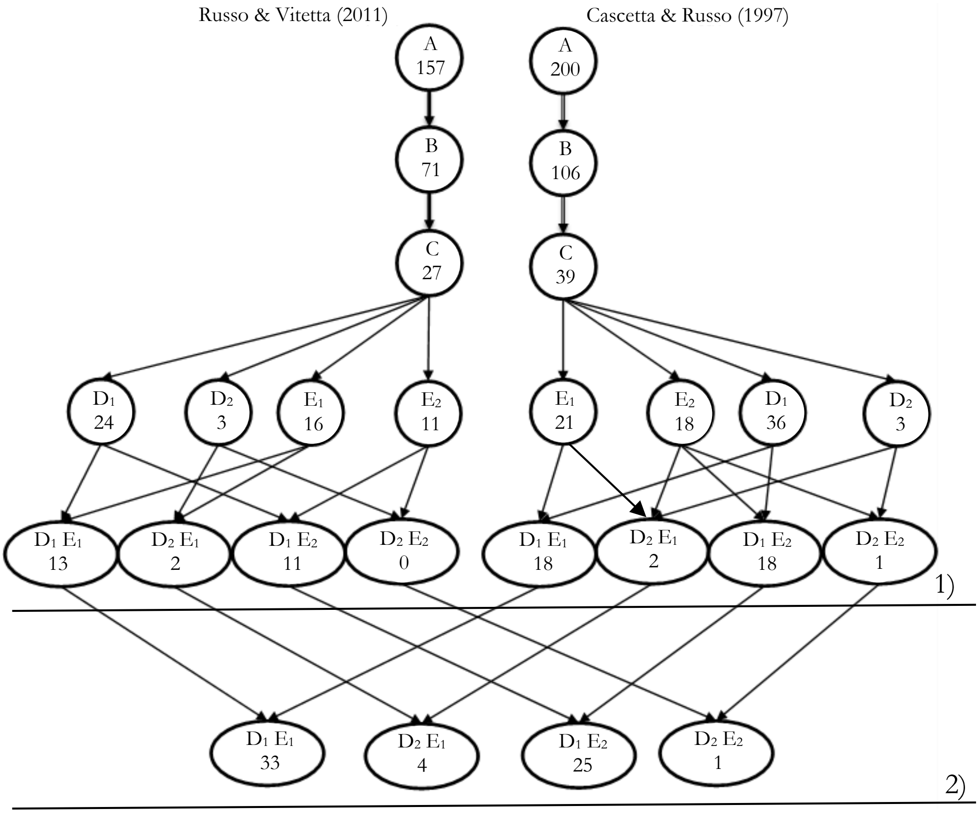


Fig. 1.8. Snowball sampling result

1. Analysis of individual paper

Step A. This step (A) concerns the identification in the selected scientific databases of the works that cite the papers under analysis. The paper (Cascetta & Russo, 1997) is cited in 200 works, whereas the paper of (Russo & Vitetta, 2011) is cited by 157 works.

Step B. The second step (B) concerns the elimination of duplications, as the citation is present in the different scientific databases. This merging operation reduces the number to 106 works for the paper (Cascetta & Russo, 1997) while to 71 works for the (Russo & Vitetta, 2011).

Step C. The third step (C) concerns a filtering operation from the sample obtained in Step B. The applied filter concerns works with quantitative results in terms of calibrated parameters. This filter operation reduces the sample to 39 works for the paper (Cascetta & Russo, 1997), while 27 works for the (Russo & Vitetta, 2011).

Step D. The fourth step (D) is sub-divided into 2 sub-levels, namely filters concerning quantitative results in terms of aggregated calibrations (D1) and quantitative results in terms of disaggregated calibrations (D2).

Step D_1. This step (D_1) allows to obtain, respectively with regard to the paper (Cascetta & Russo, 1997) a reduction of the sample to 36 works, while the paper (Russo & Vitetta, 2011) is reduced to 24 citations.

Step D_2. This step (D_2) allows to obtain, respectively with regard to the paper (Cascetta & Russo, 1997) a reduction of the set to 3 works, while the set is reduced to 3 works for the paper (Russo & Vitetta, 2011).

Step E. The fifth step (E) concerns the application of a filter to the sample extracted with Step C. The filter concerns aggregate calibrations in urban systems (E_1), and aggregate calibrations in extra-urban/regional/national systems (E_2).

Step E_1. The step (E_1) allows to obtain, respectively, a reduction of the set to 21 works for the paper (Cascetta & Russo, 1997), whereas the is reduced to 16 citations for the paper (Russo & Vitetta, 2011).

Step E_2. This step (E_1) allows to obtain, respectively with regard to the paper (Cascetta & Russo, 1997) a reduction of the sample to 18 works, while the paper (Russo & Vitetta, 2011) is reduced to 11 citations.

Step D_1E_1. The sixth step (D_1E_1) concerns the union of the samples extracted from sample C, i.e. sample D_1 and sample E_1. Therefore, the merging operation results between aggregate calibrations and urban systems. The operation allows to obtain 18 works that cite the paper of (Cascetta & Russo, 1997), while the paper of Russo & Vitetta, 2011 is cited by 13 works.

Step D_1E_2. The seventh step (D_1E_2) concerns the union of the samples extracted from sample C, thus sample D_1 and sample E_2. Therefore, the merging operation results between aggregate calibrations and extra-urban/regional/national systems. The operation allows to obtain 18 works for the paper of Cascetta & Russo, 1997, while the paper of Russo & Vitetta, 2011 is cited by 11 works.

Step D_2E_1. Step eight (D_2E_1) concerns the union of the samples extracted from sample C, hence sample D_2 and sample E_1. The merging operation concerns disaggregated calibrations and urban systems. The operation allows to obtain 2 works for the paper of Cascetta & Russo, 1997, while the paper of Russo & Vitetta, 2011 is cited by 2 works.

Step D_2E_2. The seventh step (D_2E_2) concerns the union of the samples extracted from sample C, thus sample D_2 and sample E_2. Therefore, the merging operation results between disaggregated calibrations and extra-urban/regional/national systems. The operation allows to obtain only 1 work for the paper of Cascetta & Russo, 1997, while the paper Russo & Vitetta, 2011 is cited by 0 works.

2. Analysis and comparison between pairs of papers

Step D_1E_1 (Cascetta & Russo, 1997; Russo & Vitetta, 2011). The step (D_1E_1) concerns the union of the samples extracted from sample C of both papers, hence sample D_1 and sample E_1. Therefore, the merging operation results between aggregate calibrations and urban systems. The operation allows to obtain 33 works.

Step D_1E_2 (Cascetta & Russo, 1997; Russo & Vitetta, 2011). This step (D_1E_2) concerns the union of the samples extracted from sample C of both papers, thus sample D_1 and sample E_2. Therefore, the merging operation results between aggregate calibrations and extra-urban/regional/national systems. The operation in question allows to obtain 25 works.

Step D_2E_1 (Cascetta & Russo, 1997; Russo & Vitetta, 2011). The step (D_2E_1) in question concerns the union of the samples extracted from sample C of both papers, hence sample D_2 and sample E_1. The union results between disaggregated calibrations and urban systems. The operation in question allows you to obtain 4 works.

Step D_2E_2 (Cascetta & Russo, 1997; Russo & Vitetta, 2011). The step (D_2E_2) in question concerns the union of the samples extracted from sample C of both papers, hence sample D_2 and sample E_2. The union results between disaggregated calibrations and extra-urban/regional/national systems. The operation in question allows you to obtain 1 work.

A total of 354 works were identified and considering the simultaneous presence of the work 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 calibration: 54 works with aggregated calibration and 6 works with aggregated calibration.
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 (Tab. 1.4) is reported below. The last right column of Tab. 1.4 reports the publication that is the source of each publication examined (Russo, Musolino, et al., 2025).

Tab.1.4. 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 & 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 & Vitetta, 2011)
Integrated travel demand models for evacuations: a bridge between social science and engineering	Russo F, Chilà G	2014	(Russo & 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 & Russo, 1997)

Source: Russo et al., (2025)

1.3. State of the art on fares supply modelling

1.3.1. Fare definitions

The fare may be defined as follows: "the money that you pay to travel by bus, plane, taxi, etc." (Oxford Learner's Dictionaries, 2025), or "the money a person pays to travel on a bus, train, boat, or airplane or in a taxi" (The Britannica Dictionary, 2025). Fare structures are mainly classified into static (e.g., local rail services) and dynamic (e.g., HSR services, air services) (Fig. 1.10) (Zheng & Liu, 2016; Russo et al., 2024).

The dynamic structure was first introduced in the 1970s in the airline market for scheduled services. The dynamic fare structure generally depends on the ticket variations over time from the first day of availability of tickets to the last one (day of trip). The variations depend mainly on several phenomena, for instance the behavior of travelers and of commercial (pricing policy and commercial strategies of the company) factors, of the interaction between demand and supply. The above-mentioned variables are different in relation to the country and the companies involved, but generally result to be, for instance: the typology of ticket (Economy, Business, …), the day of trip, the day of ticket purchasing, the travelled distance, the purchasing channel, the period of availability of the ticket, the train capacity (Meng et al., 2022). The dynamic structure of fares was introduced in the HSR more recently from the air context. The characteristics outlined above are interpreted in different ways by the actors involved, such as travelers and transport companies. The main difference between the two actors is that the companies generally consider the fare value in order to maximize revenues in relation to the competitive market, whereas the travelers mainly consider the fare value in order to minimize the monetary costs of the trip or maximize utility (Russo et al., 2024).

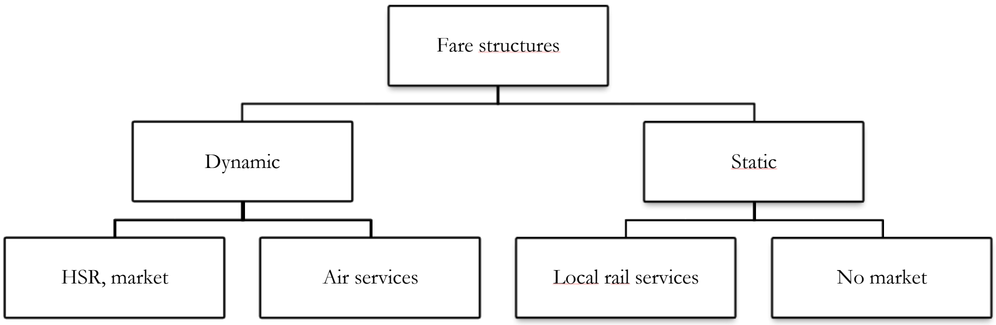


Fig. 1.9. Classification of fare structures

In the field of transport supply, the dynamic pricing received great attention from the transport companies, due to their objectives connected to revenue management maximization and seat allocation optimization. There is wide scientific literature about the fare analysis in the air transport context, whereas in the rail transport is less explored and detailed. In addition, dynamic fares play a key role in demand regulation, thus the variation of the fare value has a direct impact on the ticket purchasing and run choice behavior of the main two typologies of travelers: price-based and time-based. The fare value is the result of a combination of the pricing policy decisions of transport companies and of travelers' behavior (Russo et al., 2024; Russo et al., 2025), determining the competition among the mode and/or company choices in the intercity mobility at national scale.

1.3.2. Fare structure models

The existing literature shows that the fare structure modelling, especially in the transportation market, is mainly oriented towards two objectives: revenue management (RM) and seats allocation (SA). On the other hand, the traveler's perspective typically is based on the travel costs reduction (travel time or monetary cost) (Russo et al., 2024; Russo et al., 2025). Therefore, values of fares are generally the result of the maximization of revenue management objectives, given the demand and the category of travelers (McAfee & te Velde, 2006; Pels & Rietveld, 2004). The dynamics of the fares is also used to regulate the demand in order to maximize the seat allocation (Wang et al., 2016; Yan et al., 2020). The dynamic fare structure (Khwanpruk et al., 2021; Zhang et al.,

2017) generally depends on the ticket variations over time from the first day of availability to the last one (day of trip). The literature classifies the travelers' ticket choice behavior upon two different elements which identify two users' classes: time-based user and price-based user (Xiao et al., 2008; Khwanpruk et al., 2021); On one hand, the time-based one chooses a run with the minimum travel time, even with short distance between day of trip and day of ticket purchasing, regardless of the ticket fare (typically a business user), thus have a higher willingness to pay. On the other hand, the price-based user goal is to minimize the monetary cost of the travel (typically a non-business user); weighing less the travel time, to the point of changing day of trip in order to reduce the travel monetary cost, thus with a lower willingness to pay (Russo et al., 2025).

Tab. 1.5 contains some selected papers, classified by authors, year of publication, mode analyzed (HSR or Air), country of the case study, type of analysis (user behavior or fare estimation) and objective of the analysis (revenue management or seat allocation). The table shows that papers are mainly finalized to revenue management or seat allocation.

Tab.1.5. Literature review on dynamic fare structure in rail and air markets

Paper	Year	HSR	Air	Country	User behaviour	Fare estimation	RM/SA
Pels E., Rietveld P.	2004		x				x
McAfee P.R., te Velde V.	2006		x				x
Xiao YB., et al	2008		x	China	x	x	x
Koenigsberg O., et al.	2008		x	UK		x	x
Malighetti P., et al	2009		x	EU		x	x
Yao E., et al	2013	x		China	x		x
Hetrakul P., Cirillo C.	2014		x	.	x	x	x
Malighetti P., et al	2015		x	EU		x	x
Wang X., et al	2016	x					x
Wan Y., et al	2016		x			x	x
Zheng J., Liu J.	2016	x		China			x
Zheng J., et al	2017	x		China			x
Morlotti C. et al	2017		x	EU		x	x
Selcuk A., Avsar Z.	2019	x		EU	x		x
Khwanpruk S., et al	2021	x				x	x
Meng H., et al	2022	x					x

With: RM=Revenue Management; SA=Seat Allocation.

a. Air market

Pels & Rietveld (2004) examine the airline pricing on the London–Paris route, as well as it is focused on the interactions between low-cost and traditional carriers. An important question in the work is whether carriers react to the price adjustment of other airline companies. It results that easyJet company react to the fares of British Midland. Whereas conventional carriers do not react to the price adjustments of the low-cost companies. Therefore, some carriers seem to reduce their fares when potential competitors increase their fares. The authors understand this by supposing that it may occur because the price movement of the potential competitor may be a signal of market saturation. All carriers increase their fares as the departure date approaches.

McAfee & te Velde (2006) work regards an exhaustive economic framework concerning the dynamic price distinction under capacity constraints and demand uncertainty in the airline sector. It is shown how airline companies adjust the fares on the basis of future sales, which varies with time and seats available. In the paper is has been carried out surveys of the theoretical literature on dynamic price distinction and compare the theories with new data from airline pricing behavior.

Xiao et al. (2008) investigate travelers' behavior about ticket purchasing by means of a classification based on two main criteria, which are: price-based choice, where the travelers choose a flight with the lowest fare available, independently of the departure time; and time-based choice, where the travelers choose the flight on the basis of their desired departure time, regardless to the amount of the fare. The case study concerns the definition of the pricing strategy of an airline company by optimizing the revenue management on multiple flights.

Koenigsberg et al. (2008) deal with low-cost fare strategies in airline transportation by analyzing the scenario of a last-minute and low-fare pricing policy in the context of short-haul flights. The structure of fares is analyzed in the European case study, and it is based on the easyJet pricing policy in contrast to traditional airline pricing policies. The analysis of ticket fares over a period of observation shows the variation of the fares in relation to the seat occupation. The observation of fares purchasing shows that particularly relevant fare variations happen during the most distant periods from the day of trip, characterized by lower average fares (typical behavior of the non-business users), whereas the highest fares variation is observed approaching the flight day, thus during closest to the day of trip. The empirical results of the analysis show that the fares value, organized in several segments, depends on the remaining capacity at a given time over the selling period, introducing the phenomenon of last-minute discounting, hence an additional selling period in which the airline has unsold seats.

Malighetti et al. (2009) paper deepened the pricing policy of a low-cost airline company, thus the Ryanair case study and in particular are analyzed the European flights. It is specified a model in order to estimate the optimal fare curve for each flight, hence a family of hyperbolic price functions. It is stated that the pricing policy of the company is based on the Revenue Management objective. Therefore, the work provides evidence about how low-cost flights use advanced dynamic pricing systems, to maximize revenues with the maintaining of low average fares.

Hetrakul & Cirillo (2014) analyze the travelers' behavior about ticket purchasing with the aim of the identification of the optimal Revenue Management (RM) strategy. The results of fare analysis show that, on average, the fare variations over time are related to the distance between the day of trip and the day of ticket purchasing, to the departure time during the day, to the day of the week. The authors assume that users' behavior about ticket purchasing is based on the fare variations over time and on personal attitudes towards the trip, such as departure time of day, day of week, and trip distance.

Malighetti et al. (2015) deals with the EasyJet pricing policy by estimating fare structures that take into account the effects of the demand. The objective of the paper concerns the identification of elements of competition based on the observed fares and the role of demand in the pricing strategy. The paper shows that the competition reduces average fares (e.g., in the analysis period between 2008 and 2009 of almost 10%). The fares offered, for different routes and departure days, were surveyed from the company website. The results show that the ticket fares are higher and less discounted when there is direct competition among airlines companies on a specific origin-destination relationship. The parameters of the calibrated model show an increasing of the slope of the fare curve with the decreasing of the distance between the day of trip and day of ticket purchasing.

Wan et al. (2016) paper investigate the impact of the entering of HSR services on the airlines' domestic flights and available seats, especially on specific relationships in China, Japan, and South Korea, which are subjected to the mode diversion. The study is based on a dataset which covers the period between the years 1994 and 2012. It has been developed the difference-in-difference approach in order to estimate the impact of HSR entry. The results show that the HSR entry causes different impacts, in relation whether it regards short-haul, medium-haul or long-haul on the airline services in relation to the speed and the travel distance.

Morlotti et al. (2017) investigates the price elasticity of demand in the low-cost flights in the European context. The study analyzed the choice behavior, based on the assumption that the user travels for leisure or business reasons. They analyzed a dataset of reservations and fares offered online by the transport company EasyJet for flights during the period 8 March 2015 to 23 September 2015. The results of the fares survey show that, on average, on all the flights considered, the fares increase with the decreasing of the distance between the day of purchasing and the day of trip, in particular during last ten days. Therefore, the results show that flights early in the morning and during the working days are more business users oriented, whereas the leisure users typically prefer to travel on weekend days.

b. HSR market

Yao et al. (2013) analyze the High Speed Rail (HSR) pricing strategies with the aim of revenue maximizing throughout seat allocation strategies. A discrete choice model (Nested Logit) was calibrated in order to identify the relationship between ticket fare and HSR market share, based on stated preferences data. The results show that the probability of the users' choice behaviour to choose the HSR depends on the travel time and travel cost of each available alternative (HSR, conventional rail, air, and road), the profession and age of travellers and the trip purpose. The authors performed a sensitivity analysis on the marker share according to different assumptions of fare variations.

Zheng & Liu (2016) handle different fare structures in order to optimize fares with the aim of maximizing revenues of companies, taking into account the effects on demand. The case study analyses the time-series of ticket-sale data in the Chinese HSR market context. The authors stated that the travelers' tickets purchasing is conditioned by various factors, such as fare purchasing behavior of travelers, travel distance, number of runs that serve the same origin-destination relationship, trip purpose.

Zheng et al. (2017) analyzes the fare structures and optimal timing of pricing, considering the role of transport demand. The paper shows that company pricing strategy, in terms of number of seats that is predetermined and changeless, and number of fare classes, has impacts on the travelers' fare choice behavior. The case study regards Chinese HSR market and it is based on historical ticket data analysis of HSR lines, hence have been considered a selling period of the tickets and the demand density, defined as the number of tickets sold divided by ticket selling hours in a day.

Selcuk & Avsar (2019) analyze the dynamic pricing in the context of air transportation with the aim of revenue management. They investigate the relationship between the aggregate demand and the fares' evolution over time. The implementation of a procedure for real-time dynamic pricing in a Turkish case study shows the influence on the pricing policy, in terms of price variation, on the number of the available seats.

Khwanpruk et al. (2021) paper examined the implementation of dynamic pricing in the HSR services, in particular the work was developed within the Thailand's rail context. The specified model takes into account the time of the ticket purchasing and the quantity of reservations. The model was applied for the objective of Revenue Management, in addition was considered the pricing policy of the company. The obtained results showed the correlation between the dynamic fares and the distance travelled.

Meng et al. (2022) work explores the application of the dynamic pricing on the fares, combined with ticket allocation, in the context of HSR services. The study developed a stochastic optimization framework, taking into account operators' risk preferences, demand uncertainty and capacity constraints. The proposed model specifies pricing and allocation policies as functions of demand, booking-time information, and travelers' choice probabilities. The results showed that including risk attitudes considerably impacts on the optimal dynamic fare structures and increase the Revenue Management performance.

1.4. Literature gaps and research contribution

Several studies have been carried out on HSR demand models in scientific literature along 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 intra-modal levels (e.g., service, company and run choice).

The competition among the different mode-services-companies-runs alternatives depends also on the structures of fares. They are generally classified into static and dynamic over time. On one hand, the static fares are generally referred to the local rail services and to the bus services. On the other hand, the dynamic fares historically have characterized the airline services and, in the last decades, it has been adopted also in the HSR ones.

According to the above considerations that arise from the literature review, 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 literature and implies the possibility of using different attributes, while intra-modal competition, less studied, mainly refers mainly to attributes connected to fares and departure times.

There are two important themes in intra-modal competition:

- the run choice model;
- the evolution of dynamic fares.

The two themes are studied separately in literature. The possibility of using innovative methods for collecting data concerning the evolution of fares allows the two themes to be combined and unified into the wider research line called: “reverse assignment”.

The thesis responds to RQ1 by proposing an innovative method for collecting information based on the acquisition of data from big data, which contains a great volume and variety of historical information about the characteristics of the transport supply (fares). The proposed framework could supports the identification of users’ choice of the HSR run from observation of HSR fares.

The thesis tries to respond to RQ2 by combining the two lines of research concerning the run choice models; on the demand side, the dynamic fare functions, on the supply side. The proposed framework has been validated by means of the specification-calibration of a choice model in the run dimension and of fare structure function.

2. Mathematical formulation of run choice and fare supply modelling

The chapter presents general definitions and notations of variables used in the data-bases and built models.

Specifically, the main reference variables are defined, namely the day of trip and the day of ticket purchasing, on this basis the statement of the problem has been analyzed. Therefore, all the variables used in the mathematical formulations specified within the thesis are defined. The variables relating respectively to the run choice model and to the model of fare structures are defined, therefore regarding the relationship analyzed, the mode and service considered, the companies relating to the runs analyzed and the years in which the survey was executed.

2.1. Definition and general notation of variables

The paragraph presents the definitions and the general notations of the variables used in the data-bases and models developed. Fig. 2.1 describes the statement of the problem. On one hand, the variable t is the day of trip, the day when the user decides to make a trip from an origin to a destination. On the other hand, the variable p is the day of ticket purchasing, thus the day on which the user decides to purchase the ticket of a run (service). Both t and p are dates of calendar. The difference between t and p, k, is defined as the distance, measured in days, between the day (date) of trip and the day (date) of ticket purchasing. In the remaining part of the thesis, the day of ticket purchasing will be indicated in terms of distance between the day of trip and the day of ticket purchasing, k (Fig. 2.1).

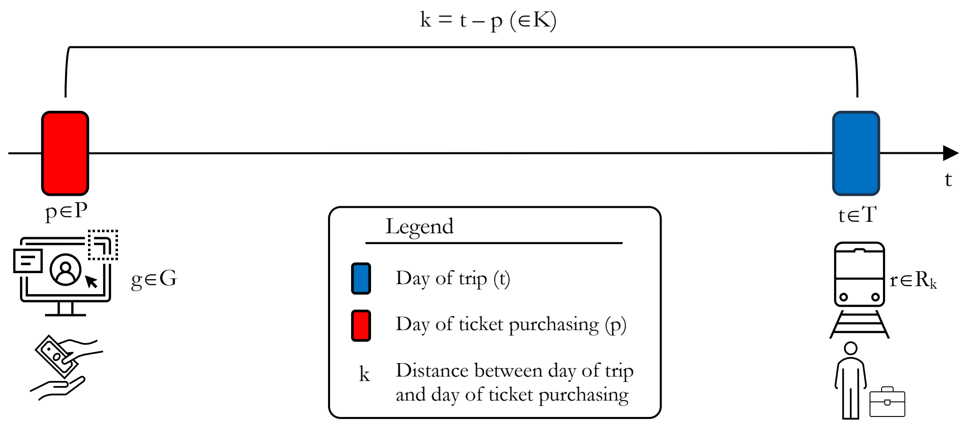


Fig. 2.1. Statement of the problem of fare collection

Let's:

- i, user
- T, set of days of trip
- $t \in T$, day of trip of user i
- $\overline{T} = |T|$ = cardinality of T
- K^t, set of days of ticket purchasing before t
- $k^t \in K^t$, day of ticket purchasing, or distance between the day (date) of trip and the day (date) 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)
- $\overline{K}^t = |K^t|$ = cardinality of K^t
- R^t_k, set of runs available at day k for travelling at day of trip t
- $r^t_k \in R^t_k$, generic run belonging to the set R^t_k
- $\overline{R}^t_k = |R^t_k|$ = cardinality of R^t_k, or maximum number of available runs during K^t
- G^t_k, set of fare types (classes) for all runs and all companies at day of trip t available in the day of ticket purchasing k
- g^t_k, generic type (class) of fare $g^t_k \in G^t_k$
- $\overline{G}^t_k = |G^t_k|$ = cardinality of G^t_k
- w^t_k (r, g), fare of run r at day k for fare type (class) g

- W^t_k [$\overline{R}^t$, $\overline{G}^t$], matrix of runs and fares, composed of a number of rows equal to the number of runs, $\overline{R}^t$, of all companies at day k, and a number of columns equal to the total number of fares (classes) offered by all companies
- $\Delta\, w^t_k$ [r^t, g^t], difference between the value of the ticket at day k^t, $w_k(r, g)$, and the value of the ticket at day k+1, $w_{k+1}(r, g)$
- $\Delta\, W^t_k$ [$\overline{R}^t$, $\overline{G}^t$], matrices differences, evidencing all the modified data referring to the day-to-day ticket series
- $w'^t_k(r^t, g^t)$ =code of fare mutation for each run r and in each day k, defined as:
- $$w'^t_k(r^t, g^t) = \begin{cases} 0, \text{ if } \Delta w^t_k(r^t, g^t) = 0 \\ 1, \text{ if } \Delta w^t_k(r^t, g^t) \neq 0 \end{cases}$$

It is possible to introduce simplifications, for instance $k^t \rightarrow k$ if this doesn't cause confusion; in the same way r, g and so on.

2.2. Specification of variables of the run choice model

The paragraph presents the specification of variables of the run choice model. The examples given refer to the following domain of analysis:

- Relationship: Rome-Milan;
- Mode/service: High Speed Rail (HSR);
- Companies: Trenitalia Frecciarossa, NTV Italo;
- Years: 2022, 2023.
- T, set of days of trip

Example:

T = {01/09/22; 08/09/22; 08/09/23; 08/09/23}

- t∈T, set of day of trip

Example:

t_1= 01/09/22 ∈ T

t_2= 08/09/23 ∈ T

- K^t, set of days of ticket purchasing before t;

  Example:

  K^{t1}={01/08, 02/08, …., 31/08}

  $\overline{K}^{t_1}$=31

  K^{t2}={01/08, 02/08, …., 07/09}

  $\overline{K}^{t_2}$=38

- $k^t \in K^t$, 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);

  $k^{t1} \in K^{t1}$

  Example:

  k^{t1}_{10} = 22/08/22

- R^t_k, set of runs available at day k for travelling at day of trip t;

  t_1=01/09/22

  k_{10}=22/08/22

  Example:

  R^{t1}_{10} = {9606, 9608, 9612, 9618, 9620, 9622, 9968, 9972, 9603, 9605, 9607, 9604, 9611, 9613, 9615, 9967, 99671, 9998}

- $\overline{R}^t_k$, cardinality of R^t_k, or number of available runs,

  Example:

  $\overline{R}^1_{10}$ = 18

- $r^t_k \in R^t_k$, available runs at day k (belonging to the set R^t_k)

  Example:

  $r^1_{10} = r^1_{10}[4]$ = 9618

- $\overline{G}^t_k$, maximum number of fare types (classes) for all runs and all companies at day of trip t available in the day of ticket purchasing k;

  Example:

  G^{t1}_{k10} = {1, …, 21, 22, …., 33} = {“Super Economy-Standard”, …, “Base-Executive”, “Low Cost-Smart”, …., “Flex-Salotto”}

- $\overline{G}$, cardinality of G^t_k;

  Example:

  $\overline{G}^1_{10}, \forall r = 33$

- g^t_k, generic type (class) of fare $g^t_k \in G^t_k$;

  Example:

  g^t_k = "Super Economy-Standard" (Trenitalia) = $1 \in G^{t1}_{k10}$

- $w^t_k(r, g)$, fare of run r at day k for fare type (class) g;

  Example:

  $w^1_{10}(9606, 1) = 49{,}9$ €

  $w^1_{10}(9618, 1) = 51{,}9$ €

- $W_k[\overline{R}, \overline{G}]$, matrix of runs and fares, composed of a number of rows equal to the number of runs, $\overline{R}$, of all companies at day k, and a number of columns equal to the total number of fares (classes) offered by all companies.

  Example:

  $W^1_{10}=[18, 21]$

- $\Delta\, w_k[r, g]$, difference between the value of the ticket at day k, $w_k(r, g)$, and the value of the ticket at day k+1, $w_{k+1}(r, g)$;

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

  Example:

  $\Delta\, w^1_{10}\ (9606, g=1) = w^1_{10}(9606, g=1) - w^1_{11}(9606, g=1) = 52{,}9€ - 49{,}9€ = 4{,}0$ €

- $\Delta W_k[\overline{R},\ \overline{G}]$, matrices differences, evidencing all the modified data referring to the day-to-day ticket series;

  Example:

  $\Delta W^1_{10}\,[18, 33] = [\ldots, 4{,}0\ €, \ldots.]$

- $w'_k(r, g)$, fare mutation for each run r and in each day k

  Example:

  $w'_{k10,11}[9606, 1] = 1$

2.3. Specification of variables of the fare supply model

The paragraph presents the specification of variables of the fare supply model. The examples given refer to the following domain of analysis:

- Relationship: Rome-Milan;
- Mode/service: High Speed Rail (HSR);
- Companies: Trenitalia Frecciarossa, NTV Italo;
- Years: 2022, 2023.

Let's:

- T, set of days of trip

  It is possible to partition T in different sets.

  Example:

  $T=\{T_a; T_b\}=\{16/08/22, \ldots, 06/09/22, 16/08/23, \ldots, 06/09/23\}$

  $T_a=\{16/08/22, \ldots, 06/09/22\}$

  $T_b=\{16/08/23, \ldots, 06/09/23\}$

- $t \in T$, day of trip

  Example:

  $t^a_1 \in T_a$; $t^a_1=16/08/22$; $t^a_2=17/08/22$; $t^a_3=18/08/22$

- K, set of days of ticket purchasing before t;

  Example:

  $K^{ta1}= K^{16/08/22} = \{15/08/22, 14/08/22, \ldots, 26/07/22\}$

- $k^t \in K^t$, 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);

  Example:

  $k^{ta1}_2 = k^{16/08/22}_2 = 14/08/22$

- R^t_k, set of runs available at day k for travelling at day of trip t

Example:

$R^1_{10} = R^{16/8/22}_{6/8/22} = \{9508, 6636, 9634, 6770, 9560, 5642, 9515, 6637, 9631, 6767, 9567, 4479\}$

- $\overline{R}^t_k$, cardinality of R^t_k, or number of available runs

Example:
$\overline{R}^4_{10} = \overline{R}^{16/8/22}_{6/8/22} = 10$

- $r^t_k \in R^t_k$, available runs at day k (belonging to the set R^t_k)

  $r^t_k \in R^{16/8/22}_{6/9/22}$

Example:
$r^{16/8/22}_{6/8/22}[4] = 6770$

- G^t_k, maximum number of fare types (classes) for all runs and all companies at day of trip t available in the day of ticket purchasing k;

Example:
$G^{t1}_{k=10} = \{1,\ldots, 21, 22, \ldots, 33\}$ = {"Super Economy-Standard", …, "Base-Executive", "Low Cost-Smart", …, "Flex-Salotto"}

- g^t_k, generic type (class) of fare $g \in G^t_k$;
  Esempio:
  g^t_k = "Super Economy-Standard" (Trenitalia) = $1 \in G^{t1}_{k=10}$

- $w_k(r, g)$, fare of run r at day k for fare type (class) g;

Example:
$w^1_{10}(9508, 1) = 49{,}9$ €
$w^1_{10}(6770, 1) = 65{,}9$ €

- $W_k[\overline{R}, \overline{G}]$, matrix of runs and fares, composed of a number of rows equal to the number of runs, $\overline{R}$, of all companies at day k, and a number of columns equal to the total number of fares (classes) offered by all companies.

Example:
$W_{k=10} = [12, 33]$

3.

Innovative data collection of fares

The chapter gives a description of the surveys of fare carried out. The survey was designed and conducted over three different years: 2022, 2023 e 2025, in order to extract the information related to the fare supplied of the chosen run, thus belonging to the HSR transport companies. Therefore, the data-base was designed in order to store the surveyed data. The executed surveys have provided data that has been stored in data-bases, from which samples were extracted and then used to calibrate the fare structure models.

The surveys were carried out for the following different samples:

- The first survey was carried out to obtain a sample for the estimation of the run choice model, then a disaggregated sample (paragraph 3.1.1);
- The second survey was carried out to obtain a sample for the estimation of the run choice model, then a disaggregated sample (paragraph 3.1.2);
- The third survey was carried out to obtain a sample for the estimation of the cost function in supply, then an aggregate sample (paragraph 3.2).

3.1. Survey and data-base building for run choice modelling

The survey has been executed over three years: 2022, 2023 and 2025; by considering the following dates as p and t (Fig. 3.1):

- years 2022 and 2023:
  - p=02/08-07/09;
  - t=01/09, 08/09.
- year 2025:
  - p=06/06-18/06;
  - t=11/06, 12/06, 17/06, 18/06, 24/06, 25/06.

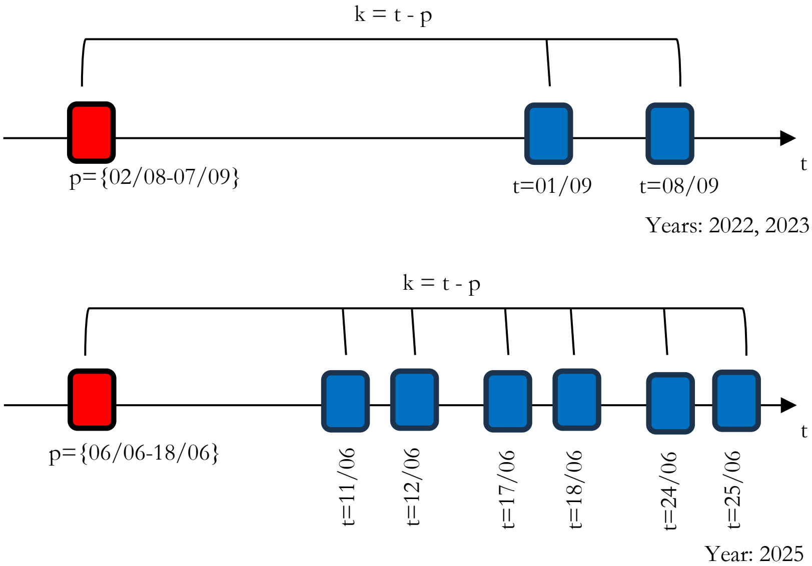


Fig. 3.1. Scheme of key dates considered in the surveys

The surveys have been designed and executed for the years 2022 and 2023, considering two fixed days of trip (t) and a set of days of ticket purchasing (p). Whereas the design of the survey was improved in the year 2025 in order to obtain the same, or more, amount of data by reducing the survey time. In particular, the set of days (p) was reduced, and the number of days (t) was incremented reducing the survey extension to the days of the ticket purchasing (p). The survey over the different years allows to obtain the following samples of users (Fig. 3.2):

- Minimum prototype sample, from the 2023 survey and composed of 30 users;
- Median prototype sample, from the 2023 survey and composed of 141 users;
- Full sample, from the 2022 survey and composed of 591 users;
- Full sample, from the 2023 survey and composed of 592 users;
- Full sample, from the 2025 survey and composed of 635 users.

Fig. 3.2 reports the general configuration of the survey executed in the different years and the samples extracted.

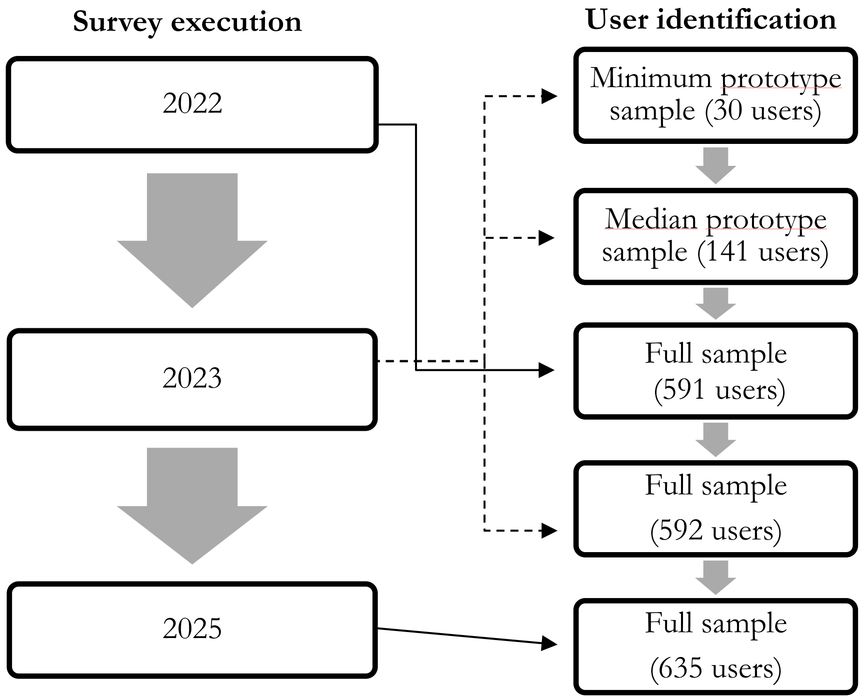


Fig. 3.2. Survey execution and users' identification

3.1.1. Survey: years 2022 and 2023

The paragraph deals with the survey of HSR ticket fares starting from the consultation of the websites of two Italian HSR transport companies (Trenitalia and NTV). The characteristics detected are reported below (Fig. 3.3).

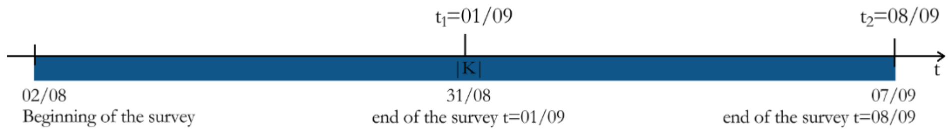


Fig. 3.3. Temporal characteristics: disaggregated survey

Temporal reference

- Years
  - 2022;
  - 2023.
- Days of trip (t):
  - 01/09;

    - 08/09.
- Days of ticket purchasing (p): Tab. 3.1.

Tab.3.1. Survey period for run choice modelling: years 2022 and 2023

Set of days of ticket purchasing (p)	Day of trip (t)
01/08-31/08	01/09
01/08-07/09	08/09

Characteristics of the service:

- Mode/service: High Speed Rail (HSR);
- Companies:
    - Trenitalia, services “Frecciarossa”;
    - NTV, services “Italo”.
- Analyzed relationship:
    - Rome-Milan;
    - Milan-Rome.

Set of runs analyzed

- Time slot: 05:00-10:00;
- Maximum number of intermediate stops: 1.

Set of fares considered

- Trenitalia:
    - 21 daily fares, collected at 11 a.m.;
- Italo NTV:
    - 12 daily fares, collected at 11 a.m.

A screenshot was taken for each element of the defined sample. The captured screenshots are stored in a data-base in relation to:

- Day of trip;
- Day of ticket purchasing;
- Company.

The data-base is organized with a three-dimensional matrix, with axis given by the precedent variables, where the single element is a matrix of values given by the screenshot. Fig. 3.4 shows an example of a screenshot taken for each of the two transport companies considered.

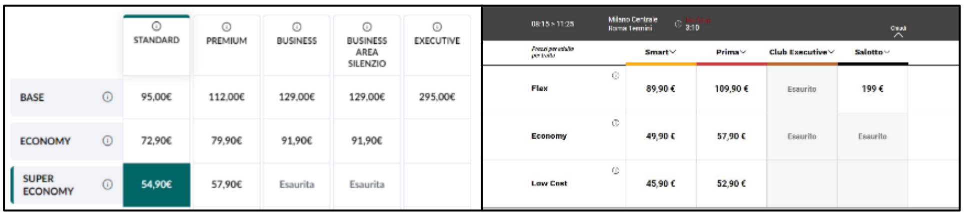


Fig. 3.4. Example of screenshot (Trenitalia and Italo NTV companies)

3.1.2. Survey: year 2025

The paragraph concerns the survey of HSR ticket prices starting from the consultation of the websites of the two HSR transport companies. The characteristics detected are reported below.

Temporal reference

- Year:
  - 2025.
- Days of trip (t):
  - 11/06;
  - 12/06;
  - 17/06;
  - 18/06;
  - 24/06;
  - 25/06.
- Days of ticket purchasing (p):

Tab.3.2. Survey period for run choice modelling: year 2025

Set of days of ticket purchasing (p)	Day of trip (t)
06/06-10/06	11/06
06/08-11/06	12/06
06/08-16/06	17/06
06/08-17/06	18/06
06/08-18/06	24/06
06/08-18/06	25/06

Characteristics of the service

- Mode/service: High Speed Rail (HSR);
- Companies:

- Trenitalia, services “Frecciarossa”;
- NTV, services “Italo”.

- Analyzed relationship:
  - Rome-Milan;
  - Milan-Rome.

Set of runs analyzed:

- Time slot: 05:00-10:00;
- Maximum number of intermediate stops: 1.

Set of fares considered

- Trenitalia Frecciarossa:
  - 21 daily fares, collected at 11 a.m.;
- NTV Italo:
  - daily fares, collected at 11 a.m.

A screenshot was taken for each element of the defined sample. The captured screenshots are stored in a data-base in relation to:

- Day of trip;
- Day of ticket purchasing;
- Company.

The data-base is organized with a three-dimensional matrix, with axis given by the precedent variables, where the single element is a screenshot. Fig. 3.5 shows an example of a screenshot, for each of the two transport companies considered.

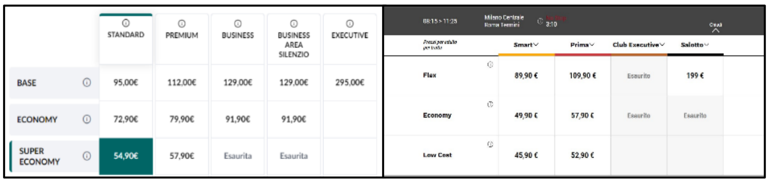


Fig. 3.5. Example of screenshot (Trenitalia and Italo NTV companies)

3.2. Survey and data-base building for fare supply modelling

The paragraph concerns the survey of the HSR ticket fares starting from viewing of the two HSR transport companies' websites. The variables described in Fig. 2.1, namely the day of trip (t) and the day of ticket purchasing (p) are reported in Fig. 3.6.

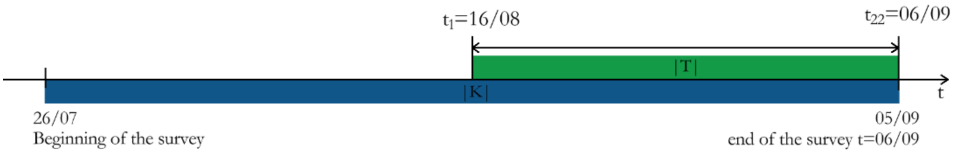


Fig. 3.6. Temporal characteristics: aggregated survey

Temporal reference
- 2022;
- 2023.

- Days of trip (t):
  - From 16/8 to 6/9 (22 days);
- Days of ticket purchasing (Tab. 3.3):

Tab.3.3. Survey period for fare supply modelling

Set of days of ticket purchasing (p)	Day of trip (t)
26/07-15/08	16/08
26/07-16/08	17/08
26/07-17/08	18/08
26/07-18/08	19/08
26/07-19/08	20/08
26/07-20/08	21/08
26/07-21/08	22/08
26/07-22/08	23/08
26/07-23/08	24/08
26/07-24/08	25/08
26/07-25/08	26/08
26/07-26/08	27/08
26/07-27/08	28/08
26/07-28/08	29/08
26/07-29/08	30/08
26/07-30/08	31/08
26/07-31/08	01/09
26/07-01/09	02/09
26/07-02/09	03/09
26/07-03/09	04/09
26/07-04/09	05/09
26/07-05/09	06/09

Characteristics of the service:

- Mode/service: High Speed Rail (HSR);
- Companies:
    - Trenitalia, services “Frecciarossa”;
    - NTV, services “Italo”.
- Relationship:
    - Rome-Milan;
    - Milan-Rome.
- Set of runs analyzed (1 run for each of the following time slot):
    - 06:00-9:00;
    - 12:00-14:00;
    - 19:00-21:00.

Set of fare considered:

- Trenitalia, Frecciarossa:
    - “Super Economy-Standard”.
- NTV, Italo:
    - “Low Cost-Smart”.

If the fare to be collected is not available, the “closest” fare is collected according to the sequence presented in Tab. 3.4.

Tab.3.4. List and coding of fares supplied

Trenitalia		Cod
1	“Super Economy-Standard”	SE_1
2	“Super Economy-Standard Area Silenzio”	SE_2
3	“Super Economy-Premium”	SE_3
4	“Super Economy-Business”	SE_4
5	“Super Economy-Business Area Silenzio”	SE_5
6	“Super Economy-Working Area”	SE_6
7	“Super Economy-Executive”	SE_7
8	“Economy-Standard”	EC_1
9	“Economy-Standard Area Silenzio”	EC_2
10	“Economy-Premium”	EC_3
11	“Economy-Business”	EC_4
12	“Economy-Business Area Silenzio”	EC_5
13	“Economy-Working Area”	EC_6
14	“Economy-Executive”	EC_7
15	“Base-Standard”	BA_1
16	“Base-Standard Area Silenzio”	BA_2
17	“Base-Premium”	BA_3
18	“Base-Business”	BA_4
19	“Base-Business Area Silenzio”	BA_5
20	“Base-Working Area”	BA_6
21	“Base-Executive”	BA_7
Italo NTV		
22	“Low Cost-Smart”	LC_1
23	“Low Cost-Prima”	LC_2
24	“Low Cost-Club Executive”	LC_3
25	“Low Cost-Salotto”	LC_4
26	“Economy-Smart”	EC_1
27	“Economy-Prima”	EC_2
28	“Economy-Club Executive”	EC_3
29	“Economy-Salotto”	EC_4
30	“Flex-Smart”	FL_1
31	“Flex-Prima”	FL_2
32	“Flex-Club Executive”	FL_3
33	“Flex-Salotto”	FL_4

The data-base is then organized into a three-dimensional matrix:

- Day of trip(t);
- Day of ticket purchasing (p);
- Run-company (r).

The single element of the matrix is composed of the collected price and the type of fare.

4.
Run choice modelling

The chapter regards the run choice modelling, thus are specified the steps regarding the proposed framework, which is composed of three main parts which identify the chosen run by the user. Therefore, are specified the general model for the run choice, with the specification of the likelihood and the estimator, and all the models calibrated in the following with the calculation of the descriptive statistics of the sample. It is important to underline that the calibrations are carried out on the basis of the extracted samples described in chapter 3. Moreover, several hypotheses regarding the choice set building are specified and used for the calibration of the parameters.

The chapter is organized as follows:

- identification of the chosen run (par. 4.1);
- run choice model: specification (par. 4.2);
- run choice model: calibration (par. 4.3).

4.1. Identification of the chosen run

This paragraph presents the part of the proposed framework related to the identification of the run chosen by a user. The hypothesis considered is that the users' choices can be identified in relation to the day-to-day ticket evolution observed between two consecutive days of ticket purchasing. The day-to-day ticket evolution potentially contains a huge quantity of information that needs a specific collection method.

The first part of the framework is composed of the three following steps (Fig. 4.1).

Step 1. Ticket coding: in which every fare structure of each 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.

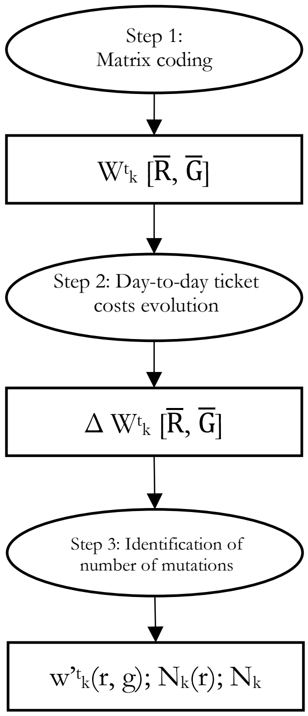


Fig. 4.1. Steps of the framework for the identification of chosen run

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 step 1 is a matrix fares W_k [R, G] with $k \in K$ (Tab 4.1)

Tab.4.1. Example of general matrix of fares

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1234	80€	90€	90€	95€	95€	100€	100€	100€	110€	110€	110€	110€	120€	120€	130€

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:

$$\Delta W_k[R, G] = W_k[R, G] - W_{k+1}[R, G] \qquad (4.1)$$

obtained as difference between the matrix of ticket costs, $W_k[R, G]$ at day k and the matrix of ticket costs $W_{k+1}[R, G]$ at previous day k+1.

The generic element of the above matrix $\Delta W_k[R, G]$ 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, $w_{k+1}(r, g)$.

$$\Delta w_k(r, g) = w_k(r, g) - w_{k+1}(r, g) \qquad (4.2)$$

The output of Step 2 is a matrix fares ΔW_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.

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

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

$$w'_k(r, g) = \begin{cases} 0, \text{ if } \Delta w_k(r, g) = 0 \\ 1, \text{ if } \Delta w_k(r, g) \neq 0 \end{cases} \qquad (4.3)$$

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

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

$$N_k(r) = \sum_g w'_k(r, g) \qquad (4.4)$$

with $N_k(r)$, number of ticket changes associated to run r (ϵ 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:

$$N_k = \sum_r N_k(r) = \sum_r \sum_g w'_k(r, g) \qquad (4.5)$$

4.2. Run choice model: specification

This paragraph presents the second part of the proposed framework related to the run choice modelling, which is composed of the two following steps (Fig. 4.2), as follows.

Step 4: Choice set building (par. 4.2.1).

Step 5: Run choice modelling (4.2.2) and Likelihood formulation (4.2.3).

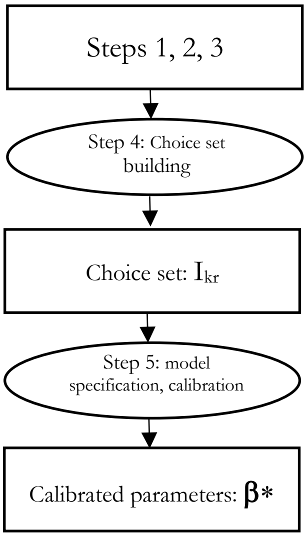


Fig. 4.2. Steps of the framework for the run choice modelling

4.2.1. Choice set building

The paragraph deals with the construction of the choice set. Four hypotheses are considered below regarding the construction of the choice set.

<u>Hypothesis 1.</u> This case considers, without affecting generality, the choice set of runs generated taking into account the Desired Arrival Time (DAT) of user at destination.

Given the generic run choice $w'_k(r, g)=1$, the choice set of the user that chooses the run r at 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.

Taking into account 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. 4.3):

$$DAT = \delta_z : \min_z |\tau_r - \delta_z| \tag{4.6}$$

where τ_r is the expected arrival time of the chosen run r.

It is assumed that the choice set is composed of 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 assumption 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:

Case 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;

Case 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.

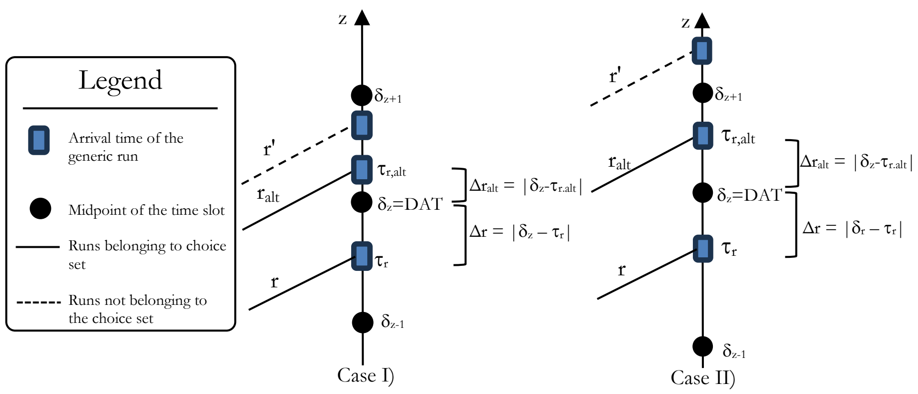


Fig. 4.3. Examples of runs of the choice set: hypothesis 1

Hypothesis 2. The second hypothesis relies upon the same considerations about DAT (hypothesis 1); however, the alternative run (r_{alt}) is identified according to the following condition:

$$\Delta r_{alt} \geq \Delta r \quad (4.7)$$

where $\Delta r_{alt} = |\tau_{ralt} - \delta_z|$ is the penalty, or time interval between the expected arrival time of the alternative run r_{alt}, τ_{ralt}, and the average point of the time period, δ_z. The condition of eq. (4.7) ensures that the penalty, Δr, associated to the chosen run is smaller than the penalty, Δr_{alt}, associated to the alternative run.

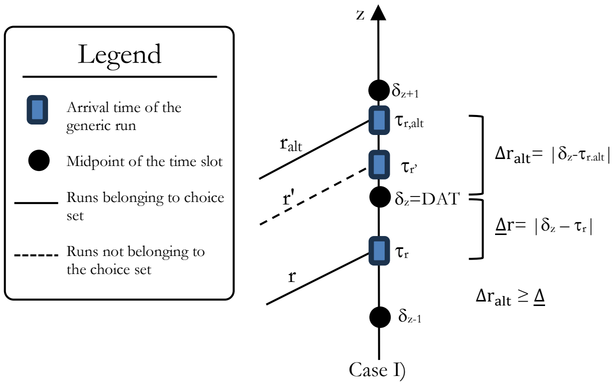


Fig. 4.4. Examples of runs of the choice set: hypothesis 2

Hypothesis 3. The third hypothesis relies upon the same considerations about DAT (hypothesis 1), but in this case the choice set is composed by three alternatives. The chosen run is the one associated to the fare variation; the first alternative run is on the opposite side from the chosen run and closest to the DAT; the second alternative run is on the side of and immediately after the chosen run respect to the DAT.

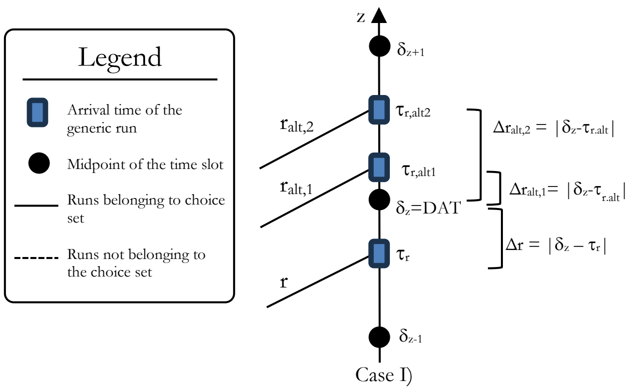


Fig. 4.5. Examples of runs of the choice set: hypothesis 3

Hypothesis 4. The fourth hypothesis relies upon the same considerations about DAT (hypothesis 1), though the alternative run (r_{alt}) is identified according to the following condition:

$$\Delta r_{alt} \geq \underline{\Delta} \tag{4.8}$$

where $\underline{\Delta}$ is a given threshold (e.g. 0 minutes, 5 minutes, 10 minutes, ….). The condition of eq. (4.8) ensures that no alternative runs having a penalty greater than $\underline{\Delta}$ could belong to the choice selected.

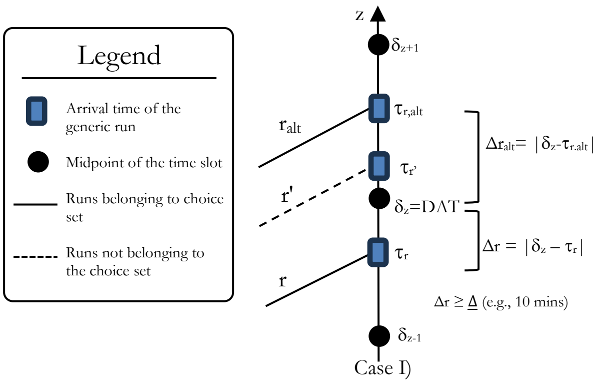


Fig. 4.6. Examples of runs of the choice set: hypothesis 4

4.2.2. Run choice model

The specified choice models rely on random utility theory (McFadden, 1972; Ben-Akiva & 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^i_r.

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. Whereas, it is possible to express the probability that the run $r \in I_{kr}$, will be chosen by user i as the probability, p^i_r, that the perceived utility of run r, U^i_r, 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:

$$p^i_r(\boldsymbol{\beta}) = \text{Prob}\,(U^i_r(\boldsymbol{\beta}) > U^i_{r'}(\boldsymbol{\beta})) \quad \forall\ r' \neq r,\ r \in I_{kr} \tag{4.9}$$

Model (4.9) may be differently specified. The specification considered here is the binomial Logit (descending from the generation of $I_{r,k}$ according to criteria a):

$$p^i_r(\boldsymbol{\beta}) = \exp(V^i_r(\boldsymbol{\beta})) / (1 + \exp(V^i_r(\boldsymbol{\beta}) - V^i_{r'}(\boldsymbol{\beta}))) \qquad r', r \in I_{kr} \qquad (4.10)$$

where:

$V^i_r(\boldsymbol{\beta}) = E[U^i_r] / \theta = \boldsymbol{\beta}^T \bullet \mathbf{x}$

$E[U^i_r]$, 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.

4.2.3. Likelihood function

The paragraph deals with the specification of the likelihood function. The calibration allows to obtain the estimation of the 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 $p^i_r(\beta)$ depends on the vector β, also the probability L of observing the entire sample is a function of the unknown parameters:

$$L(\boldsymbol{\beta})=\prod_{i=1....n} p^i_r(\boldsymbol{\beta}) \qquad (4.11)$$

The estimate of maximum likelihood, $\boldsymbol{\beta}^*$, of vector of parameters, $\boldsymbol{\beta}$, is obtained by means of eq. (4.12):

$$\boldsymbol{\beta}^* = \max L(\boldsymbol{\beta}) = \max \prod_{i=1....n} p^i_r(\beta) \qquad (4.12)$$

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

- <u>Likelihood function L for all days k ∈ K</u>

For the generic day k, recalling eq. (4.5) that defines the number of mutations for each run r, the value of L_k is equal to:

$$L_k = \prod_{r \in R_k} p_{kr}^{N_k(r)} \qquad (4.13)$$

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. (4.13) and the overall L function, for all days k∈K, is given by:

$$L = \prod_{k \in K} L_k = \prod_{k \in K} \prod_{r \in R_k} {p_{kr}}^{N_k(r)} \qquad (4.14)$$

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

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:

$$L_{kr}(g) = \prod_{g \in G} p_{kr}\ (g) \qquad \forall g\text{: } w'(r,g) = 1 \qquad (4.15)$$

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:

$$L_k = \prod_{r \in R_k} L_{kr}(g) = \prod_{r \in R_k} \prod_{g \in G} p_{kr}\ (g) \qquad (4.16)$$

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

$$L = \prod_{k \in K} L_k = \prod_{k \in K} \prod_{r \in R_k} \prod_{g \in G} p_{kr}\ (g) \qquad (4.17)$$

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 and the class of fare is variable. The second could be the case where both the run and the class of the fare are variable. In a more general way, it is possible to consider also the variation of day of trip t.

Fig. 4.4 summarizes the research contributions of the thesis in the process of choice set building and of the likelihood specification. The first contribution concerns the identification of number of mutations (paragraph 4.1). The second contribution concerns the specification of the run choice model, in terms of choice set building based on DAT (paragraph 4.2.1), specification of the likelihood function (paragraph 4.2.3).

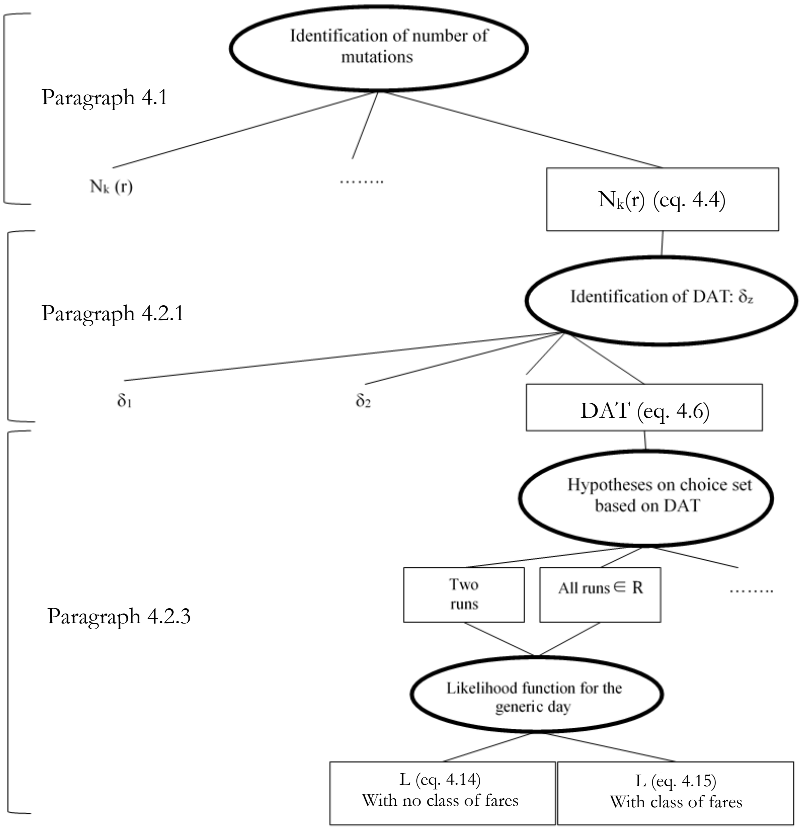


Fig. 4.7. Building process of the choice set and of the likelihood function

4.3. Run choice model: calibration

This paragraph contains the description of the calibration process that has been carried out. The attributes specified in the run choice models are described below:

- Fare, x_w [€], ticket cost associated with a run (corresponding to the day k+1 and to the fare number of the fare matrix) with a variation between day k and day k+1;

- Penalty, x_{Δ}[h], defined as the time interval between the DAT (in general δ_z for the alternative run), and the arrival time of the run r at the station τ_r, calculated with the following equation:

$$\Delta = |\tau_r - \delta_z| \quad [h] \tag{4.18}$$

- Time, x_t [h], $t = h_{arrival} - h_{departure}$ [h], travel time on board from origin to destination, calculated as the difference between the time of arrival and the time of departure at stations;
- Stops, x_s, number of intermediate stops between departure and arrival stations;
- Company, x_{co}, HSR company of the chosen run and the alternative run (0=Trenitalia, 1=Italo NTV).

The parameters specified in the run choice models are described below:

- β_0, constant;
- β_w, parameter associated with the attribute x_w;
- β_t, parameter associated with the attribute x_t;
- β_{Δ}, parameter associated with the attribute x_{Δ};
- β_s, parameter associated with the attribute x_s;
- β_{co}, parameter associated with the attribute x_{co}.

The set of run choice models specified is described in Tab. 4.2.

Tab.4.2. Models specification

Model	Average utility (V_r^i)
1	$\beta_w \cdot x_w$
2	$\beta_w \cdot x_w + \beta_t \cdot x_t$
3	$\beta_w \cdot x_w + \beta_{co} \cdot x_{co}$
4	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_{\Delta} \cdot x_{\Delta}$
5	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_s \cdot x_s$
6	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_{co} \cdot x_{co}$
7	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_D \cdot x_D + \beta_{co} \cdot x_{co}$
8	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_s \cdot x_s + \beta_{co} \cdot x_{co}$
9	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_D \cdot x_D + \beta_s \cdot x_s$
10	$\beta_w \cdot x_w + \beta_t \cdot x_t + \beta_D \cdot x_D + \beta_s \cdot x_s + \beta_{co} \cdot x_{co}$
11	$\beta_0 + \beta_w \cdot x_w$
12	$\beta_0 + \beta_w \cdot x_w + \beta_t \cdot x_t$

The models were calibrated on the basis of a combination of the following:

- years of the survey: 2022, 2023, 2025;
- dimension of sample: minimum prototype, median prototype, full sample;
- hypothesis on the choice set building: 1, 2, 3, 4.

Tab. 4.3 reports the calibration framework carried out by the combination of the above-mentioned elements. The models specified according to the hypothesis 1 of choice set building were calibrated by using:

- the minimum prototype sample of the year 2023 (par 4.3.1, part a);
- the median prototype sample of the year 2023 (par. 4.3.1, part b);
- the full samples of the years 2022 (par. 4.3.1, part c), 2023 (par. 4.3.1, part d), 2025 (par. 4.3.1, part e).
- the aggregation of the full samples of the years 2023 and 2025 (par. 4.3.1, part g);
- the aggregation of the full samples of the years 2022, 2023 and 2025 (par. 4.3.1, part f).

The models specified according to the hypothesis 2 of choice set building were calibrated by using the full sample of the year 2025 (par. 4.3.2).

The models specified according to hypothesis 3 of choice set building were calibrated by using the full sample of the year 2025 (par. 4.3.3).

The models specified according to the hypothesis 4 of choice set building were calibrated by using the full sample of the year 2025 (par. 4.3.4).

Tab.4.3. Calibration framework

Hypothesis		2022	2023	2025	2023+ 2025	2022+ 2023+2025
1	Minimum prototype		4.3.1a			
1	Median prototype		4.3.1b			
1	Full sample	4.3.1c	4.3.1d	4.3.1e	4.3.1g	4.3.1f
2	Full sample			4.3.2		
3	Full sample			4.3.3		
4	Full sample			4.3.4		

4.3.1. Calibrated parameters: hypothesis 1

The paragraph presents an example concerning the construction of the choice set in hypothesis 1.

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. 4.5). The time axis z is discretized into time intervals of 30 minutes (11:00; 11:30; 12:00;) and the δ_z are identified as the midpoints of each time interval (δ_{z-1}= 11:15; δ_z= 11:45; δ_{z+1}= 12:15). 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 t_r of each run. In the example presented, the chosen run is r, with a corresponding τ_r=11:34, and the resulting DAT is δ_z=11:45. The value of penalty is $\Delta=|\delta_z - \tau_r|$ = 11:45-11:34 = 11 minutes. 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_{alt} with $\tau_{r,alt}$=10:52 and the penalty Δ_{alt} = 7 minutes. The choice set is I_{rk}={9618, 9622}.

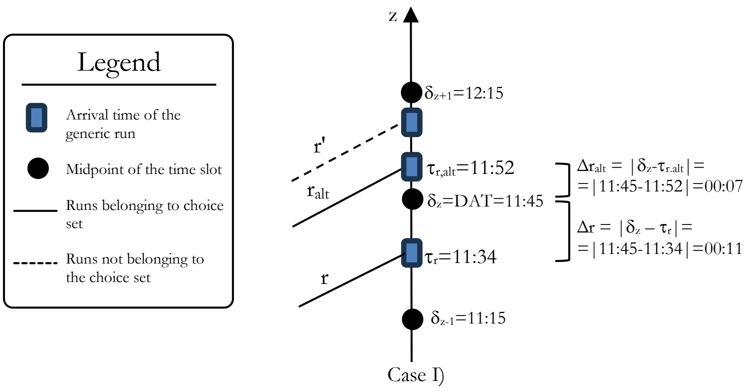


Fig. 4.8. Example of runs of the choice set: hypothesis 1

a. Minimum prototype sample: year 2023

The paragraph reports the calibration of the parameters of the run choice models specified in Tab. 4.2 according to hypothesis 1 of choice set building and based upon the data extracted during the survey of year 2023.

The minimum prototype sample has the following characteristics:

- $\bar{t}$=01/09/23

- K= {21/8/23, …,25/8/23}
- $\overline{K}$=4
- Sample dimension: 30 users

Tab. 4.4 reports a description of the minimum prototype sample composed of 30 users by considering the four days of K and the selected runs.

Tab.4.4. Minimum prototype sample

	k-2	k-1	k	k+1	
r	$N_k(r)$	$N_k(r)$	$N_k(r)$	$N_k(r)$	
9606	0	2	4	4	
9608	0	0	3	4	
9612	0	1	2	2	
9618	0	0	0	0	
9620	0	2	1	2	
9622	0	0	0	0	
9968	0	0	0	0	
9972	0	0	1	2	
N_k	0	5	11	14	30

Tab. 4.5 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the minimum prototype sample: minimum, maximum, average values and standard deviation.

Tab.4.5. Descriptive statistics of attributes: minimum prototype sample

Attribute (**x**)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	$[0; +\infty]$	2.98	3.98	3.33	0.36
Fare (x_w)	€	$[0; +\infty]$	49.90	100.90	70.58	11.85
Company (x_{co})		[0;1]	0	1	0.13	0.34
Stops (x_s)		[0;1]	0	1	0.72	0.45
Penalty (x_Δ)	h	$[-\infty; +\infty]$	0.03	0.97	0.47	0.33

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.6 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that model 9 has the corrected signs of the calibrated parameters. The values of the t-student statistic are not significant.

Tab.4.6. Calibrated parameters: minimum prototype sample

Model	β_w	β_t	β_{co}	β_Δ	β_s	$L(\boldsymbol{\beta})$
1	0.0311					-20.465
(t-value)	(0.7946)					
4	0.0298	0.6216		-2.3860		-14.211
(t-value)	(0.5709)	(0.6995)		(-2.6900)		
9	-0.114	-14.187		-18.415	-24.360	-2.0108e-07
(t-value)	(-0.0004)	(-0.0014)		(-0.0053)	(-0.0061)	
10	0.0569	-25.5690	10.6200	-26.6380	-29.3750	-1.6473e-07
(t-value)	(-0.0001)	(-0.0027)	(0.009)	(-0.0046)	(-0.0054)	

b. Median prototype sample: year 2023

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building.

The median prototype sample has the following characteristics:

- $\bar{t}$=01/09/23
- K= {01/08/2023, …, 31/08/2023}
- K=30
- Sample dimension: 141 users

Tab. 4.7 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the median prototype sample: minimum, maximum, average value and standard deviation.

Tab.4.7. Descriptive statistics of attributes: median prototype sample

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.98	3.98	3.33	0.36
Fare (x_w)	€	[0; +∞]	44.90	249.00	87.77	26.12
Company (x_{co})		[0;1]	0	1	0.12	0.32
Stops (x_s)		[0;1]	0	1	0.69	0.46
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.10	0.47	0.30	0.16
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.10	0.47	0.30	0.15

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.8 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that all models (except for model 7) have the corrected signs of the calibrated parameters. Some values of the t-student statistic are not significant.

Tab.4.8. Calibrated parameters: median prototype sample

Model	β_0	β_w	β_t	β_Δ	β_s	$L(\boldsymbol{\beta})$
1		-0.0237				-96.944
(t-value)		(-1.0575)				
2		-0.0289	-0.4898			-95.576
(t-value)		(-1.1348)	(-1.6353)			
4		-0.0242	-0.7188	-0.9304		-94.972
(t-value)		(-1.0132)	(-1.9632)	(-1.0885)		
5		-0.0327	-1.1391		-0.7743	-91.899
(t-value)		(-1.1662)	(-2.8773)		(-2.5915)	
7		-0.0813	-1.1808	4.8215	-2.1348	-88.834
(t-value)		(-1.9917)	(-2.8503)	(2.2795)	(-3.0881)	
11	-0.1061	-0.0248				-96.749
(t-value)	(-0.6240)	(-1.0632)				
12	-0.2073	-0.03411	-0.5939			-94.916
(t-value)	(-1.1408)	(-1.2116)	(0.3142)			

c. Full sample: year 2022

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building.

The full prototype sample of the year 2022 has the following characteristics:

- T= {01/09/22; 08/09/22}
- K^{t1}= {01/09, 07/08, …., 31/08}
- $\overline{K}^{t_1}$=30
- K^{t2}= {01/09, 07/08, …., 07/09}
- $\overline{K}^{t_2}$=38

Tab. 4.9 reports the number of users identified per day, direction and company belonging to the full sample of year 2022.

Tab.4.9. Number of users identified: full sample of year 2022

t	Direction	Company	Numbers of users identified (mutations)
01/09	MI-RO	Trenitalia	160
		Italo NTV	27
01/09	RO-MI	Trenitalia	126
		Italo NTV	21
08/09	MI-RO	Trenitalia	111
		Italo NTV	21
08/09	MI-RO	Trenitalia	109
		Italo NTV	16
		Total	591

Tab. 4.7 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users

belonging to the full sample composed of 591 users: minimum, maximum, average values and standard deviation.

Tab.4.10. Descriptive statistics of attributes: sample of year 2022

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.98	3.98	3.20	0.25
Fare (x_w)	€	[0; +∞]	45.90	195.90	77.67	17.62
Company (x_{co})		[0;1]	0	1	0.19	0.39
Stops (x_s)		[0;1]	0	1	0.77	0.42
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.07	0.50	0.28	0.14
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.07	0.47	0.21	0.14

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.11 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that the fare, time and penalty parameters of the presented models have never corrected signs.

Tab.4.11. Calibrated parameters: full sample of year 2022

Model	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1	0.0385					-403.27
(t-value)	(3.2886)					
2	0.0445	0.5906				-400.09
(t-value)	(3.6645)	(2.4842)				
3	0.0181		-0.4092			-401.26
(t-value)	(1.2919)		(-3.1899)			
4	0.0818	1.6051		5.1500		-341.38
(t-value)	(5.4342)	(6.0533)		(9.4409)		
5	0.0805	1.2282			0.7652	-388.02
(t-value)	(5.1846)	(4.4818)			(4.7364)	
6	0.0197	0.6751	-0.4996			-397.22
(t-value)	(1.3710)	(2.7885)	(-2.4314)			
7	0.0255	1.8488	-1.1003	5.5795		-331.29
(t-value)	(1.5346)	(6.7563)	(-4.5251)	(9.9660)		
8	0.0283	1.8992	-1.4171		1.2883	-373.55
(t-value)	(1.7048)	(5.9238)	(-5.1291)		(6.1777)	
9	0.0452	1.0592		7.5130	-1.1398	-331.53
(t-value)	(2.8932)	(3.7198)		(9.4389)	(-4.3393)	
10	0.0242	1.3961	-0.6889	6.8646	-0.7248	-328.68
(t-value)	(1.4991)	(4.2791)	(-2.3601)	(8.4052)	(-2.2880)	

d. Full sample: year 2023

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building. The full prototype sample of year 2023 has the following characteristics:

- T= {01/09/23; 08/09/23}
- K^{t1}= {01/09, 07/08, …., 31/08}
- $\overline{K}^{t_1}$=30
- K^{t2}= {01/09, 07/08, …., 07/09}
- $\overline{K}^{t_2}$=38

Tab. 4.12 reports the number of users identified per day, direction and company belonging to the full sample of year 2022 composed of 592 users.

Tab.4.12. Number of users identified: full sample of year 2023

t	Direction	Company	Numbers of users identified (mutations)
01/09	MI-RO	Trenitalia	124
		Italo NTV	42
01/09	RO-MI	Trenitalia	133
		Italo NTV	8
08/09	MI-RO	Trenitalia	114
		Italo NTV	45
08/09	MI-RO	Trenitalia	102
		Italo NTV	24
		Total	592

Tab. 4.13 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the full sample of year 2023 composed of 592 users: minimum, maximum, average value and standard deviation.

Tab.4.13. Descriptive statistics of attributes: full sample of year 2023

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.98	3.98	3.22	0.26
Fare (x_w)	€	[0; +∞]	39.90	295.00	89.21	26.79
Company (x_{co})		[0;1]	0	1	0.22	0.41
Stops (x_s)		[0;1]	0	1	0.77	0.42
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.07	0.50	0.27	0.15
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.07	0.47	0.22	0.14

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.14 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1, 2, 3, 5 have corrected signs of calibrated parameters. The values of the t-student statistic are generally not significant.

Tab.4.14. Calibrated parameters: full sample of year 2023

Model	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1	-0.0242					-405.01
(t-value)	(-2.8534)					
2	-0.0487	-0.1138				-404.89
(t-value)	(-3.6645)	(-0.5007)				
3	-0.0247		-0.6001			-399.38
(t-value)	(-2.8670)		(-3.1899)			
4	-0.0218	0.3236		2.3254		-389.31
(t-value)	(-2.6538)	(1.3247)		(5.3650)		
5	-0.0249	-0.1206			-0.0106	-404.89
(t-value)	(-2.7420)	(-0.4960)			(-0.0786)	
6	-0.0508	2.2654	-0.6997			-382.88
(t-value)	(-3.7314)	(5.5923)	(-3.6019)			
7	-0.0492	0.4090	-0.8692	0.3510		-397.07
(t-value)	(-3.6875)	(1.4493)	(-3.7895)	(2.1004)		
8	-0.0526	0.6665	-0.8433		2.6817	-379.55
(t-value)	(-3.8160)	(2.5812)	(-4.1387)		(6.0160)	
9	-0.0474	-0.0014		5.0093	-1.2192	-372.14
(t-value)	(-3.9896)	(-0.0054)		(7.4587)	(-5.5490)	
10	-0.0545	0.1311	-0.2571	4.7409	-1.0587	-371.60
(t-value)	(-3.8829)	(0.4573)	(-1.0258)	(6.6375)	(-3.9297)	

e. Full sample: year 2025

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building. The full prototype sample of year 2025 has the following characteristics:

- T= {11/06/25; 12/06/25; 17/06/2025; 18/06/2025; 24/06/2025; 25/06/2025}
- K^{t1}= {06/06, 07/06, …., 10/06}
- $\bar{K}^{t_1}=5$
- K^{t2}= {06/06, 07/06, …., 11/06}
- $\bar{K}^{t_2}=6$
- K^{t3}= {06/06, 07/06, …., 16/06}
- $\bar{K}^{t_3}=11$
- K^{t4}= {06/06, 07/06, …., 17/06}
- $\bar{K}^{t_4}=12$
- K^{t5}= {06/06, 07/06, …., 18/06}
- $\bar{K}^{t_5}=13$

- K^{t_6}= {06/06, 07/06, …., 18/06}
- $\overline{K}^{t_6}$=13
- Sample dimension: 635 users

Tab. 4.15 reports the number of users identified per day, direction of trip and company belonging to the full sample of year 2025 composed of 635 users.

Tab.4.15. Number of users identified: full sample of year 2025

t	Direction	Company	Numbers of users identified (mutations)
11/06	MI-RO	Trenitalia, Italo NTV	26
11/06	RO-MI	Trenitalia, Italo NTV	25
12/06	MI-RO	Trenitalia, Italo NTV	26
12/06	RO-MI	Trenitalia, Italo NTV	39
17/06	MI-RO	Trenitalia, Italo NTV	46
17/06	RO-MI	Trenitalia, Italo NTV	108
18/06	MI-RO	Trenitalia, Italo NTV	64
18/06	RO-MI	Trenitalia, Italo NTV	101
24/06	MI-RO	Trenitalia, Italo NTV	50
24/06	RO-MI	Trenitalia, Italo NTV	64
25/06	MI-RO	Trenitalia, Italo NTV	31
25/06	RO-MI	Trenitalia, Italo NTV	45
		Total	635

Tab. 4.16 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the full sample of year 2025 composed of 635 users: minimum, maximum, average value and standard deviation.

Tab.4.16. Descriptive statistics of attributes: full sample of year 2025

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	$[0; +\infty]$	2.87	3.67	3.14	0.14
Fare (x_w)	€	$[0; +\infty]$	49.90	295.00	102.92	23.66
Company (x_{co})		[0;1]	0	1	0.30	0.46
Stops (x_s)		[0;1]	0	1	0.69	0.46
Penalty (chosen run) (x_Δ)	h	$[-\infty; +\infty]$	0.03	0.50	0.19	0.11
Penalty (alternative run) (x_Δ)	h	$[-\infty; +\infty]$	0.03	0.50	0.17	0.09

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.17 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1 and 2 have corrected signs of calibrated parameters. The values of the t-student statistic are generally not significant.

Tab.4.17. Calibrated parameters: full sample of year 2025

Model	β_w	β_t	β_{co}	β_Δ	β_s	$L(\boldsymbol{\beta})$
1	-0.0074					-439.48
(t-value)	(-1.1071)					
2	-0.0073	0.1659				-426.97
(t-value)	(-1.0991)	0.4107				
3	-0.0027		-0.6261			-439.40
(t-value)	(-0.4197)		(-4.8864)			
4	-0.0697	0.3222	2.2571			-431.46
(t-value)	(-1.0282)	(0.7848)	(3.8670)			
5	-0.0079	-0.8351		0.3926		-435.01
(t-value)	(-1.1553)	(-1.5610)		(-2.9415)		
6	-0.0018	0.8427			-0.7021	-425.03
(t-value)	(-0.2907)	(1.9587)			(-5.2376)	
7	-0.0004	1.1699	2.9465		-0.8456	-412.47
(t-value)	(-0.0682)	(2.6346)	(4.8307)		(-5.9232)	
8	0.00002	-0.0963		0.0868	-1.0773	-408.19
(t-value)	(0.0028)	(-1.7833)		(5.6738)	(-7.0256)	
9	-0.0076	-0.8476	2.5659	0.4779		-425.20
(t-value)	(-1.0891)	(-1.5803)	(4.2635)	(3.5033)		
10	0.0028	-0.9857	3.9939	1.1208	-1.3903	-387.07
(t-value)	(0.4210)	(-1.8181)	(6.1602)	(6.8701)	(-8.0830)	

f. Full sample: years 2022, 2023 and 2025

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building. The full prototype sample obtained by the aggregation of the full samples of years 2022, 2023 and 2025 has the following characteristics:

- T= {01/09/22; 08/09/22; 01/09/23; 08/09/23; 11/06/25; 12/06/25; 17/06/2025; 18/06/2025; 24/06/2025; 25/06/2025}
- K^{t1}= {01/09, 07/08, …., 31/08}
- $\overline{K}^{t_1}=30$
- K^{t2}= {01/09, 07/08, …., 07/09}
- $\overline{K}^{t_2}=38$
- K^{t3}= {01/09, 07/08, …., 31/08}
- $\overline{K}^{t_3}=30$
- K^{t4}= {01/09, 07/08, …., 07/09}
- $\overline{K}^{t_4}=38$
- K^{t5}= {06/06, 07/06, …., 10/06}
- $\overline{K}^{t_5}=5$
- K^{t6}= {06/06, 07/06, …., 11/06}
- $\overline{K}^{t_6}=6$
- K^{t7}= {06/06, 07/06, …., 16/06}
- $\overline{K}^{t_7}=11$

- K^{t8}= {06/06, 07/06, …., 17/06}
- $\overline{K}^{t_8}$=12
- K^{t9}= {06/06, 07/06, …., 18/06}
- $\overline{K}^{t_9}$=13
- K^{t10}= {06/06, 07/06, …., 18/06}
- $\overline{K}^{t_{10}}$=13
- Sample dimension: 1818 users

Tab. 4.18 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the complete sample composed of 1818 users obtained by the aggregation of the full samples of the years 2022, 2023 and 2025: minimum, maximum, average value and standard deviation.

Tab.4.18. Descriptive statistics of attributes: complete sample (hypothesis 1)

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.87	3.67	3.14	0.14
Fare (x_w)	€	[0; +∞]	49.90	295.00	102.92	23.66
Company (x_{co})		[0;1]	0	1	0.30	0.46
Stops (x_s)		[0;1]	0	1	0.69	0.46
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.03	0.50	0.19	0.11
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.03	0.50	0.17	0.09

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.19 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1 and 2 have corrected signs of calibrated parameters with significant values of t-student statistics (except for one parameter).

Tab.4.19. Calibrated parameters: complete sample

Model	β_0	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1		-0.0051					-1259.3
(t-value)		(-1.2485)					
2		-0.0047	0.1652				-1238.6
(t-value)		(-1.1611)	(1.1096)				
3		-0.0111		-0.5158			-1258.7
(t-value)		(-2.3640)		(-6.3119)			
4		-0.0019	0.6689	3.0578			-1193.2
(t-value)		(-0.4553)	(4.2241)	(10.7767)			
5		-0.0033	0.2051		0.1947		-1254.4
(t-value)		(-0.8051)	(1.3723)		(2.9258)		
6		-0.0107	0.8427			-0.7021	-1234.7
(t-value)		(-2.2811)	(1.9587)			(-5.2376)	
7		-0.0095	1.1397	3.6360		-0.8384	-1147.9
(t-value)		(-1.9742)	(6.7917)	(12.2964)		(-9.1552)	
8		-0.0096	0.7519		0.6096	-0.9683	-1205.6
(t-value)		(-2.0398)	(4.6435)		(7.4589)	(-9.5195)	
9		-0.0029	0.6735	3.3899	-0.1896		-1190.2
(t-value)		(-0.7103)	(4.2687)	(10.6921)	(-2.4365)		
10		-0.0090	1.2150	3.3315	0.2341	-1.3903	-1144.6
(t-value)		(-1.8840)	(7.0918)	(10.5228)	(2.5855)	(-8.0830)	
11	23.2030	2.00E-07					-1.523e-07
(t-value)	(0.0091)	(0.0000)					
12	22.2030	2.00E-07	4.00E-08				-1.523e-07
(t-value)	(0.0090)	(0.0000)	(0.0000)				

g. Full sample: years 2023 and 2025

The paragraph reports the calibration of the parameters of the run choice models specified according to hypothesis 1 of choice set building. The prototype sample obtained by the aggregation of the full samples of years 2023 and 2025 have the following characteristics:

- T= {01/09/23; 08/09/23; 11/06/25; 12/06/25; 17/06/2025; 18/06/2025; 24/06/2025; 25/06/2025}
- K^{t1}= {01/09, 07/08, …., 31/08}
- $\overline{K}^{t_1}$=30
- K^{t2}= {01/09, 07/08, …., 07/09}
- $\overline{K}^{t_2}$=38
- K^{t3}= {06/06, 07/06, …., 10/06}
- $\overline{K}^{t_3}$=5
- K^{t4}= {06/06, 07/06, …., 11/06}
- $\overline{K}^{t_4}$=6
- K^{t5}= {06/06, 07/06, …., 16/06}
- $\overline{K}^{t_5}$=11

- K^{t6}= {06/06, 07/06, …., 17/06}
- $\overline{K}^{t_6}$=12
- K^{t7}= {06/06, 07/06, …., 18/06}
- $\overline{K}^{t_7}$=13
- K^{t8}= {06/06, 07/06, …., 18/06}
- $\overline{K}^{t_8}$=13
- Sample dimension: 1227 users

Tab. 4.20 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the complete sample composed of 1227 users obtained by the aggregation of the full samples of the years 2023 and 2025: minimum, maximum, average value and standard deviation.

Tab.4.20. Descriptive statistics of attributes: sample of years 2023 and 2025

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.87	3.98	3.18	0.21
Fare (x_w)	€	[0; +∞]	39.90 €	295.00 €	96.30 €	26.13 €
Company (x_{co})		[0;1]	0	1	0.26	0.44
Stops (x_s)		[0;1]	0	1	0.73	0.44
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.03	0.50	0.23	0.14
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.03	0.50	0.19	0.12

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.21 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1, 2 and 3 have corrected signs of calibrated parameters with significant values of t-student statistics (except for one parameter).

Tab.4.21. Calibrated parameters: sample of years 2023 and 2025

Model	β_0	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1		-0.0151					-845.73
(t-value)		(-2.8217)					
2		-0.0151	-0.0258				-832.94
(t-value)		(-2.8215)	(-0.1310)				
3		-0.0188		-0.4686			-845.72
(t-value)		(-3.1556)		(-4.9773)			
4		-0.0140	0.3374	2.3076			-821.76
(t-value)		(-2.6427)	(1.6320)	(6.6817)			
5		-0.0143	-0.0528		0.1731		-843.28
(t-value)		(-2.7020)	(-0.2667)		(2.2063)		
6		-0.0185	0.2619			-0.5014	-832.13
(t-value)		(-3.1144)	(1.2726)			(-5.1341)	
7		-0.0178	0.7786	2.7455		-0.6632	-799.90
(t-value)		(-2.9702)	(3.5649)	(7.7093)		(-6.4466)	
8		-0.0176	0.3675		0.4987	-0.8023	-817.23
(t-value)		(-2.9497)	(1.7665)		(5.3829)	(-7.0323)	
9		-0.0141	0.3408	2.3215	-0.0099		-821.75
(t-value)		(-2.6429)	(1.6330)	(6.3538)	(-0.0848)		
10		-0.0172	0.7931	2.4169	0.3207	-0.8374	-794.44
(t-value)		(-2.8671)	(3.5968)	(6.5605)	(3.2944)	(-7.1803)	
11	22.2030	4.00E-09					-2.794e-07
(t-value)	(0.0117)	(0.0000)					
12	22.2030	1.00E-07	5.00E-08				-2.794e-07
(t-value)	(0.0117)	(0.0000)	(0.0000)				

4.3.2. Calibrated parameters: hypothesis 2

The paragraph concerns the calibration of the specified run choice models assuming the hypothesis 2 for the choice set building (see par. 4.2.1). An example of choice set building according to hypothesis 2 is reported in Fig. 4.9.

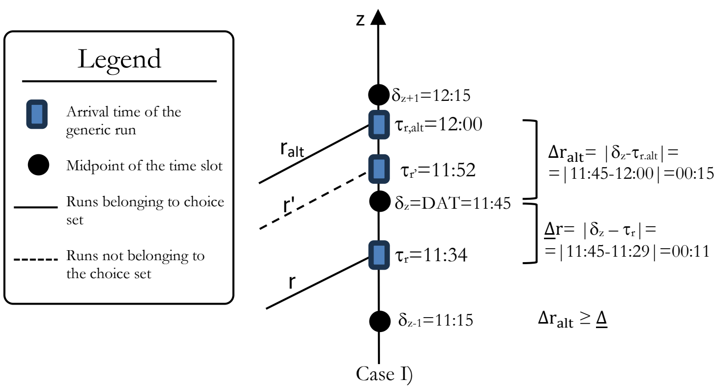


Fig. 4.9. Example of runs of the choice set: hypothesis 2

Tab. 4.18 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the complete sample composed of 635 users belonging to the full sample of year 2025: minimum, maximum, average value and standard deviation.

Tab.4.22. Descriptive statistics of attributes: full sample of year 2025 (hypothesis 2)

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.87	3.67	3.14	0.14
Fare (x_w)	€	[0; +∞]	49.90 €	295.00 €	102.85 €	23.54 €
Company (x_{co})		[0;1]	0	1	0.26	0.44
Stops (x_s)		[0;1]	0	1	0.71	0.45
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.00	1.17	0.14	0.09
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.03	1.17	0.41	0.21

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.23 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1, 2, 3, 4, 7 and 9 have corrected signs of calibrated parameters with significant values of t-student statistics (except for one parameter).

Tab.4.23. Calibrated parameters: full sample of year 2025 (hypothesis 2)

Model	β_0	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1		-0.0069					-439.54
(t-value)		(-1.0615)					
2		-0.0069	-0.1647				-437.91
(t-value)		(-1.0068)	(-0.4001)				
3		-0.0054		-0.2529			-439.46
(t-value)		(-0.8363)		(-1.1405)			
4		-0.0228	-1.3554	-17.7692			-124.66
(t-value)		(-1.9502)	(-2.2631)	(-11.1494)			
5		-0.0071	-0.9158		0.2922		-437.22
(t-value)		(-1.0732)	(-1.6567)		(2.1042)		
6		-0.0054	-0.0029			-0.2527	-437.91
(t-value)		(-0.8357)	(-0.0068)			(-1.7559)	
7		-0.0184	-0.9908	-17.7359		-0.4841	-122.95
(t-value)		(-1.6611)	(-1.5959)	(-10.9677)		(-1.8402)	
8		-0.0048	-0.9101		0.3801	-0.3546	-434.38
(t-value)		(-0.7507)	(-1.6464)		(2.6402)	(-2.3736)	
9		-0.02249	-1.2599	-17.8387	-0.0829		-124.60
(t-value)		(-1.9491)	(-1.9160)	(-11.1593)	(-0.3413)		
10		-0.0168	-1.2434	-17.7280	0.4779	-0.8401	-122.00
(t-value)		(-1.4852)	(-1.8928)	(-10.6563)	(1.3737)	(-2.2861)	
11	22.2030	1.00E-08					-1.446e-07
(t-value)	(0.0084)	(0.0000)					
12	22.2030	1.00E-08	3.00E-06				-1.446e-07
(t-value)	(0.0084)	(0.0000)	(0.0000)				

4.3.3. Calibrated parameters: hypothesis 3

The paragraph concerns the calibration of the specified run choice models assuming hypothesis 3 for the choice set building (see par. 4.2.1). An example of choice set building according to hypothesis 2 is reported in Fig. 4.10.

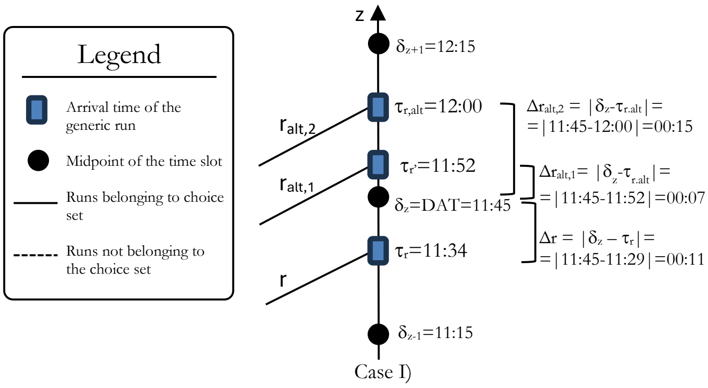


Fig. 4.10. Example of runs of the choice set: hypothesis 3

Tab. 4.24 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the complete sample composed of 635 users belonging to the full sample of year 2025: minimum, maximum, average value and standard deviation.

Tab.4.24. Descriptive statistics of attributes: full sample of year 2025 (hypothesis 3)

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.87	3.67	3.14	0.14
Fare (x_w)	€	[0; +∞]	49.90	295.00	102.92	23.66
Company (x_{co})		[0;1]	0	1	0,35	0,48
Stops (x_s)		[0;1]	0	1	0,73	0,44
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.00	0.20	0.03	0.03
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.00	0.20	0.06	0.07

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.25 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1, 2 and 3 have corrected signs of calibrated parameters. Some values of t-student statistics are not significant.

Tab.4.25. Calibrated parameters: full sample of year 2025 (hypothesis 3)

Model	β_w	β_t	β_{co}	β_Δ	β_s	$L(\boldsymbol{\beta})$
1	-0.0144					-694.71
(t-value)	(-2.215)					
2	-0.0143	-0.3037				-671.70
(t-value)	(-2.2016)	(-0.8928)				
3	-0.0062		-0.7084			-694.31
(t-value)	(-1.1172)		(-6.5370)			
4	-0.0137	-0.5224	1.2126			-692.35
(t-value)	(-2.1178)	(-1.4424)	(1.9677)			
5	-0.0145	-0.5779		0.1202		-693.91
(t-value)	(-2.2323)	(-1.2403)		(0.8884)		
6	-0.0059	0.6274			-0.7869	-670.15
(t-value)	(-1.0256)	(1.7779)			(-6.6446)	
7	-0.0059	0.5029	0. 6071		-0.7716	-669.75
(t-value)	(-1.0340)	(1.3215)	(0.8905)		(-6.4453)	
8	-0.00056	-0.2402		0.4119	-0.8674	-665.78
(t-value)	(-0.9673)	(-0.4995)		(2.9177)	(-7.2452)	
9	-0.0139	-0.8648	1.2603	0.1439		-691.78
(t-value)	(-2.1508)	(-1.7474)	(2.0519)	(1.0583)		
10	-0.0056	-0.4137	0.7321	0.4214	-0.8524	-665.19
(t-value)	(0.9742)	(-0.8111)	(1.0817)	(2.9790)	(-7.0627)	

4.3.4. Calibrated parameters: hypothesis 4

The paragraph concerns the calibration of the specified run choice models assuming the hypothesis 4 for the choice set building (see par. 4.2.1). An example of choice set building according to hypothesis 2 is reported in Fig. 4.11.

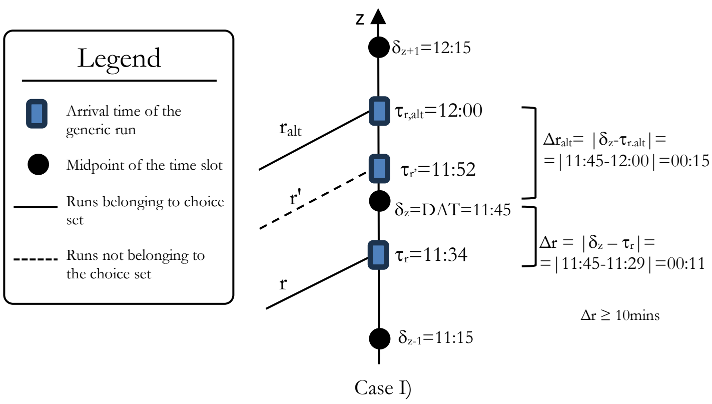


Fig. 4.11. Example of runs of the choice set: hypothesis 4

Tab. 4.26 contains some descriptive statistics of the time, fare, company, stops and penalty attributes associated to the runs selected by the users belonging to the complete sample composed of 635 users belonging to the full sample of year 2025: minimum, maximum, average value and standard deviation.

Tab.4.26. Descriptive statistics of attributes: full sample of year 2025 (hypothesis 4)

Attribute (x)	UoM	Range	Min	Max	Avg	S.D.
Time (x_t)	h	[0; +∞]	2.87	3.67	3.14	0.14
Fare (x_w)	€	[0; +∞]	49.90 €	295.00 €	103.32 €	24.24 €
Company (x_{co})		[0;1]	0	1	0.30	0.46
Stops (x_s)		[0;1]	0	1	0.69	0.46
Penalty (chosen run) (x_Δ)	h	[-∞; +∞]	0.00	0.20	0.03	0.06
Penalty (alternative run) (x_Δ)	h	[-∞; +∞]	0.00	0.20	0.06	0.07

with: UoM=unity of measure Min=minimum; Max=maximum; Avg=average; S.D.=standard deviation

Tab. 4.25 reports the calibration results of a set of specified models (see Tab. 4.2). It is worth noting that models 1 and 2 have corrected signs of calibrated parameters. Some values of t-student statistics are not significant.

Tab.4.27. Calibrated parameters: full sample of year 2025 (hypothesis 4)

Model	β_w	β_t	β_{co}	β_Δ	β_s	L($\boldsymbol{\beta}$)
1	-0.0074					-439.48
(t-value)	(-1.1071)					
2	-0.0073	0.1659				-426.97
(t-value)	(-1.0991)	(0.4107)				
3	-0.0027		-0.6261			-439.40
(t-value)	(-0.4197)		(-4.8864)			
4	-0.0008	0.2726	-6.0739			-413.08
(t-value)	(-0.1319)	(0.6587)	(-6.8467)			
5	-0.0079	-0.8351		0.3926		-435.01
(t-value)	(-1.1553)	(-1.5610)		(-2.9415)		
6	-0.0018	0.8427			-0.7021	-425.03
(t-value)	(-0.2907)	(1.9587)			(-5.2376)	
7	0.0004	0.5269	-5.1086		-0.3092	-411.15
(t-value)	(0.0608)	(1.2134)	(-5.1029)		(-1.9726)	
8	0.00002	-0.0963		0.0868	-1.0773	-408.19
(t-value)	(0.0028)	(-1.7833)		(5.6738)	(-7.0256)	
9	-0.0012	-0.7504	-6.1503	0.4281		-408.14
(t-value)	(-0.1827)	(-1.4062)	(-6.9252)	(3.1172)		
10	0.0014	-0.8721	-3.9492	0.7203	-0.7074	-400.46
(t-value)	(0.2184)	(-1.6222)	(-3.8591)	(4.5546)	(-3.9154)	

4.4. Discussion

The paragraph presents a discussion about the calibration framework of parameters of the run choice models.

As far as concerns Tab. 4.3, which reports the calibration process carried out, a comparison among a selection of calibrated models is presented below.

The first comparison (along a horizontal line in Tab .4.3) concerns the parameters of a selection of models calibrated by means of full samples 2022, 2023, 2025, 2023+2025, 2022+2023+2025. The selected models are based on the same hypothesis (1) about the choice set building (see parr. 4.3.1c; 4.3.1d; 4.3.1e; 4.3.1f; 4.3.1g).

It is important to compare the calibrated parameters of the models with the same specification. Tab. 5.28 and 5.29 contains the models with the calibrated parameters having correct sign. It is worth noting that no model of the full sample 2022, have calibrated parameters with correct sign.

Tab.4.28. Comparison of calibrated parameters: run choice model

Sample	H	Model	β_w	β_t	β_{co}	β_Δ	β_s
2022	1	-	-	-	-	-	-
2023	1	1	-0.0242	-	-	-	-
2023	1	2	-0.0487	-0.1138	-	-	-
2023	1	3	-0.0247	-	-0.6001	-	-
2023	1	5	-0.0249	-0.1206	-	-	-0.0106
2025	1	1	-0.0074	-	-	-	-
2025	1	3	-0.0027	-	-0.6261	-	-
23+25	1	1	-0.0151	-	-	-	-
23+25	1	2	-0.0151	-0.0258	-	-	-
23+25	1	3	-0.0188	-	-0.4686	-	-
22+23+25	1	1	-0.0051	-	-	-	-
22+23+25	1	3	-0.0111	-	-0.5158	-	-

with: H=hypothesis.

The comparison of calibrated parameters of model 1 shows that the model of the full sample 2023 has the fare parameter with the highest value. Model 2 has correct sign for the calibrated parameters of full samples 2023 and 2023+2025. Model 3 has correct signs of the calibrated parameters for all full samples. The model 5 of full sample 2023 has correct sign in the calibrated parameters, however there are no other models to compare with.

The second comparison (along a vertical line in Tab .4.3) is among the calibrated parameters of a selection of models specified with the different hypothesis (1-4) concerning the choice set building and calibrated by using the same full sample 2025 (see par 4.3.1e; par: 4.3.2; 4.3.3; 4.3.4).

In this case it is important to compare the same model with different hypotheses on the choice set building, thus are compared the models 1 of full sample 2025 with the hypothesis 2, 3 and 4. The calibrated parameter results higher, with correct sign, with hypothesis 3, hence the trinomial model. The comparison of models 2 of full sample 2025 with hypotheses 2 and 3 show that the calibrated parameters result higher, with correct sign, with hypothesis 3, hence the trinomial model. The comparison of models 3 of full sample 2025 with the hypotheses 2, 3 and 4 show that the calibrated parameters result higher, with correct sign, with hypothesis 3, hence the trinomial model.

Tab.4.29. Comparison of calibrated parameters: run choice model

Sample	H	Model	β_w	β_t	β_{co}	β_Δ	β_s
2025	1	1	-0.0074	-	-	-	-
2025	1	3	-0.0027	-	-0.6261	-	-
2025	2	1	-0.0069	-	-	-	-
2025	2	2	-0.0069	-0.1647	-	-	-
2025	2	3	-0.0054	-	-0.2529	-	-
2025	3	1	-0.0144	-	-	-	-
2025	3	2	-0.0143	-0.3037	-	-	-
2025	3	3	-0.0062	-	-0.7084	-	-
2025	4	1	-0.0074	-	-	-	-
2025	4	3	-0.0027	-	-0.6261	-	-

With: H=hypothesis.

4.A. Appendix

Appendix A concerns the application of the procedure specified in paragraph 4.1 and concerning the identification of fare variations, and therefore of user choice, starting from the coding of the tariff matrices supplied by the transport companies.

0. <u>Coding of the generic fare matrix</u>

It is considered the day t=01/09/2023

It is considered an interval of days K=30 and let k-1=-10 (23/8/23), k=-9 (24/8/23) belonging to K; keK. An application is reported for the Italian case and the Trenitalia company.

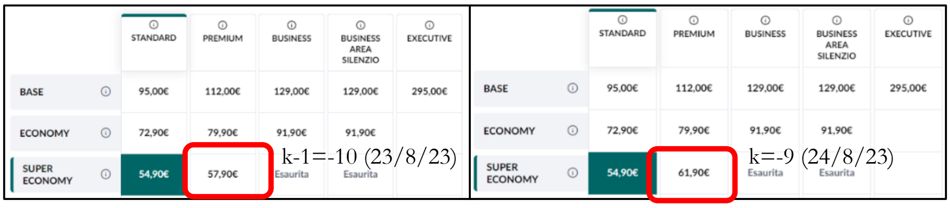


Fig. 4.12. Example of the fare matrix

The elements within the fare matrix will be:

For simplicity of example Classes (c ∈ C) and Quality of Service (q ∈ Q) will be abbreviated as follows:

CLASSES: “Base”=BA; “Economy”=EC; “Super Economy”=SE

QUALITY OF SERVICE: “Standard”=ST; “Premium”=PR; “Business”=BS; “Business Area Silenzio”=BS.AS; “Executive”=EXE

Tab.4.30. Example of fare matrix coding

	ST	PR	BS	BS.AS	EXE
BA	BA-ST	BA-PR	BA-BS	BA-BS.AS	BA-EXE
EC	EC-ST	EC-PR	EC-BS	EC-BS.AS	EC-EXE
SE	SE-ST	SE-PR	SE-BS	SE-BS.AS	SE-EXE

A numbering of fares is made starting from: BA-ST(c=1; q=1).

Tab.4.31. Example of numbering of the fare matrix elements

	ST	PR	BS	BS.AS	EXE
BA	1	2	3	4	5
EC	6	7	8	9	10
SE	11	12	13	14	15

The vector of fares $w_{r,k}$ is inserted inside the matrix W

Trenitalia, r_{k-1}=9620, k=-10; K=30.

Tab.4.32. Example of fare vector at day k inside the fare matrix

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9620	95 €	112 €	129 €	129 €	295 €	72,9 €	79,9 €	91,9 €	91,9 €	-	54,9 €	57,9 €	-	-	-

Trenitalia, r_k=9620, k=-9; K=30

Tab.4.33. Example of fare vector at day k+1 inside the fare matrix

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9620	95 €	112 €	129 €	129 €	295 €	72,9 €	79,9 €	91,9 €	91,9 €	-	54,9 €	61,9 €	-	-	-

<u>Matrix Δw'.</u> r_k=9620, with k-1=-10 e k=-9, K=30 (if $w_{k-1} - w_k > 0$ so is 1, otherwise 0). w_{k-1}-w_k>0; 12_{10}-12_9 = 57,9 € - 61,9 € = 4,0 €, then code 1 at fare number 12 of run 9620.

Tab.4.34. Example of coding fare differences between two consecutive days

run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9620	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

1. <u>Identification of number of mutations</u>

1.1) Let r=generic run; t=01/09/2023, k=generic k, k+1=generic k+1.
Simulated case
let N_k=1
let r : $N_k(r) = 1$; r' : N_k (r') = 0
for ∀ r' ≠ r, r' ∈ R_k

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9620 09:25	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

1.2) let r=generic run; t=01/09/2023, k=generic k, k+1=generic k+1.
Simulated case
Let N_k=2
Let r : $N_k(r) = 2$; r' : N_k (r') = 0
for∀ r' ≠ r, r' ∈ R_k

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9618 08:50	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

1.3) Let r=generic run; t=01/09/2023, k=generic k, k+1=generic k+1.
Simulated case
Let N_k=2

Let r : $N_k(r)=1$; r' : N_k (r')=1; r'': N_k (r'')=0; for $\forall$ r''$\neq$ r, r''$\neq$r'; r''$\in R_k$

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9608 06:50	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
9612 07:50	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

1.4) Let r=generic run; t=01/09/2023, k=generic k, k+1=generic k+1.
Simulated case
Let $N_k= 4$
Let r : $N_k(r)=2$; r' : N_k (r')=2; r'': N_k (r'')=0; per $\forall$ r''$\neq$ r, r''$\neq$r'; r''$\in R_k$

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9612 06:50	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0
9620 07:50	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0
run hh:mm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

1.5) Let r=9608 (Trenitalia), r=9612 (Trenitalia) e 9618 (Trenitalia); t=01/09/2023; k=22, k+1=21; K=30
Let $N_k= m_k$
Let r : N_k (r) = $m_{k,r}$ $\forall$ r $\in R_k$ e
$\sum_{r \in R_k} N_k(r) = N_k = m_k$

r	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
9606	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	**1**	0
9608	0	0	0	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0
9612	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0
9618	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9620	0	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0
9622	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9968	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9972	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0

2. <u>Users identification</u>

Let $N_k > 0$ for k e $N_{k'}=0$ for $\forall$ k $\neq$k'; k' $\in$ K

2.1) follows from the hypothesis 1.1)

One user chose the race r=9620 (simulated case)

Δc'	
0	$N_k(r)$
1	$N_k(9620)$
0	$N_k(r)$
0	$N_k(r)$
1	N_k

2.2) follows hypothesis 1.2)

Two users chose the run r=9618 (simulated case)

Δc'	
0	$N_k(r)$
0	$N_k(r)$
0	$N_k(r)$
2	$N_k(9618)$
2	N_k

2.3) follows hypothesis 1.3)

One user chose the generic run r=9608 and another the generic run r'=9612 (simulated case)

Δc'	
0	$N_k(r)$
1	$N_k(9608)$
1	$N_k(9612)$
0	$N_k(r)$
2	N_k

2.4) follows hypothesis 1.4)

Two users chose the generic run r=9608 and two users chose the generic run r'=9612 (simulated case)

Δc'	
0	$N_k(r)$
2	$N_k(9608)$
2	$N_k(9612)$
0	$N_k(r)$
4	N_k

2.5) follows hypothesis 1.5)

Are identified m_k users who choose in the day k; $m_{k,r}$ users choose la generic run r (real case)

Δc'	
4	$N_k(9606)$
3	$N_k(9608)$
2	$N_k(9612)$
0	$N_k(9618)$
1	$N_k(9620)$
0	$N_k(9622)$
0	$N_k(9968)$
1	$N_k(9972)$
11	N_k

3. <u>Identification of likelihood function $L=L_k$ for day k</u>

(hypotheses 3.1, 3.2, 3.3 and 3.4 refer to simulated and non-real scenarios, case 3.5 refers to a real case)

3.1) follows hypothesis 2.1)

ID	run	Choice
1	r	0
1	9620	1
1	r	0
1	r	0

$$L_k = p_{kr} = \prod p_{kr}^{N_k(r)} = p_{9,9620}^1$$

3.2) follows hypothesis 2.2)

ID	run	Choice
1	r	0
1	r	0
1	r	0
1	9618	1
2	r	0
2	r	0
2	r	0
2	9618	1

$$L_k = p_{kr} \cdot p_{kr} = p_{kr}^2 = \prod p_{kr}^{N_k(r)} = p_{kr}^2 = p_{9,9618}^2$$

3.3) follows hypothesis 2.3)

ID	run	Choice
1	r	0
1	9608	1
1	r	0
1	r	0
2	r	0
2	r	0
2	9612	1
2	r	0

$$L_k = p_{kr} \cdot p_{kr'} = \prod p_{kr} \cdot p_{kr'} = = \prod p_{kr}^{N_k(r)} \cdot p_{kr}^{N_k(r')}$$
$$= p_{9,9608}^1 \cdot p_{9,9612}^1$$

3.4) follows hypothesis 2.4)

ID	run	Choice
1	r	0
1	9608	1
1	r	0
1	r	0
2	r	0
2	9608	1
2	r	0
2	r	0
3	r	0
3	r	0
3	9612	1
3	r	0
4	r	0
4	r	0
4	9612	1
4	r	0

$$L_k = (p_{kr} \cdot p_{kr}) \cdot (p_{kr'} \cdot p_{kr'}) = \prod (p_{kr}^2 \cdot p_{kr'}^2)$$
$$= (p_{9,9608}^1 \cdot p_{9,9608}^1) \cdot (p_{8,9612}^1 \cdot p_{8,9612}^1) = p_{9,9608}^2 \cdot p_{8,9612}^2$$

3.5) follows hypothesis 2.5) this hypothesis refers to a real case (example with reference to N_k (9606)=4, with t=1/9/23, k=-9, k+1=-8, K=30

ID	run	Choice
1	9606	1
1	9608	0
1	9612	0
1	9618	0
1	9620	0
1	9622	0
1	9972	0
2	9606	1
2	9608	0
2	9612	0
2	9618	0
2	9620	0
2	9622	0
2	9972	0
3	9606	1
3	9608	0
3	9612	0
3	9618	0
3	9620	0
3	9622	0
3	9972	0
4	9606	1
4	9608	0
4	9612	0
4	9618	0
4	9620	0
4	9622	0
4	9972	0

$$L_k = \prod\nolimits_{r \in R_k} p_{kr}^{m_{k,r}} = \prod\nolimits_{r \in R_k} p_{kr}^{N_{k(r)}}$$

$$= \left(p_{9,9606}^1 \cdot p_{9,9606}^1 \cdot p_{9,9606}^1 \cdot p_{9,9606}^1\right)$$

$$\cdot \left(p_{9,9608}^1 \cdot p_{9,9608}^1 \cdot p_{9,9608}^1\right)$$

$$\cdot \left(p_{9,9612}^1 \cdot p_{9,9612}^1\right) \cdot p_{9,9620}^1 \cdot p_{9,9972}^1$$

$$= p_{9,9606}^4 \cdot p_{9,9608}^3 \cdot p_{9,9612}^2$$

$$\cdot p_{9,9620}^1 \cdot p_{9,9972}^1$$

4. <u>Likelihood function L for all days k belonging to K</u>

4.1) Let us assume that for each day kϵK there is only one mutation, in a run r=9620 with k-1=-9, k=-8 (simulated case), and therefore the 1.1; 2.1; 3.1 hold.

$$L = \prod\nolimits_{k \in K} p_{kr} = \prod\nolimits_{k \in K} p_{kr}^{N_k(r)} = p_{9,9620}^1 \cdot \ p_{8,9620}^1$$

4.2) Let's assume that for each day kϵK there are two mutations for the run r=9618 and 0 for the others, with k-1=-9, k=-8 (simulated case), 1.2, 2.2, 3.2 hold.

$$L = \prod\nolimits_{k \in K} p_{kr}^2 = \prod\nolimits_{k \in K} p_{kr}^{N_k(r)} = \left(p_{9,9618}^2\right) \cdot \left(p_{8,9618}^2\right)$$

4.3) Let's assume that for each day kϵK there are two mutations, one for the run r=9608 and one for the run r'=9612 with k-1=-9, k=-8 (simulated case).

$$L=\prod_{k\in K} p_{kr}\cdot p_{kr'}=\prod_{k\in K} p_{kr}^{N_k(r)}\cdot p_{kr}^{N_k(r')}$$
$$=\left(p_{9,9608}^{1}\cdot p_{9,9612}^{1}\right)\cdot\left(p_{8,9608}^{1}\cdot p_{8,9612}^{1}\right)$$

4.4) Let's assume that for each day kɛK there are four mutations, two for the run r=9608 and two for the run r'=9612 with k-1=-9, k=-8 (simulated case).

$$L=\prod_{k\in K}(p_{kr}\cdot p_{kr})\cdot(p_{kr'}\cdot p_{kr'})=\prod_{k\in K}⟦\left(p_{kr}^{2}⟧\cdot p_{kr'}^{2}\right)$$
$$=\prod_{k\in K} p_{kr}^{N_k(r)}\cdot p_{kr}^{N_k(r')}=$$
$$=\left[\left(p_{9,9608}^{1}\cdot p_{9,9608}^{1}\right)\cdot\left(p_{9,9612}^{1}\cdot p_{9,9612}^{1}\right)\right]$$
$$\cdot\left[\left(p_{8,9608}^{1}\cdot p_{8,9608}^{1}\right)\cdot\left(p_{8,9612}^{1}\cdot p_{8,9612}^{1}\right)\right]=$$
$$=\left(p_{9,9608}^{2}\cdot p_{9,9612}^{2}\right)\cdot\left(p_{8,9608}^{2}\cdot p_{8,9612}^{2}\right)$$

4.5) In the general case of multiple chosen runs on day k, each chosen run $m_{k,r}$ times, the function L for the generic day k is given by 3.5 and the overall function L is (this hypothesis refers to a real case, with t=1/9/23, k-2=-11, k-1=-10, k=-9, k+1=-8; K=30).

$$L=\prod_{k\in K}\prod_{r\in R_k} p_{kr}^{N_k(r)}$$
$$=\left[\left(p_{10,9606}^{1}\cdot p_{10,9606}^{1}\cdot p_{10,9606}^{1}\cdot p_{10,9606}^{1}\right)\right.$$
$$\cdot\left(p_{10,9608}^{1}\cdot p_{10,9608}^{1}\cdot p_{10,9608}^{1}\right)\cdot\left(p_{10,9612}^{1}\cdot p_{10,9612}^{1}\right)\cdot p_{10,9620}^{1}$$
$$\left.\cdot\, p_{10,9972}^{1}\right]$$
$$\cdot\left[\left(p_{9,9606}^{1}\cdot p_{9,9606}^{1}\cdot p_{9,9606}^{1}\cdot p_{9,9606}^{1}\right)\right.$$
$$\cdot\left(p_{9,9608}^{1}\cdot p_{9,9608}^{1}\cdot p_{9,9608}^{1}\right)\cdot\left(p_{9,9612}^{1}\cdot p_{9,9612}^{1}\right)$$
$$\left.\cdot\, p_{9,9620}^{1}\cdot p_{9,9972}^{1}\right]$$
$$\cdot\left[\left(p_{8,9606}^{1}\cdot p_{8,9606}^{1}\cdot p_{8,9606}^{1}\cdot p_{8,9606}^{1}\right)\right.$$
$$\cdot\left(p_{8,9608}^{1}\cdot p_{8,9608}^{1}\cdot p_{8,9608}^{1}\right)\cdot\left(p_{8,9612}^{1}\cdot p_{8,9612}^{1}\right)$$
$$\left.\cdot\, p_{8,9620}^{1}\cdot p_{8,9972}^{1}\right]$$
$$=\left(p_{10,9606}^{4}\cdot\ p_{10,9608}^{3}\cdot\ p_{10,9620}^{2}\cdot p_{10,9620}\cdot p_{10,9972}\right)$$
$$\cdot\left(p_{9,9606}^{4}\cdot\ p_{9,9608}^{3}\cdot\ p_{9,9612}^{2}\cdot p_{9,9620}\cdot p_{9,9972}\right)$$
$$\cdot\left(p_{8,9606}^{4}\cdot\ p_{8,9608}^{3}\cdot\ p_{8,9612}^{2}\cdot p_{8,9620}\cdot p_{8,9972}\right)$$

5. <u>General L function for run choice and fare</u>

If it is intended to insert the fares and calibrate respectively also to the choice of tariff it is needed to do:

5.1) $L = \prod_{k \in K} \prod_{r \in R_k} \prod_{w \in r} p_{kr}{}^{N_k(r,w)}$

With $N_k(r,w)$ in that case always equal to1

	k-2	k-1	k	k+1	
	(k-12)-(k-11)	(k-11)-(k-10)	(k-10)-(k-9)	(k-9)-(k-8)	
r	21_22	22_23	23_24	24_25	
	$N_k(r)$	$N_k(r)$	$N_k(r)$	$N_k(r)$	
9606	0	2	4	4	
9608	0	0	3	4	
9612	0	1	2	2	
9618	0	0	0	0	
9620	0	2	1	2	
9622	0	0	0	0	
9968	0	0	0	0	
9972	0	0	1	2	
N_k	0	5	11	14	30

This hypothesis refers to a real case, with t=1/9/23, k-2=11, k-1=10, k=9, k+1=8; K=30.

$$L = \prod_{k \in K} \prod_{r \in R_k} \prod_{w \in r} p_{kr}^{N_k(r,w)} =$$

$$= \left[\left(p^1_{10,9606,13} \cdot p^1_{10,9606,14}\right) \cdot p^1_{10,9612,8}\right.$$
$$\left.\cdot \left(p^1_{10,9620,11} \cdot p^1_{10,9620,12}\right)\right] \cdot$$
$$\cdot \left[\left(p^1_{9,9606,11} \cdot p^1_{9,9606,12} \cdot p^1_{9,9606,13} \cdot p^1_{9,9606,14}\right)\right.$$
$$\cdot \left(p^1_{9,9608,7} \cdot p^1_{9,9608,8} \cdot p^1_{9,9608,9}\right) \cdot\cdot \left(p^1_{9,9612,11} \cdot p^1_{9,9612,12}\right)$$
$$\left.\cdot p^1_{9,9620,12} \cdot p^1_{9,9972,5}\right] \cdot$$
$$\cdot \left[\left(p^1_{8,9606,11} \cdot p^1_{8,9606,12} \cdot p^1_{8,9606,13} \cdot p^1_{8,9606,14}\right)\right.$$
$$\cdot \left(p^1_{8,9608,11} \cdot p^1_{8,9608,12} \cdot p^1_{8,9608,13} \cdot p^1_{8,9608,14}\right) \cdot$$
$$\cdot \left(p^1_{8,9612,11} \cdot p^1_{8,9612,12}\right) \cdot \left(p^1_{8,9620,11} \cdot p^1_{8,9620,12}\right)$$
$$\left.\cdot \left(p^1_{8,9972} \cdot p^1_{8,9972}\right)\right]$$

6.1. <u>Choice Set (1) with known DAT</u>

<u>6.1.a</u>

6.3.1a <u>(follows hypothesis 6.1)</u>

$V_{kr} = w_{kr} \cdot \Delta_w$

$$p_{kr} = \frac{\exp(V_{kr})}{\sum_r \exp(V_{kr})}$$

5.

Fare supply modelling

The chapter presents the specification, calibration and validation of the framework aimed at the fare modelling building process. Are analyzed the fare structures, thus the evolution of the fares, by means of considering a single run and set of runs, clustered by time slots.

The chapter is organized into four paragraphs, as follows. Paragraph 5.1 contains the specification of the applied framework for the fares analysis. Paragraph 5.2 contains the examples of the fares observations, by means of single runs and sets of runs, analyzed as k varies and t fixed. Paragraph 5.3 regards the examples of the fares observations, by means of single runs and sets of runs, analyzed as t varies and k fixed. Paragraph 5.4 regards the examples of the fares observations, by means of single runs and sets of runs, analyzed as k varies and t fixed. Paragraph 5.5 regards the pilot analysis on a single run, analyzed as t varies and k fixed and as k varies and t fixed.

Paragraph 5.6 contains the calibrated parameters of the specified fare models, which regard the single run and the set of runs. Finally, Appendix B contains additional model specifications.

5.1. Framework

The framework is composed of three main steps (Fig. 5.1):

Step 1: Definition of domain;

Step 2: Fare observation;

Step 3: Fare modelling.

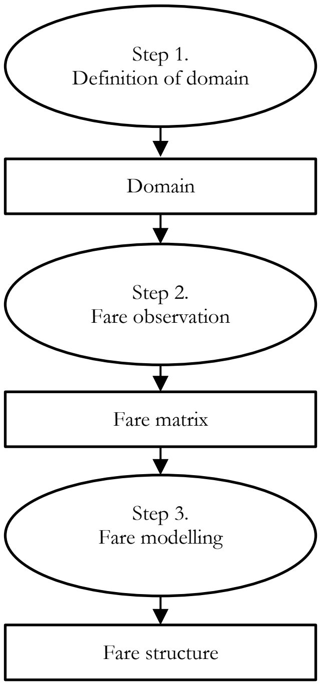


Fig. 5.1. Steps of the framework for fare supply modelling

Step 1: Definition of domain**.** The first step concerns the definition of the domains of the variables to be surveyed.

- T, set of days of trip. It is possible to consider an exhaustive approach, for instance considering one year in order to observe the seasonal pattern of fares. Otherwise, a selective approach is considered, for instance selecting a specific period of the year in order to focus on a specific patterns of fares (e.g., vacation period during summer time).
- K, interval of days of ticket purchasing before t. It is important to consider a sufficient extended period in order to capture the dynamic pattern of fares, if there are any, along t and k.
- G, number of fare types (classes) for operating runs and companies at day of trip t. The selection of the fares to be surveyed is based on several criteria, for instance the cheapest fare among the different class of fares.
- R^t_k, set of runs available for user at day k. It is possible to consider the whole scheduled runs travelling in the day of trip t (exhaustive approach), otherwise a sample of scheduled runs (selective approach) selected according to one or more criteria. For instance, one or more

time windows of the day when it is supposed to have the higher level of traffic, and one or more reference runs travelling (departing or arriving) inside the identified time windows selected according to same attributes, such as the number of intermediate stops, the minimum on-board travel time, or other attributes.

Step 2: Fare observation**.** The observation of fares evolution may be executed according to different modalities, for instance when data are directly provided by the transport company, otherwise by means of a direct observation of available public data. The latter may be implemented by data acquisition from the companies' websites, manually or with the support of automatic procedures, for instance scraping procedures. The input of Step 2 is the domain of analysis, as well as the output is the observed fares, generally organized into a matrix for taking into account the several characteristics associated to complementary services associated to each fare. The fare selected is the minimum value of ticket available of the considered run r operated by a transport company for every day of ticket purchasing $k \in K$. Therefore, the resulting observed diagram is the envelope of the minimum fare values among all the supplied fares along the days k. Fig. 5.2 e 5.3 report two examples of plots of observed fares along the two variables t and k considered.

Step 3: fare modelling**.** The general equation of fare considered is given below:

$$w = w\,(t, k, g) \qquad (5.1)$$

with *w*(), fare function to be specified.

The eq. (5.1) degenerates into the following equation in the Cartesian plan w-k:

$$w = w\,(\bar{t}, k, \bar{g}) \qquad (5.2)$$

when the travel day t is fixed: $t=\bar{t}$ and the fare g is fixed: $g = \bar{g}$.

The eq. (5.1) degenerates into the following equation in the Cartesian plan w-t:

$$w = w\,(t, \bar{k}, \bar{g}) \qquad (5.3)$$

when the purchase day k is fixed: $k=\bar{k}$ and the fare g is fixed: $g=\bar{g}$.

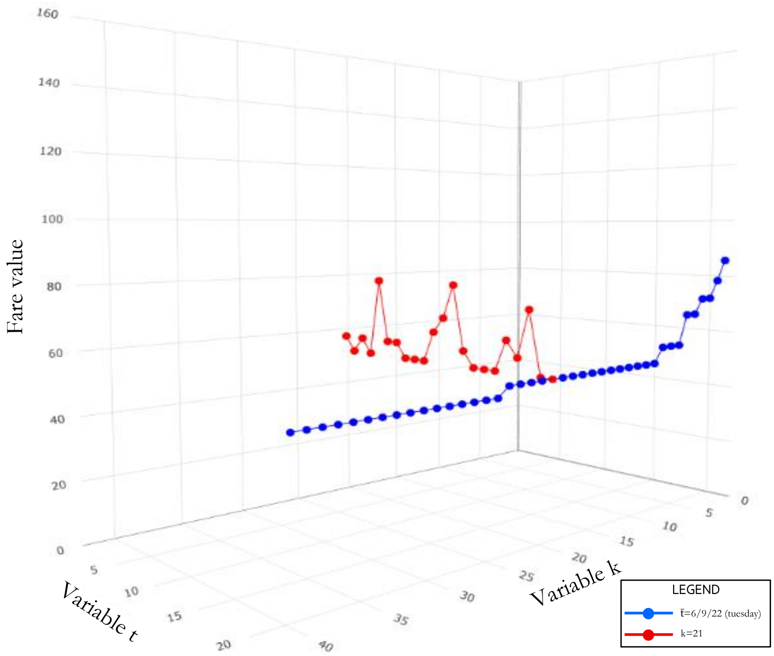


Fig. 5.2. Example of observed diagrams: single days

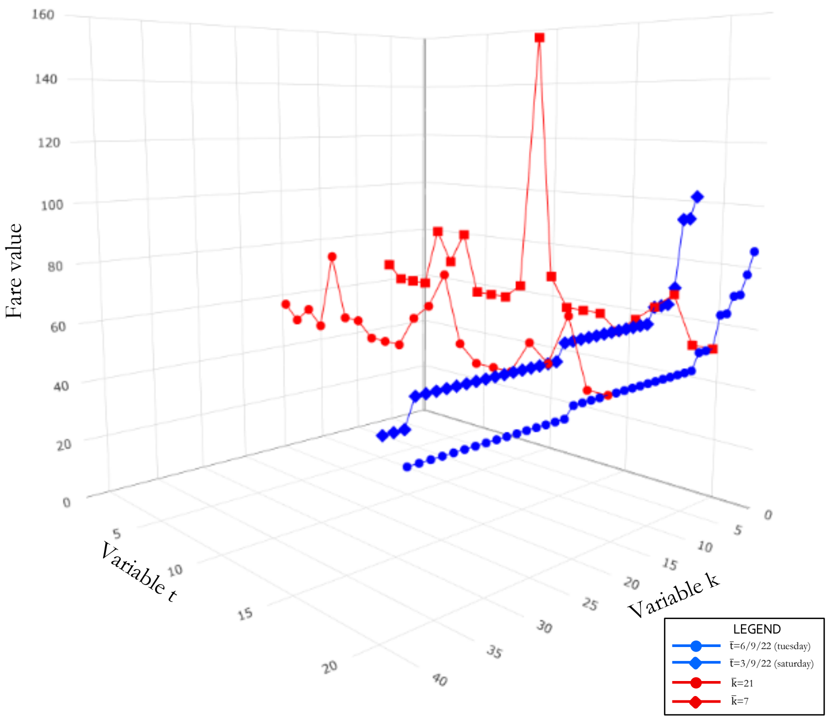


Fig. 5.3. Example of observed diagrams: couple of days

The chapter is organized as follows. Paragraph 5.2 contains the two-dimensional analysis of fares observed in the w-k Cartesian plan for t fixed (step 1). Paragraph 5.3 concerns the two-dimensional analysis of the fares observed in the w-t Cartesian plan with k fixed (Step 2). Paragraph 5.4 contains the results of calibration of the fare function w = *w* (k) from eq. (5.2) (Step 3). Finally, Appendix B contains several alternative specifications of two-dimensional fare functions.

5.2. Domain of analysis

The domain of analysis was built according to a selective approach, and it refers to the relationship Rome-Milan (Italy) considering both directions of travel, the runs selected which may be single runs or set of runs, the time slots identified during the day.

According the above assumptions, the following elements are considered:

- Single runs:
  - Direction Rome-Milan:
    - ✓ run 9634 (Trenitalia);
    - ✓ run 6770 (NTV).
  - Direction Milan-Rome:
    - ✓ run 9631 (Trenitalia);
    - ✓ run 9981 (NTV).
- Set of runs, composed of shuttle runs, without intermediate stops or at least one intermediate stop, travelling inside some selected time windows of the day when higher level of traffic are expected:
  - Direction Rome-Milan:
    - ✓ 6:00-9:00, run 9508 (Trenitalia) and run 6636 (NTV);
    - ✓ 12:00-14:00, run 9634 (Trenitalia) and run 6770 (NTV);
    - ✓ 19:00-21:00, run 9560 (Trenitalia) and run 9962 (NTV).
  - Direction Milan-Rome:
    - ✓ 6:00-9:00, run 9515 (Trenitalia) and run 8111 (NTV);
    - ✓ 12:00-14:00, run 9631(Trenitalia) and run 9981 (NTV);
    - ✓ 19:00-21:00, run 9961 (Trenitalia) and run 9963 (NTV).

The schematization described above is reported in Tab. 5.1.

Tab.5.1. Identification of time slots and selected runs

Direction	Time slot		Selected runs	
	Number	[hh:mm]	Trenitalia	NTV
RO-MI	1	06:00-09:00	9508	6636
	2	12:00-14:00	9634	6770
	3	19:00-21:00	9560	9962
MI-RO	1	06:00-09:00	9515	8111
	2	12:00-14:00	9631	9981
	3	19:00-21:00	9961	9963

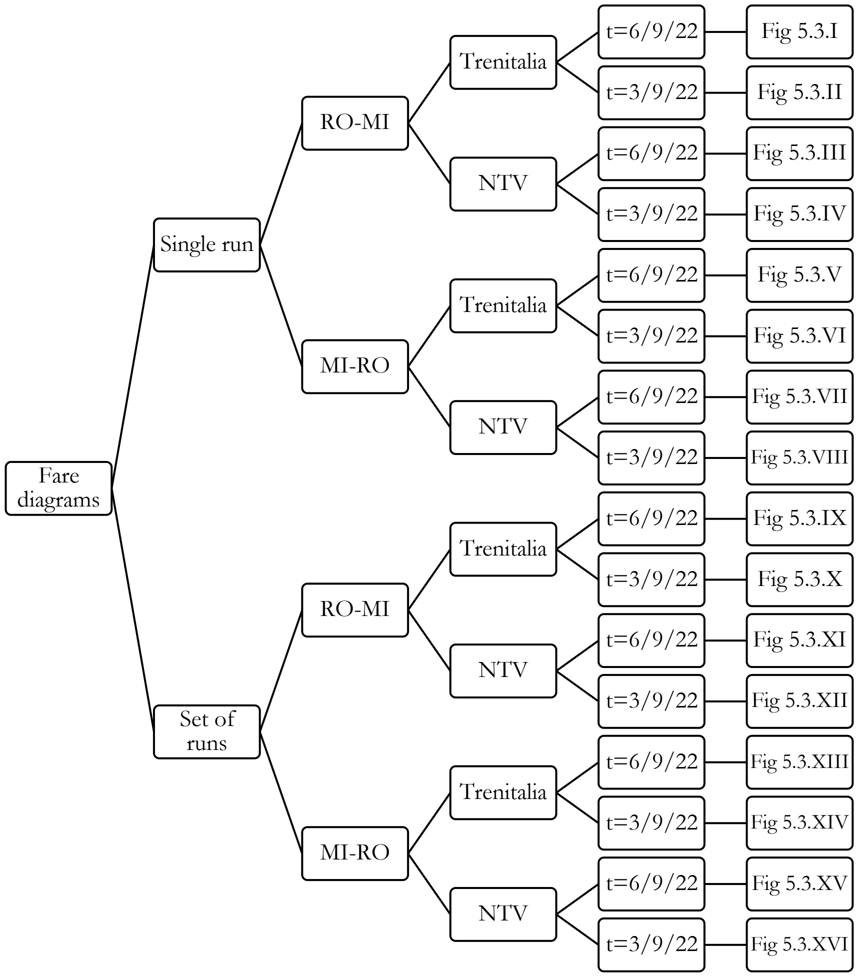
5.3. Fares in the plane: w-k

The paragraph presents a two-dimensional analysis of the observed fares in the w-k Cartesian plan, for t fixed ($t=\bar{t}$).

The survey of observed fares was conducted by consulting the websites of the railway transport companies Trenitalia, for the HSR "Frecciarossa" services and NTV for the HSR "Italo" services.

Fig. 5.4 and Tab. 5.2.a-b report the schemes of fare observations, which are classified into:

- single run vs. set of runs;
- direction: Rome-Milan (RO-MI) and Milan-Rome (MI-RO);
- transport company: Trenitalia and NTV;
- weekday: Tuesday (t=06/09/2022) vs. day of weekend: Saturday (t=03/09/2022),
- type of fare: "Super Economy-Standard" (SE-ST) fare for company Trenitalia company and "Low Cost-Smart" (LC-SM) fare for NTV;
- domain of observation, or cardinality $|K|$: 42 days for weekday vs 38 days for weekend day.



MI=Milan; RO=Rom**e**; 06/09/22, Tuesday; 03/09/22, Saturday

Fig. 5.4. Survey scheme of observed fares diagrams

Tab.5.2a. Survey scheme of fares diagrams (part a)

Run	Direction	Company	Day	Figure	t	|K|	g
Single run	RO-MI	Trenitalia	Tuesday	5.3.I	6/9/22	42	SE-ST
			Saturday	5.3.II	3/9/22	38	SE-ST
		NTV	Tuesday	5.3.III	6/9/22	42	LC-SM
			Saturday	5.3.IV	3/9/22	38	LC-SM
	MI-RO	Trenitalia	Tuesday	5.3.V	6/9/22	42	SE-ST
			Saturday	5.3.VI	3/9/22	38	SE-ST
		NTV	Tuesday	5.3.VII	6/9/22	42	LC-SM
			Saturday	5.3.VIII	3/9/22	38	LC-SM
Set of runs	RO-MI	Trenitalia	Tuesday	5.3.IX	6/9/22	42	SE-ST
			Saturday	5.3.X	3/9/22	38	SE-ST
		NTV	Tuesday	5.3.XI	6/9/22	42	LC-SM
			Saturday	5.3.XII	3/9/22	38	LC-SM
	MI-RO	Trenitalia	Tuesday	5.3.XIII	6/9/22	42	SE-ST
			Saturday	5.3.XIV	3/9/22	38	SE-ST
		NTV	Tuesday	5.3.XV	6/9/22	42	LC-SM
			Saturday	5.3.XVI	3/9/22	38	LC-SM

With: SE-ST="Super Economy-Standard"; LC-SM="Low Cost-Smart"

Tab.5.2b. Survey scheme of fares diagrams (part b)

Figure	Run	Arrival time	Time slot (*)
5.3.I	9634	15:58	2
5.3.II	9634	15:58	2
5.3.III	6770	16:35	2
5.3.IV	6770	16:35	2
5.3.V	9631	16:15	2
5.3.VI	9631	16:15	2
5.3.VII	9981	15:25	2
5.3.VIII	9981	15:25	2
5.3.IX	9508, 9634, 9560	9:50, 15:58, 22:50	1, 2, 3
5.3.X	9508, 9634, 9560	9:50, 15:58, 22:50	1, 2, 3
5.3.XI	6636, 6770, 9962	10:40, 16:35, 00:20	1, 2, 3
5.3.XII	6636, 6770, 9962	10:40, 16:35, 00:20	1, 2, 3
5.3.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	1, 2, 3
5.3.XIV	9515, 9631, 9961	10:49, 16:15, 22:10	1, 2, 3
5.3.XV	8111, 9981, 9963	10:19, 15:25, 00:19^{+1}	1, 2, 3
5.3.XVI	8111, 9981, 9963	10:19, 15:25, 00:19^{+1}	1, 2, 3

(*)see Tab 5.1.

5.3.1. Single runs

The group of figures 5.5.I- 5.5.IV depicts the fare observed of single runs along the direction Rome-Milan, whereas the group of Figures 5.5.V-5.5.VIII the single runs along the direction Milan-Rome.

The different observed diagrams of fare are characterized by common elements. The fare structure presents two main regions:

- stable region, more distant from the day of trip, is characterized by a sub-horizontal evolution of fares, where prices are almost stable;
- unstable region, closer to the day of trip, where the evolution of fares shows an increasing positive gradient due to the variation in ticket fares.

It is possible, in general, to identify a threshold which correspond to one or a group of days of ticket purchase; the latter separate the two regions above described.

Some observed diagrams diverge from the common structure described above as the observed curves depicted in Fig. 5.6.II and Fig. 5.10.VI, which need to be further investigated.

It is possible to define formally two branches of the fare diagram with stable and unstable values of fares and a critical value of k, as defined below:

- stable branch is the portion of the fare diagram such that the variations

Δw do not repeat between two consecutive values of k;

- unstable branch is the portion of the fare diagram such that the variations Δw are repeated at least between two consecutive values of k;
- critical value of k is the threshold such that a relevant change in the evolution of the fare from constant to variable is observed. Tab. 5.3 reports the extension of stable and unstable regions and the critical values of k.

Tab.5.3. Characteristics of fare diagrams: constant and variable patterns

Figure	|K|	Constant pattern	k critical	Variable pattern
5.3.I	42	32	10	10
5.3.II	38	38	8	8
5.3.III	42	35	7	7
5.3.IV	38	27	11	11
5.3.V	42	32	10	10
5.3.VI	38	29	9	9
5.3.VII	42	35	7	7
5.3.VIII	38	31	7	7

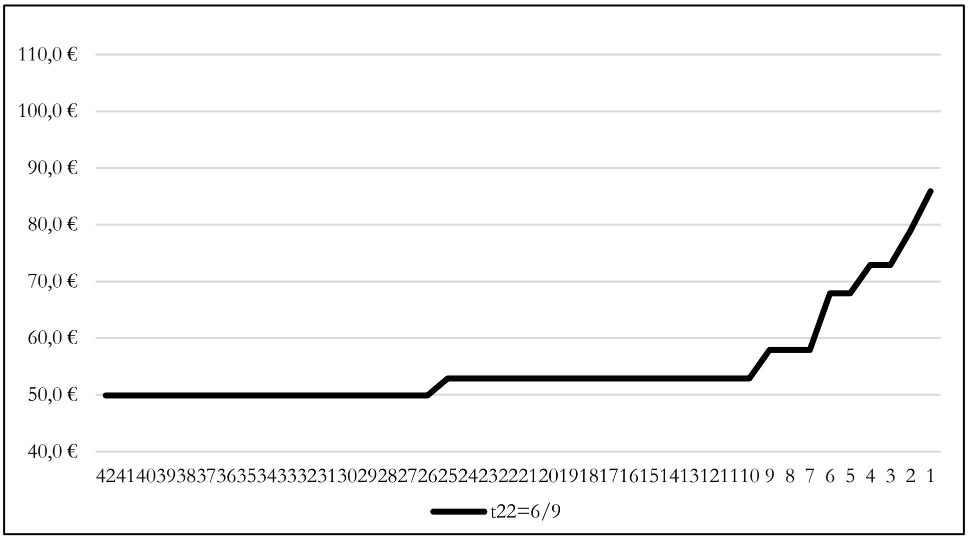


Fig. 5.3.I. Observed fare diagram of single run (Trenitalia 9634, Tuesday)

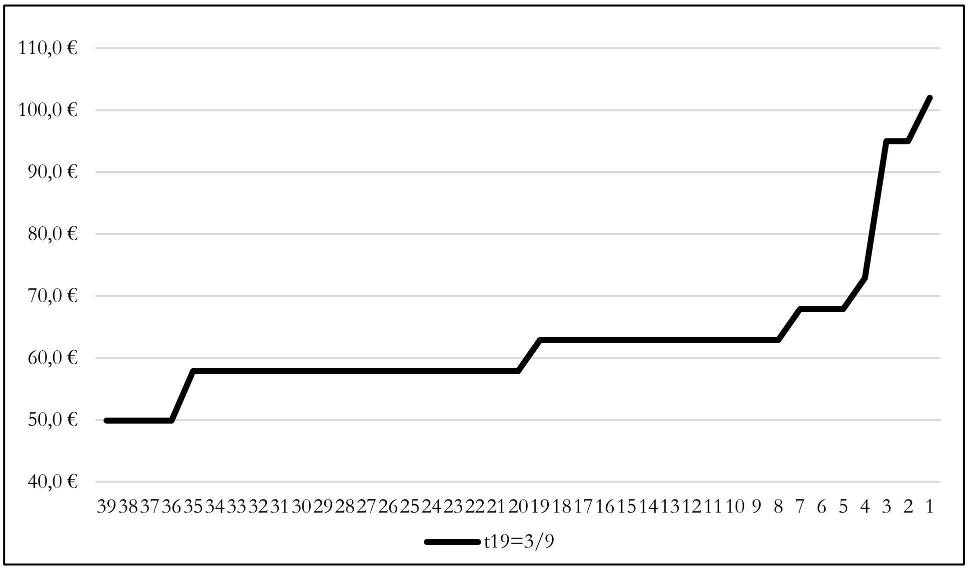


Fig. 5.3.II. Observed fare diagram of single run (Trenitalia 9634, Saturday)

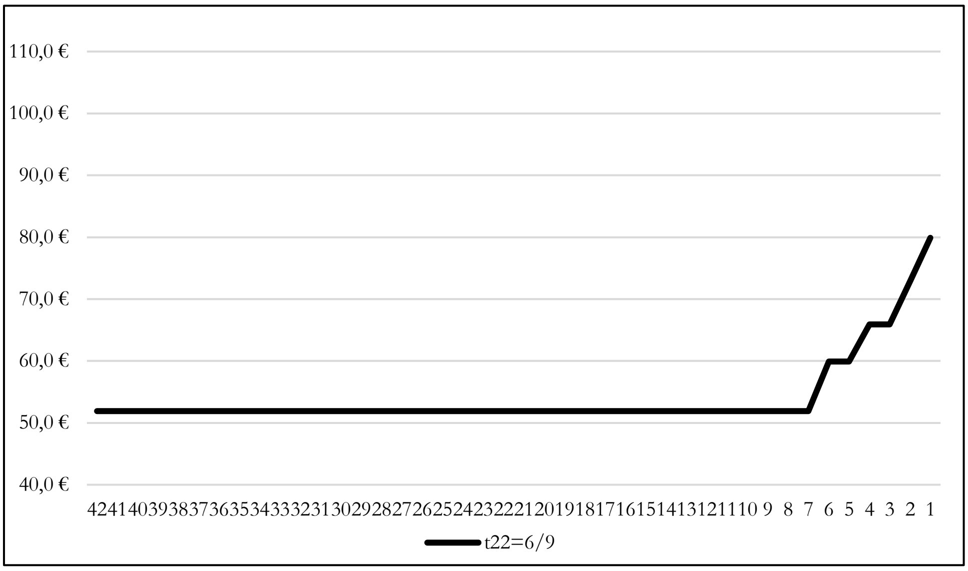


Fig. 5.3.III. Observed fare diagram of single run (NTV 6770, Tuesday)

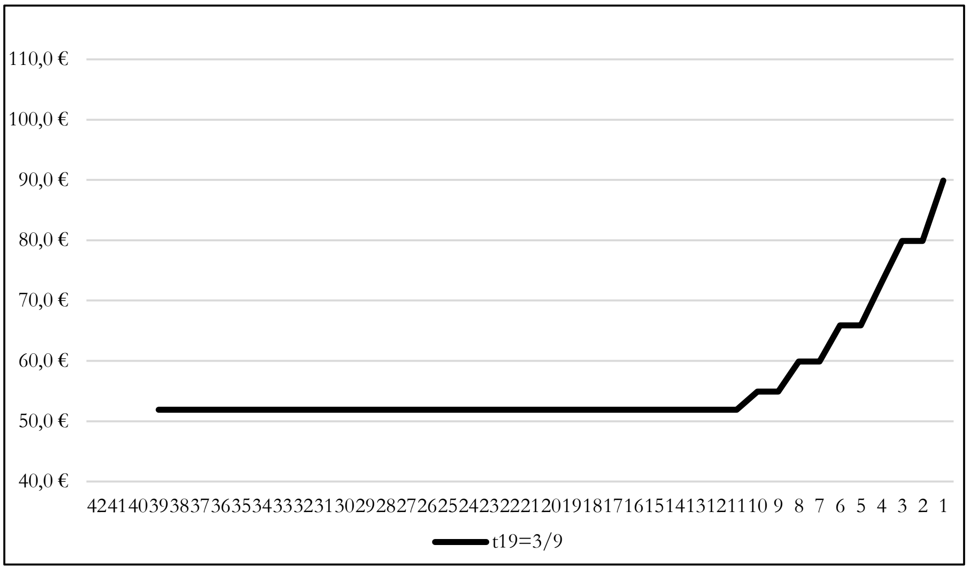


Fig. 5.3.IV. Observed fare diagram of single run (NTV 6770, Saturday)

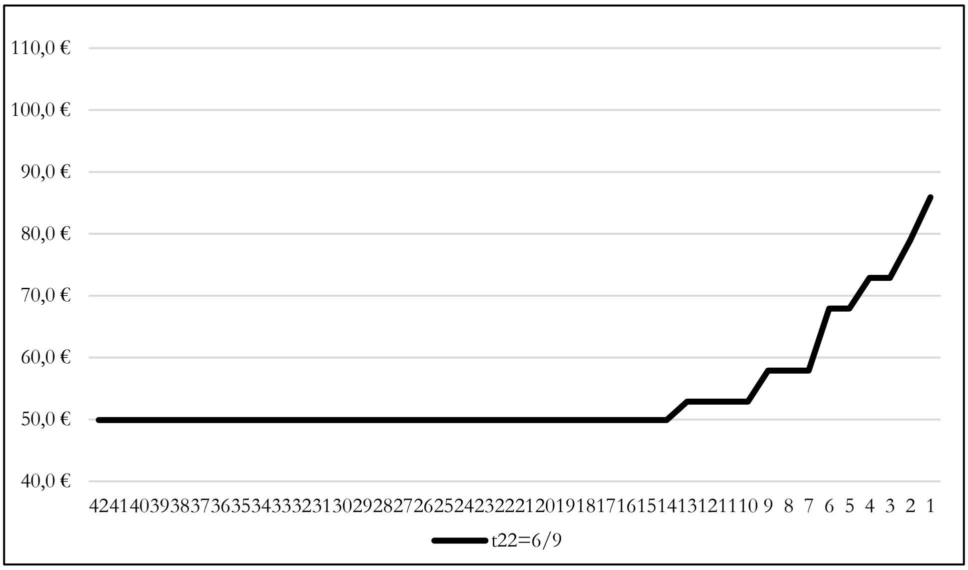


Fig. 5.3.V. Observed fare diagram of single run (Trenitalia 9631, Tuesday)

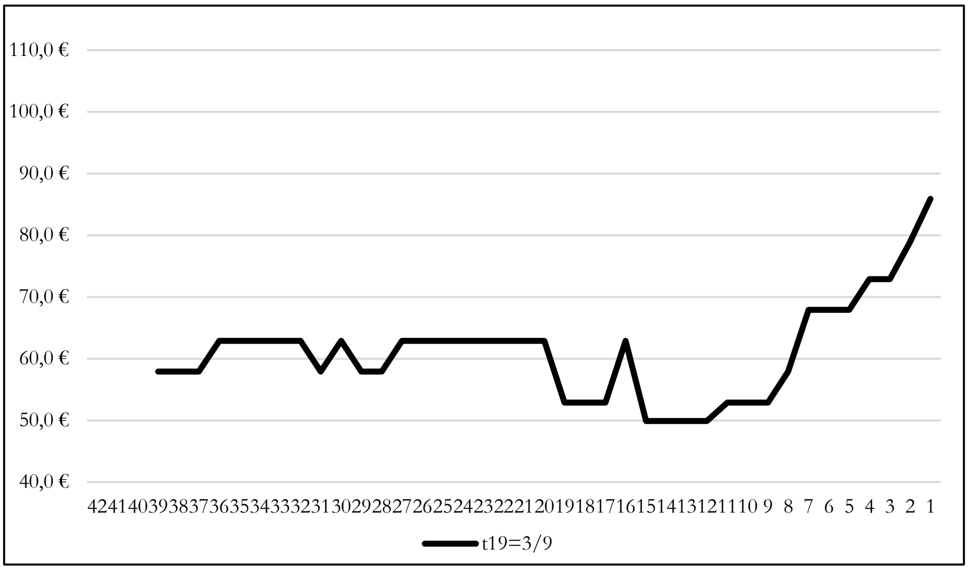


Fig. 5.3.VI. Observed fare diagram of single run (Trenitalia 9631, Saturday)

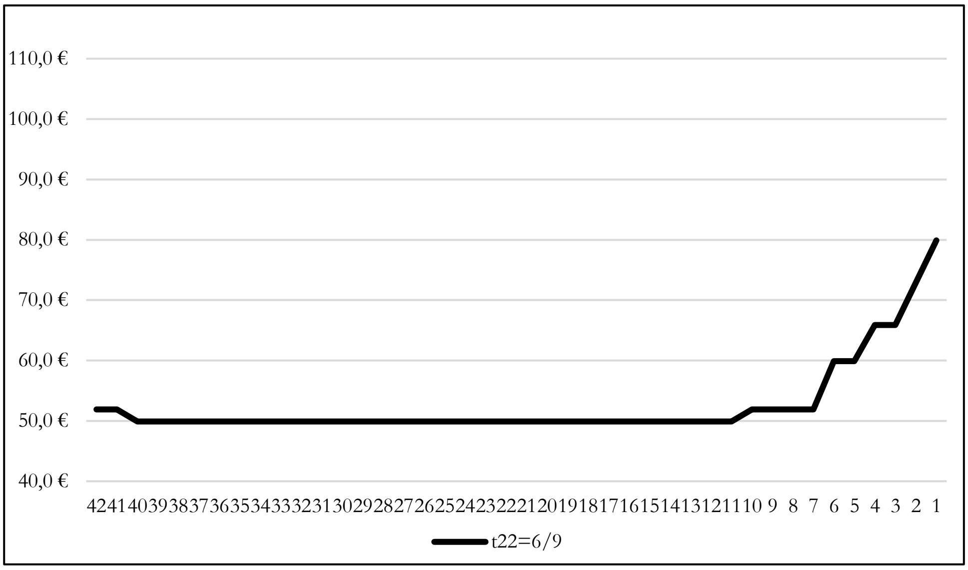


Fig. 5.3.VII. Observed fare diagram of single run (NTV 9981, Tuesday)

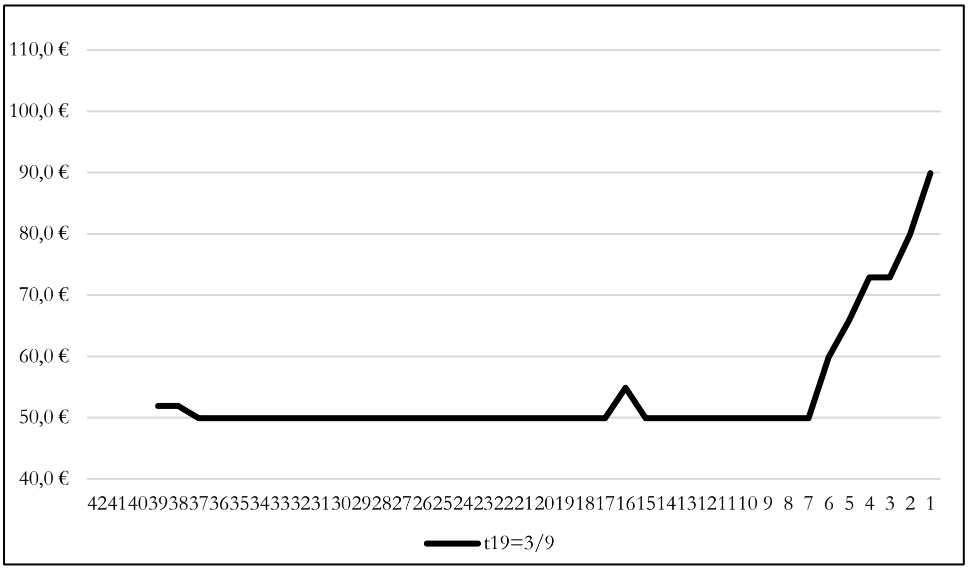


Fig. 5.3.VIII. Observed fare diagram of single run (NTV 9981, Saturday)

5.3.2. Set of runs

The figures 5.3.IX-5.3.XII show the observed diagrams of fares associated to set of runs along the direction Rome-Milan, whereas Figures 5.3.XIII-5.3.XVI the observed diagrams of fares associated to sets of runs along the direction Milan-Rome.

The observations of fares related to set of runs confirm the previous considerations of par. 5.2.1 about the different pattern in the two identified regions: stable and unstable.

The fares observation of set of runs show that the values of fares overlap in some cases and diverge in other cases. The cases where there is a high level of overlapping are depicted in Fig. 5.3.XVIII, Fig 5.3.IX and Fig. 5.3.XV, whereas the difference in values of fares are more evident in Fig.5.3.X and Fig. 5.3.XIV.

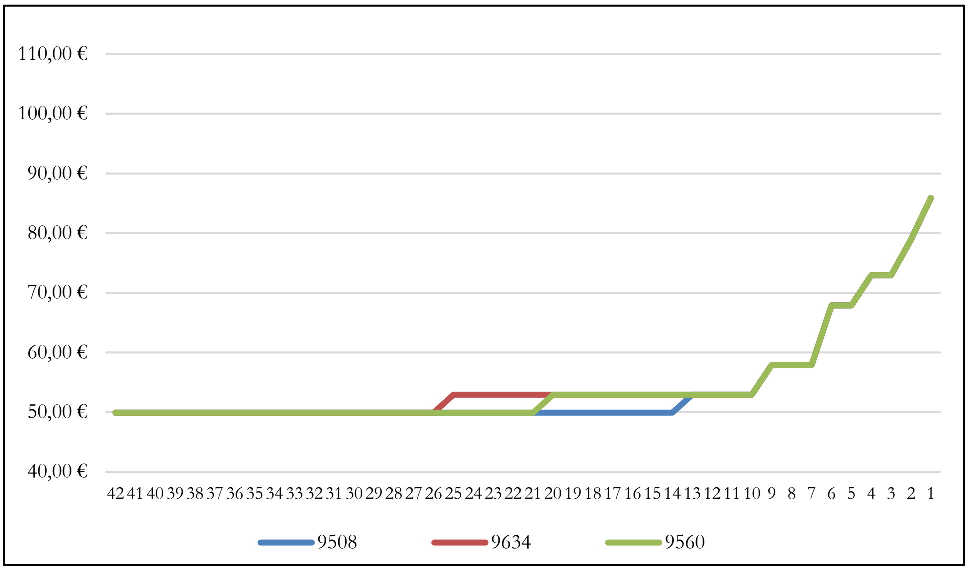


Fig. 5.3.IX. Observed fare diagrams of set of runs (Trenitalia 9508, 9634 and 9560, Tuesday)

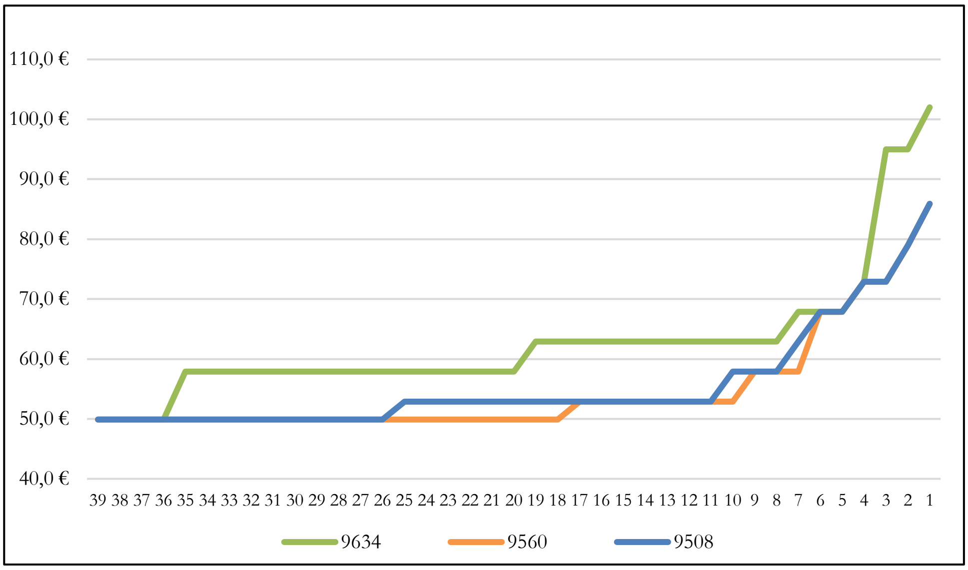


Fig. 5.3.X. Observed fare diagrams of set of runs (Trenitalia 9508, 9634 and 9560, Saturday)

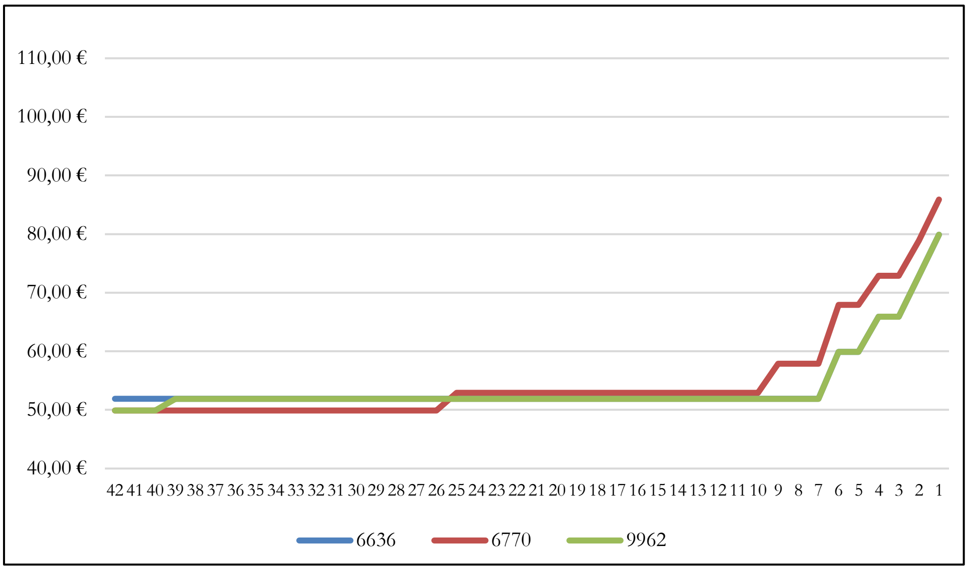


Fig. 5.3.XI. Observed fare diagrams of set of runs (NTV 6636, 6770 and 9962, Tuesday)

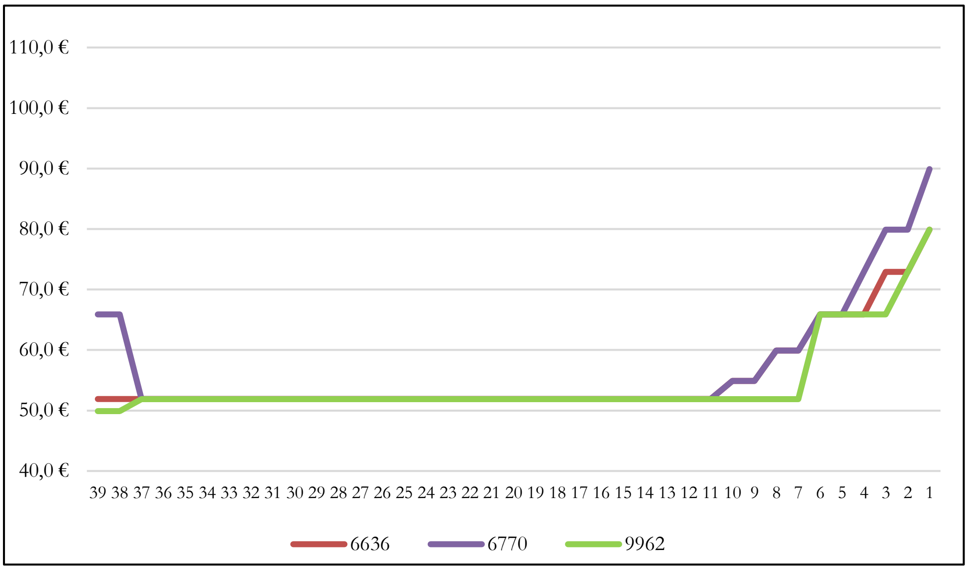


Fig. 5.3.XII. Observed fare diagrams of set of runs (NTV 6636, 6770 and 9962, Saturday)

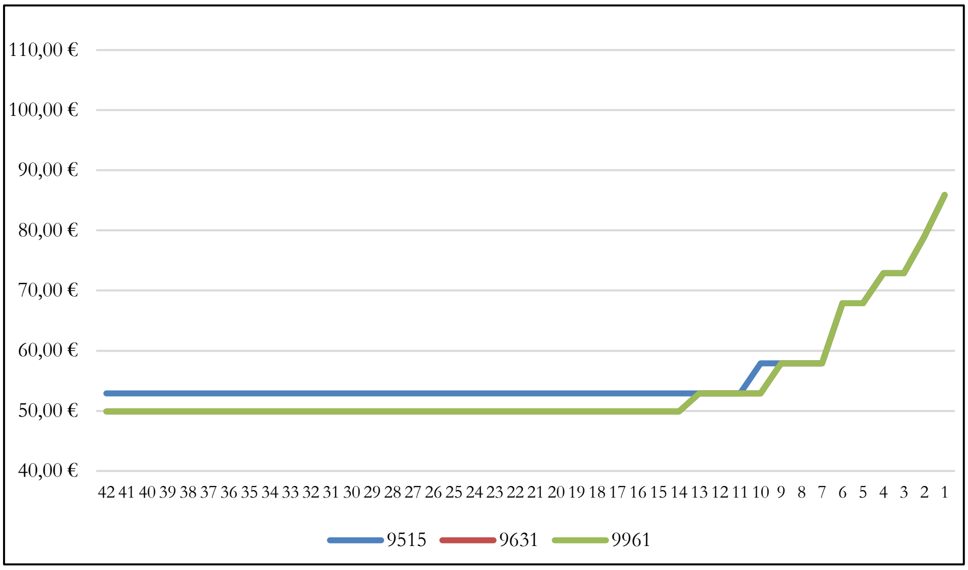


Fig. 5.3.XIII. Observed fare diagrams of set of runs (Frecciarossa 9515, 9631 and 9961, Tuesday)

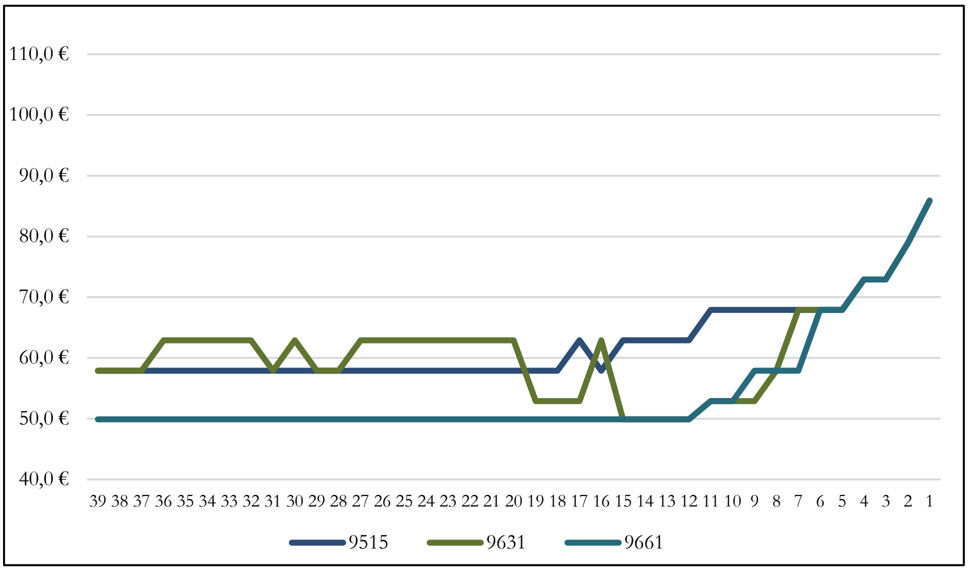


Fig. 5.3.XIV. Observed fare diagrams of set of runs (Frecciarossa 9515, 9631 and 9961, Saturday)

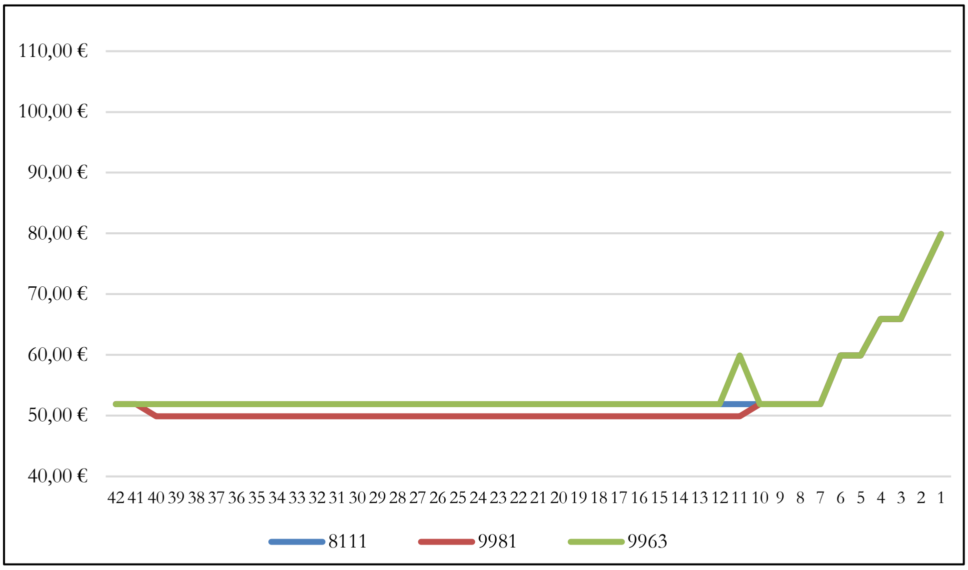


Fig. 5.3.XV. Observed fare diagrams of set of runs (NTV 8111, 9981 and 9963, Tuesday)

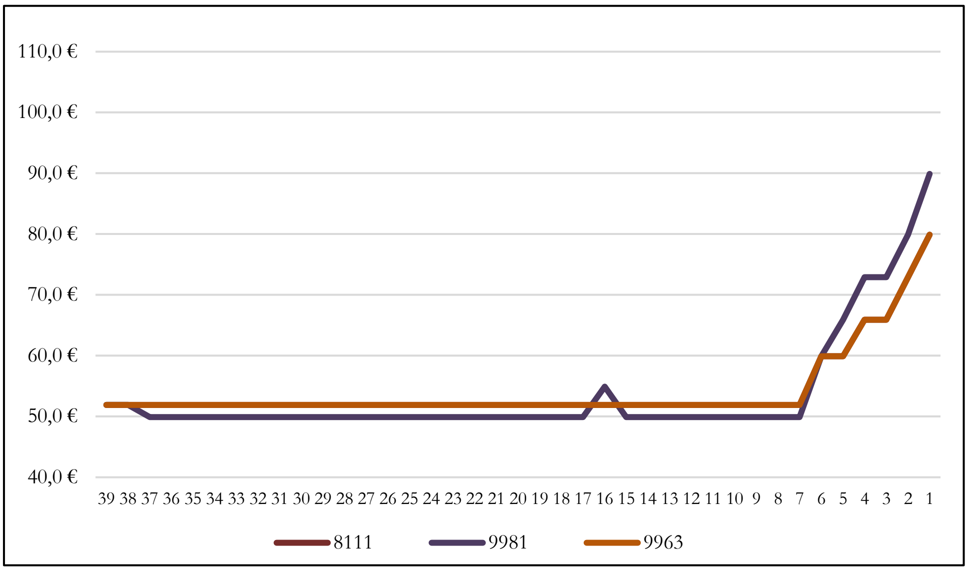


Fig. 5.3.XVI. Observed fare diagrams of set of runs (NTV 8111, 9981 and 9963, Saturday)

5.4. Fares observation in the plane: w-t

The paragraph contains observation of fares in the Cartesian plan w-t with k fixed.

The survey was conducted by consulting the websites of the railway transport companies Trenitalia, for the HSR "Frecciarossa" services and NTV for the HSR "Italo" services.

Fig. 5.4 and Tab. 5.4.a-b report the schemes of the fare observations, which are classified into:

- single run vs. set of runs;
- direction: Rome-Milan (RO-MI) and Milan-Rome (MI-RO);
- transport company: Trenitalia and NTV;
- day of ticket purchase: k=21 for the single runs and: k=7, 5 and 3 for the sets of runs,
- type of fare: "Super Economy-Standard" (SE-ST) fare for company Trenitalia company and "Low Cost-Smart" (LC-SM) fare for NTV;
- domain of observation, or cardinality |T|: 42 days for weekday vs. 38 days for weekend day.

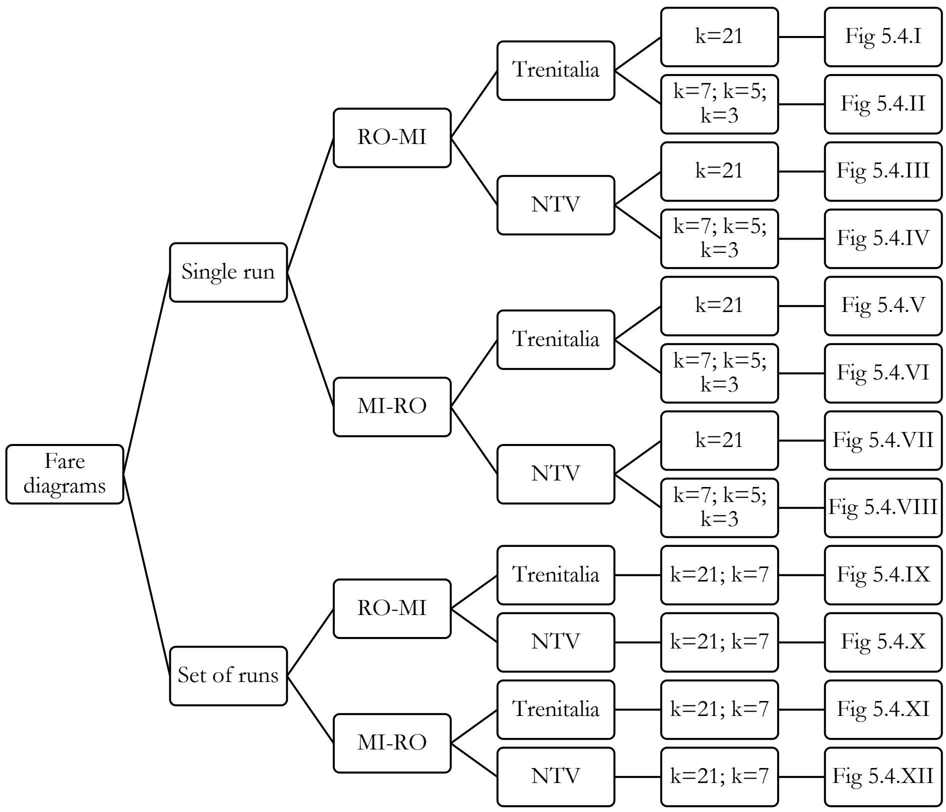


Fig. 5.21. Survey scheme of observed fare diagrams

Tab.5.4a. Survey scheme of fare diagrams (part a)

Run	Direction	Company	Figure	$\|T\|$	$\bar{k}$	g
Single run	RO-MI	Trenitalia	5.4.I	22	21	SE-ST
			5.4.II	22	7,5,3	SE-ST
		NTV	5.4.III	22	21	LC-SM
			5.4.IV	22	7,5,3	LC-SM
	MI-RO	Trenitalia	5.4.V	22	21	SE-ST
			5.4.VI	22	7,5,3	SE-ST
		NTV	5.4.VII	22	21	LC-SM
			5.4.VIII	22	7,5,3	LC-SM
Set of runs	RO-MI	Trenitalia	5.4.IX	22	21,7	SE-ST
		NTV	5.4.X	22	21,7	LC-SM
	MI-RO	Trenitalia	5.4.XI	22	21,7	SE-ST
		NTV	5.4.XII	22	21,7	LC-SM

With: SE-ST=Super Economy-Standard; LC-SM=Low Cost-Smart.

Tab.5.4b. Survey scheme of fare diagrams (part b)

Run	Direction	Company	Arrival time(*)	Time slot
Single run	RO-MI	Trenitalia	15:58	2
			15:58	2
		NTV	16:35	2
			16:35	2
	MI-RO	Trenitalia	16:15	2
			16:15	2
		NTV	15:25	2
			15:25	2
Set of runs	RO-MI	Trenitalia	9:50, 15:58, 22:50	1, 2, 3
		NTV	9:50, 15:58, 22:50	1, 2, 3
	MI-RO	Trenitalia	10:40, 16:35, $00{:}20^{+1}$	1, 2, 3
		NTV	10:40, 16:35, $00{:}20^{+1}$	1, 2, 3

With: +1=time referred to the following day; (*) see Tab. 5.5

5.4.1. Single runs

The group of Figures 5.3.I- 5.4.IV depicts the observed diagrams of fares for the single runs along the direction Rome-Milan, whereas the group of Figures 5.4.V-5.4.VIII related to the single runs along the direction Milan-Rome.

The different observed diagrams of fare have different characterizations according to the day of ticket purchasing and to the companies.

According to the day of ticket purchasing two main elements emerge:

- single run at k=21: the observed diagrams show a trend with higher values of fares in proximity of the weekend days (Saturday and Sunday) in case of Trenitalia company; while this trend is not present in the observed diagrams of NTV companies;

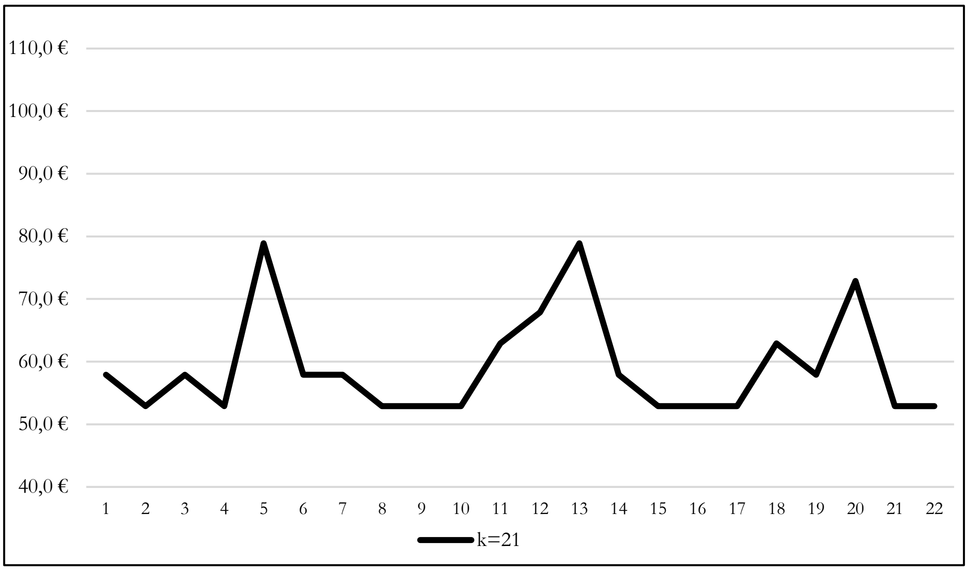
- single run at k= 7, 5, 3: the observed diagrams show higher values of fares as the day of trip approaching (value of k decreasing) along the whole domain considered for both companies Trenitalia and NTV.

In one case (Fig. 5.4.II and Fig. 5.4.IV) there is a decreasing trend of the fares along the domain analyzed, that needs to be further investigated.

Tab 5.5 shows the average values of fares at the day of ticket purchase analyzed and the presence of peaks of values in the weekend days.

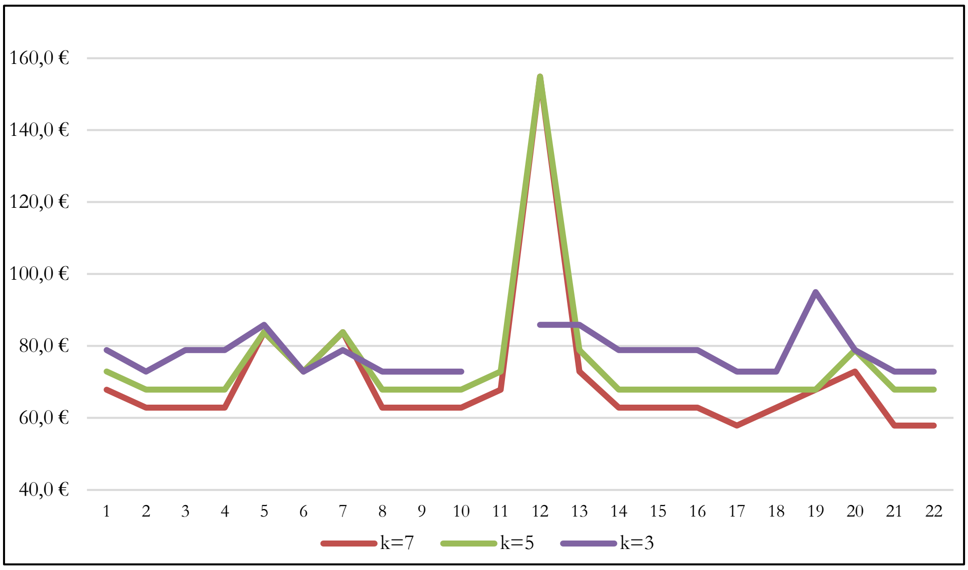
Tab.5.5. Characteristics of observed fare diagrams: average values of fare at different days of ticket purchasing

Figure	Average fare k=21 [€]	Peaks at weekend days (Yes/No)	Average fare at k=7 [€]	Average fare at k=5 [€]	Average fare at k=3 [€]
5.4.I	59.1	Yes	-	-	-
5.4.II	-	No	70.4	75.0	78.1
5.4.III	52.4	No	-	-	-
5.4.IV	-	No	56.7	61.6	72.4
5.4.V	51.9	Yes	-	-	-
5.4.VI	-	No	62.9	68.5	76.4
5.4.VII	50.9	No	-	-	-
5.4.VIII	-	No	56.4	62.0	73.0



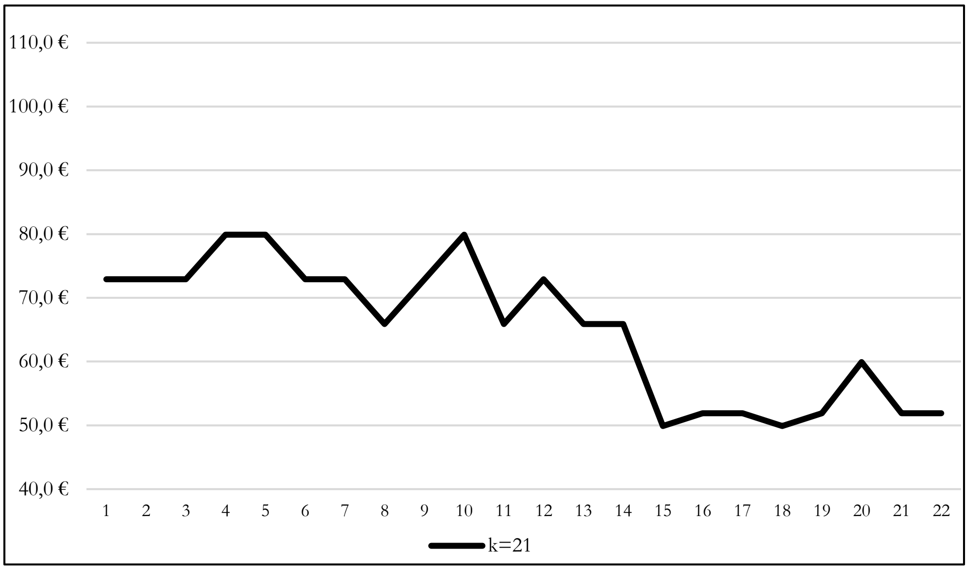
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.I. Observed fare diagram of single day of purchase (Trenitalia 9634, k=21)



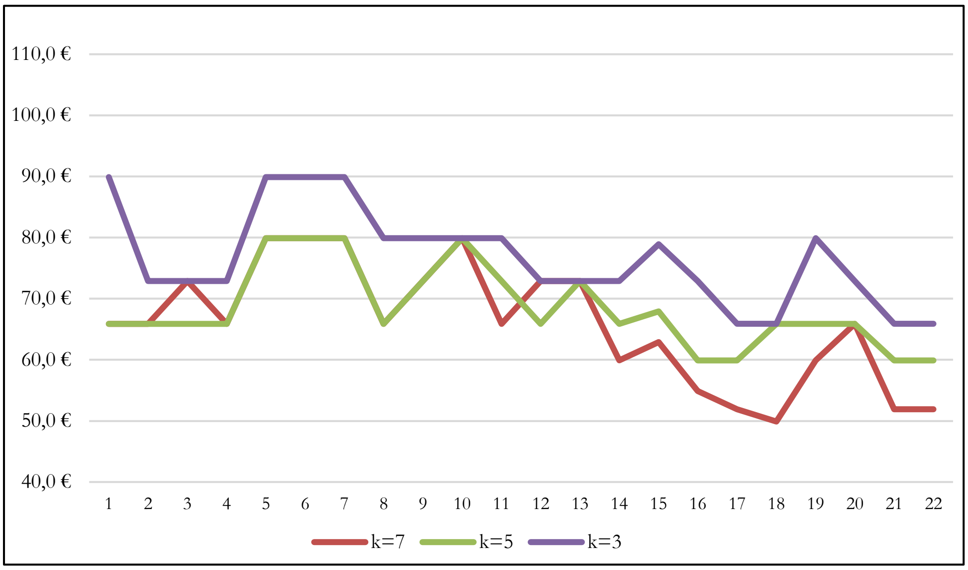
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.II. Observed fare diagrams of group of days of purchase (Trenitalia 9634, k=7, 5, 3)



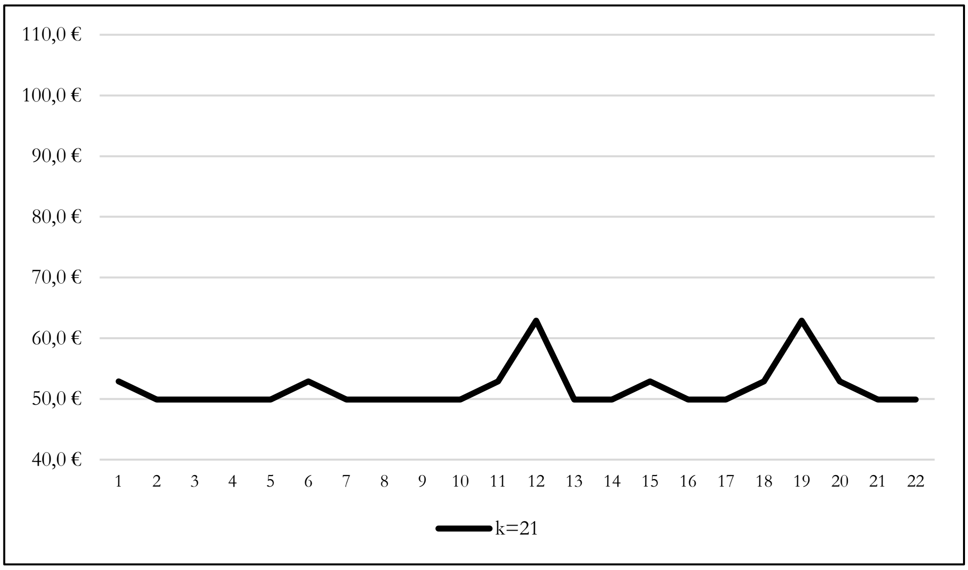
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.III. Observed fare diagram of single day of purchase (NTV 6770, k=21)



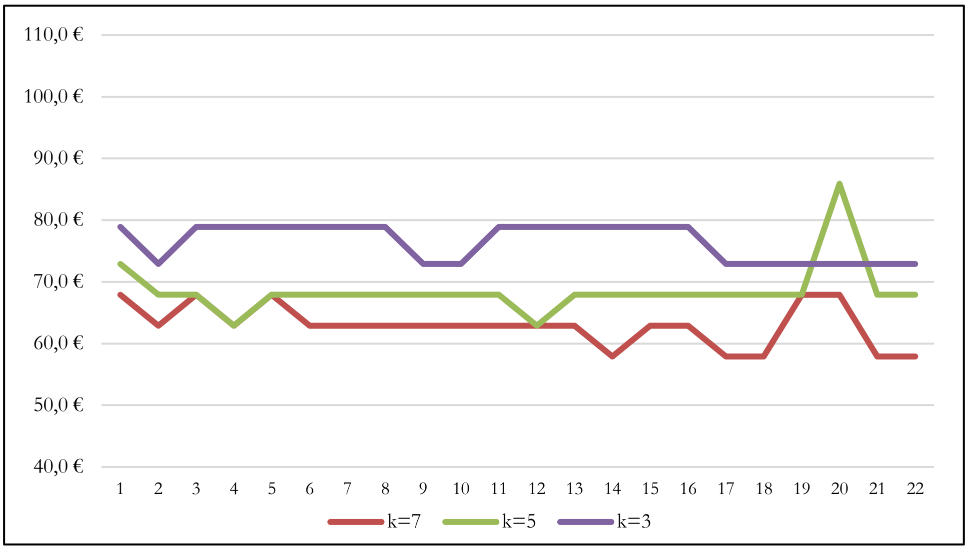
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.IV. Observed fare diagrams of group of days of purchase (NTV 6770, k=7, 5, 3)



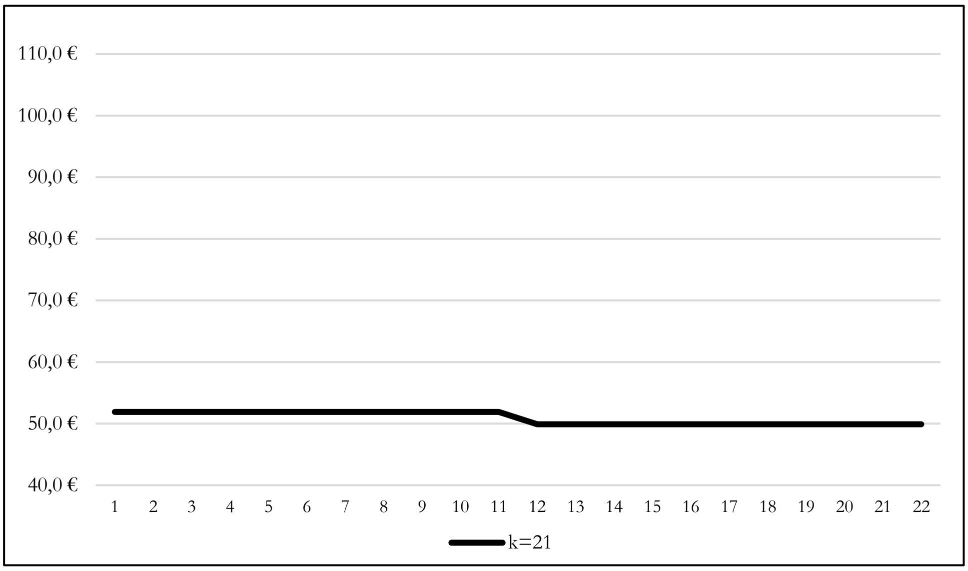
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.V. Observed fare diagram of single day of purchase (Trenitalia 9631, k=21)



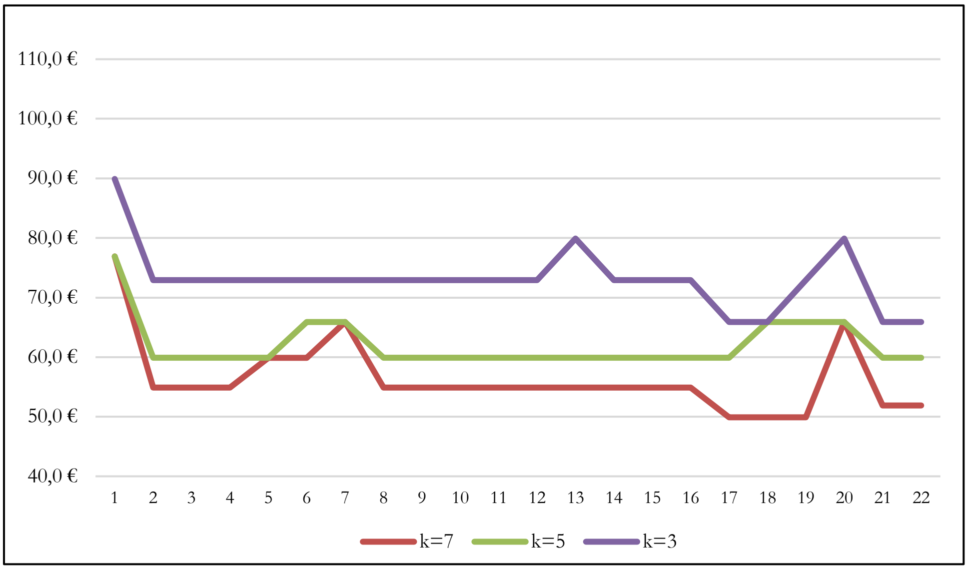
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.VI. Observed fare diagram of group of days of purchase (Trenitalia 9631, k=7, 5, 3)



t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.VII. Observed fare diagram of single day of purchase (NTV 9981, k=21)



t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
t	16/8	17/8	18/8	19/8	20/8	21/8	22/8	23/8	24/8	25/8	26/8	27/8	28/8	29/8	30/8	31/8	01/9	02/9	03/9	04/9	05/9	06/9
t	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar
K	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
k	26/7	27/7	28/7	29/7	30/7	31/7	1/8	2/8	3/8	4/8	5/8	6/8	7/8	8/8	9/8	10/8	11/8	12/8	13/8	14/8	15/8	16/8
k	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar	Mer	Gio	Ven	Sab	Dom	Lun	Mar

Fig. 5.4.VIII. Observed fare diagram of group of days of purchase (NTV 9981, k=7, 5, 3)

5.5. Pilot analysis on a single run

The paragraph describes the application of the first two steps of the framework presented in paragraph 5.1, namely: “definition of domain” and “fare observation”.

Phase 1: definition of domain. The analysis of domain is constituted by the following elements.

- Number and type of HSR fares. The "Super Economy-Standard" fare is recorded (where "Super Economy", in a row, is the fare class and "Standard", in a column, is the fare service level) supplied by Trenitalia, which is the cheapest ticket available among the cheapest (in a row) class fares offered each day. Fares under special conditions are excluded, for example: elderly or young passengers.
- Day of trip. The pilot case study considers a single day "t", assuming that a traveler chooses to make one trip per day t=03/09/2022 from Rome to Milan.

- Day of ticket purchasing. The observed set of ticket purchasing days "p", concerns the period from July 26, 2022 to September 2, 2022, in which the survey of the selected fare on the selected day of trip is conducted.
- Distance between day of trip (t) and day of ticket purchasing (p). The variable "k" is calculated as the difference between the day of trip and the day of ticket purchasing (k=t-p).
- Number of HSR runs. The run r=9634 of the Rome-Milan route, operated by the Trenitalia company belonging to the 6:00-9:00 time slot, which considers peak of traffic volume, was analyzed. Run 9634 is identified by the minimum intermediate stop criteria; therefore, it is considered a representative run in the defined time window.

Phase 2: fare observation. Fares were directly observed on the Trenitalia company website in relation to the definition of domain.

The above-mentioned fare (“Super Economy-Standard”) is the objective of survey during the identified period. In case of absence of the selected fare, on a day k, it is detected the next fare belonging to the same class (for example, "Super Economy" or "Economy"), and furthermore in case of absence of the last fare belonging to the same class, the cheapest fare of the higher class was detected (e.g. "Super Economy-Premium").

Diagram in Fig. 5.4 shows the structure of fares detected with the above-mentioned criteria along variable k considering the day of trip t=03 September 2022. If it is defined a ticket purchase date, p=August 16, 2022, the distance between the given ticket purchase date and the given day of trip, then the interval k=3/09-16/08=21.

Diagram of Fig. 5.5 confirms also in this case that the fare structure presents two main patterns. A stable pattern is more distant from the day of trip t=03/09/2022 where ticket prices are constant. An unstable pattern is closer to the day of trip: t=03/09/2022, where the evolution of fares shows an increasing positive gradient due to the variation in ticket fares.

Moreover, it is possible to identify a threshold, corresponding to one or a group of days of ticket purchase k, which defines the border between the two regions indicated above.

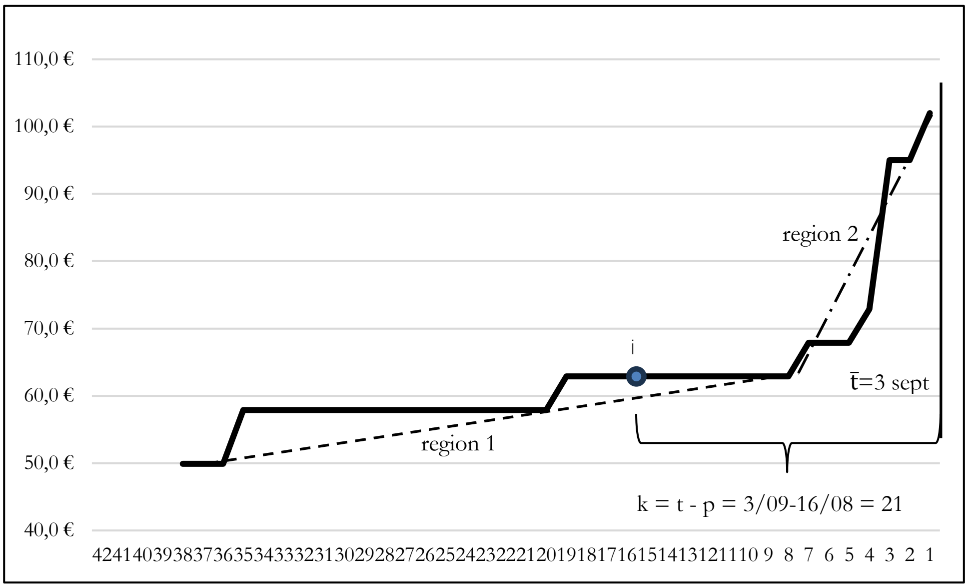


Fig. 5.5. Pilot observed fare diagram (t=03/09/2022)

The diagram of Fig. 5.6 shows the trend of the fares detected along the domain of the days of trip taking into account day of ticket purchasing: k= 21. The diagram indicates a trend whereby the fare values increase in the proximity of weekend days and decrease during working days. This trend needs to be further studied with the support of a wider range of observations.

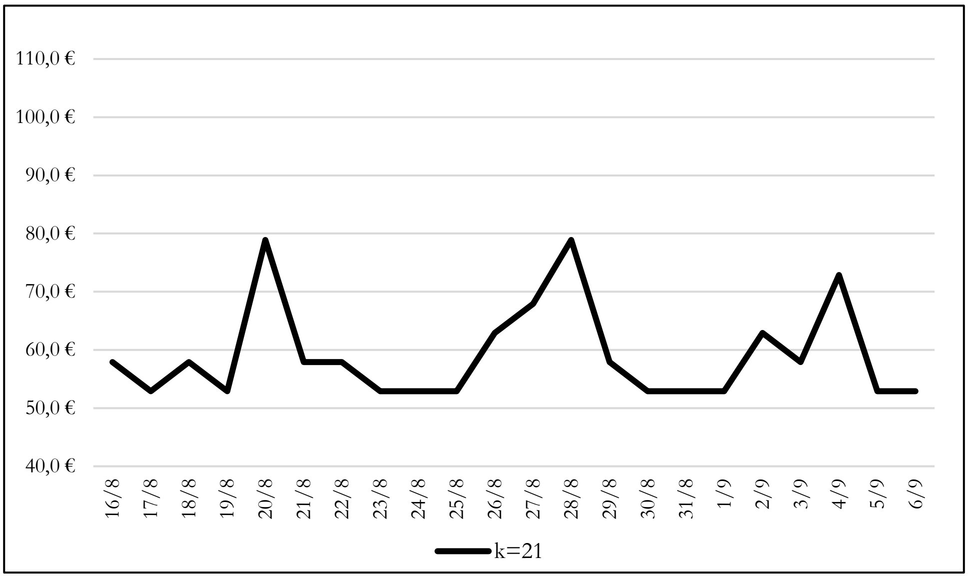


Fig. 5.6. Pilot observed fare diagram (k=21)

Finally, Tab.5.6 contains some descriptive statistics calculated for the time series presented in Fig. 5.5 and Fig. 5.6.

Tab.5.6. Descriptive statistics of fares (see Fig. 5.5 and Fig 5.6)

	Figure 5.5 (k)	Figure 5.6 (t)
n.obs	37	22
Avg	64.8€	59.1€
Var	25.2	70.2
Min	62.9€	52.9€
Max	85.9€	78.9€
avrΔ (27/07-26/08)	1.35	n.a.
avrΔ (27/08-02/09)	-	0.2
avrWE	66.2€	63.9€
avrWD	64.1€	57.3€

With: n.obs= number of observations; Avg=average; Var=variance; Min=minimum; Max=maximum; avrΔ=average gradient; avrWE= average in weekend days; avrWD= average in workdays; n.a.=not available.

5.6. Fare modelling: specification and calibration of fare function: w=w(t)

The paragraph regards the specification and calibration of a fare model w=*w*(k). The specification of the fare model and of the estimator is reported in par. 5.6.1. Therefore, the calibrated parameters of the models related to the single run and of sets of runs are presented respectively in par. 5.6.2 and in par. 5.6.3.

5.6.1. Model specification

The general functional form of the fare model is the following:

$$\mathrm{w} = w\,(\mathrm{k}'; \boldsymbol{\alpha}) + \varepsilon \qquad (5.1)$$

with:

w=dependent variable

k'= independent variable

$\boldsymbol{\alpha}$=vector of unknown parameters

w( )=fare function

ε=error term

where

$$\mathrm{k}' = |\mathrm{k} - \overline{\mathrm{K}}|$$

is a variable transformation of the variable k, day of ticket purchasing.

The values of the domain of variable k are two:

- $\overline{K} = 42$ for t= 6/9/22
- $\overline{K} = 38$ for t=3/9/22

The domains of the different variables are reported in the following.

The independent variable k belongs to the set of natural numbers and the dependent variable w belongs to the set of the real numbers:

$$w(k): \mathbb{N} \rightarrow \mathbb{R}$$

where k and w may assume positive values:

$$k \in [0;\ +\infty];\ w \in [0; +\infty]$$

The independent variable k' belongs to the set of natural numbers and the dependent variable w belongs to the set of the real numbers.

$$w(k'): \mathbb{N}' \rightarrow \mathbb{R}$$

where k' and w may assume positive values:

$$k' \in [0, +\infty];\ \mathbb{R} \quad \in [0, +\infty]$$

The general function, of eq. (5.1), may assume different specifications, reported in Tab. 5.7 and 5.8 below.

Tab.5.7. Specifications of the model for single run

Model	Specification
	PHASE 1
0	$w = \alpha_0$
1	$w = \alpha_0 + \alpha_1 \cdot k'$
2	$w = \alpha_0 + \alpha_2 \cdot k'^2$
3	$w = \alpha_0 + \alpha_3 \cdot k'^3$
4	$w = \alpha_0 + \alpha_4 \cdot k'^4$
5	$w = \alpha_0 + \alpha_5 \cdot k'^5$
	PHASE 2
6	$w = \alpha_0 + \alpha_1 \cdot k' + \alpha_2 \cdot k'^2$
7	$w = \alpha_0 + \alpha_2 \cdot k'^2 + \alpha_3 \cdot k'^3$
8	$w = \alpha_0 + \alpha_3 \cdot k'^3 + \alpha_4 \cdot k'^4$
9	$w = \alpha_0 + \alpha_4 \cdot k'^4 + \alpha_5 \cdot k'^5$
	PHASE 3
10	$w = \alpha_0 + \alpha' \cdot k'^{\alpha_1}$
11	$w = \alpha_0 + \alpha' \cdot k'^{\alpha_1} + \alpha'' \cdot k'^{\alpha_2}$

Tab.5.8. Specifications of the model for sets of runs

Model	Specification
	PHASE 4
12	$w = \alpha_0 + \alpha_2 \cdot k'^2$
13	$w = \alpha_0 + \alpha_3 \cdot k'^3$
14	$w = \alpha_0 + \alpha_4 \cdot k'^4$
15	$w = \alpha_0 + \alpha_5 \cdot k'^5$
	PHASE 5
16	$w = \alpha_0 + \alpha_{2,r1} \cdot k'^2 + \alpha_{2,r2} \cdot k'^2 + \alpha_{2,r3} \cdot k'^2$
17	$w = \alpha_0 + \alpha_{3,r1} \cdot k'^3 + \alpha_{3,r2} \cdot k'^3 + \alpha_{3,r3} \cdot k'^3$
18	$w = \alpha_0 + \alpha_{4,r1} \cdot k'^4 + \alpha_{4,r2} \cdot k'^4 + \alpha_{4,r3} \cdot k'^4$
19	$w = \alpha_0 + \alpha_{5,r1} \cdot k'^5 + \alpha_{5,r2} \cdot k'^5 + \alpha_{5,r3} \cdot k'^5$
20	$w = \alpha_0 + \alpha' \cdot k'^{\alpha}$

The vector of parameters, $\boldsymbol{\alpha}$, is calibrated by means of the Ordinary Least Squares (OLS) method. The method operates by identifying the vector $\widehat{\boldsymbol{\alpha}}$, among the configurations of vector $\boldsymbol{\alpha}$, that minimizes the sum of the squares of the deviations between the observed values of the fares (independent variable) and the values of fares estimated by the models, as reported in Tab. 5.7 and 5.8.

The optimization model is specified by equations (5.2) and (5.3).

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} S\,(\boldsymbol{\alpha}) \tag{5.2}$$

with:

S=objective function;

$\widehat{\boldsymbol{\alpha}}$=vector of calibrated parameters.

$$S\,(\boldsymbol{\alpha}) = \sum_k \left(w_{obs,k} - w_{est,k}(\boldsymbol{\alpha})\right)^2 \tag{5.3}$$

with:

$w_{obs,k}$= observed fare;

$w_{est,k}(\boldsymbol{\alpha})$=estimated fare (model)

The validation of the models is performed by the coefficient of determination (ρ^2), which is the proportion of the variation in the dependent variable that is predictable from the independent variable(s).

$$\rho^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{5.4}$$

with: SS_{res}=residual sum of square

$$SS_{res} = \sum_k \left(w_{obs,k} - w_{est,k}(\hat{\alpha})\right)^2 \tag{5.5}$$

SS_{tot}=total sum of square

$$SS_{tot} = \sum_k \left(w_{obs,k} - \overline{w}_{obs,k}\right)^2 \tag{5.6}$$

Average value of the fares

$$\overline{w}_{obs,k} = \frac{1}{n}\sum_{i=1}^{n} w_{obs,k} \tag{5.7}$$

5.6.2. Calibrated parameters: single run

The paragraph reports the calibrated parameters of the specified fare models (Tab. 5.7) and clustered in 3 calibration phases.

a. Calibration phase 1

Phase 1 concerns the calibration process of the five models 0-5 specified in the previous paragraph in Tab. 5.7.

Tab 5.9 reports the calibrated parameters of model 0. The value of the calibrated parameter α_0 lies between 52.7 € and 63.1 €. Fig. 5.7 depicts the values of the observed fares already plotted in Fig. 5.7 and of the fare model 0.

Tab.5.9. Calibrated parameters of model 0

Fig	run	Arrival time	Company	t	|K|	α_0
5.2.I	9634	15:58	Trenitalia	6/9/22	42	55.2
5.2.II	9634	15:58		3/9/22	38	63.1
5.2.III	6770	16:35	NTV	6/9/22	42	54.2
5.2.IV	6770	16:35		3/9/22	38	56.2
5.2.V	9631	16:15	Trenitalia	6/9/22	42	54.4
5.2.VI	9631	16:15		3/9/22	38	61.1
5.2.VII	9981	15:25	NTV	6/9/22	42	52.7
5.2.VIII	9981	15:25		3/9/22	38	53.8

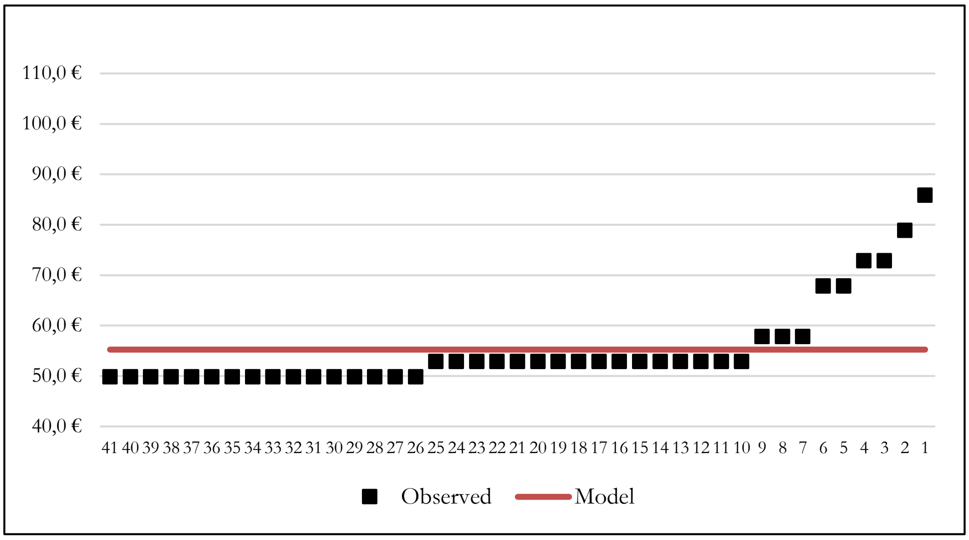


Fig. 5.7. Calibrated parameters of model 0

Tab 5.10 reports the calibrated parameters of model 1. The values of the calibrated parameter α_0 lies between 43.6 € and 57.5 €, and the value of α_1 between 1.87E-01 and 7.65E-01. The value of the coefficient of determination (ρ^2) oscillates from 0.33 to 0.57. Fig. 5.8 presents the plotted values of the observed fares and of the fare curve estimated by model 1.

Tab.5.10. Calibrated parameters of model 1

Fig	run	Arrival time	Company	t	|K|	α_0	α_1	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	43.9	5.43E-01	0.56
5.2.II	9634	15:58		3/9/22	38	48.2	7.65E-01	0.57
5.2.III	6770	16:35	NTV	6/9/22	42	47.9	2.96E-01	0.33
5.2.IV	6770	16:35		3/9/22	38	45.0	5.74E-01	0.46
5.2.V	9631	16:15	Trenitalia	6/9/22	42	43.2	5.33E-01	0.50
5.2.VI	9631	16:15		3/9/22	38	57.5	1.87E-01	0.07
5.2.VII	9981	15:25	NTV	6/9/22	42	45.3	3.50E-01	0.38
5.2.VIII	9981	15:25		3/9/22	38	43.6	5.22E-01	0.37

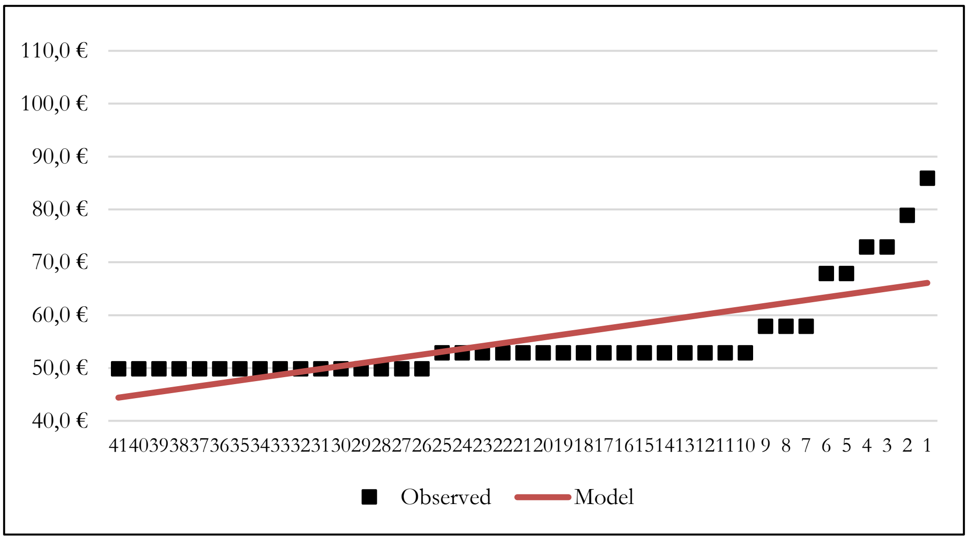


Fig. 5.8. Calibrated parameters of model 1

Tab 5.11 reports the calibrated parameters of model 2. The values of the calibrated parameter α_0 lie between 45.9 € and 57.6 €, and the value of α_2 between 7.05E-03 and 2.08E-02. The value of the coefficient of determination (ρ^2) oscillates from 0.16 to 0.73. Fig. 5.9 presents the plotted values of the observed fares and of fares estimated by model 2.

Tab.5.11. Calibrated parameters of model 2

Fig	run	Arrival time	Company	t	|K|	α_0	α_2	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	46.9	1.43E-02	0.73
5.2.II	9634	15:58		3/9/22	38	52.8	2.08E-02	0.68
5.2.III	6770	16:35	NTV	6/9/22	42	49.3	8.31E-03	0.49
5.2.IV	6770	16:35		3/9/22	38	47.8	1.69E-02	0.65
5.2.V	9631	16:15	Trenitalia	6/9/22	42	45.9	1.45E-02	0.69
5.2.VI	9631	16:15		3/9/22	38	57.6	7.05E-03	0.16
5.2.VII	9981	15:25	NTV	6/9/22	42	47.0	9.72E-03	0.55
5.2.VIII	9981	15:25		3/9/22	38	45.9	1.56E-02	0.53

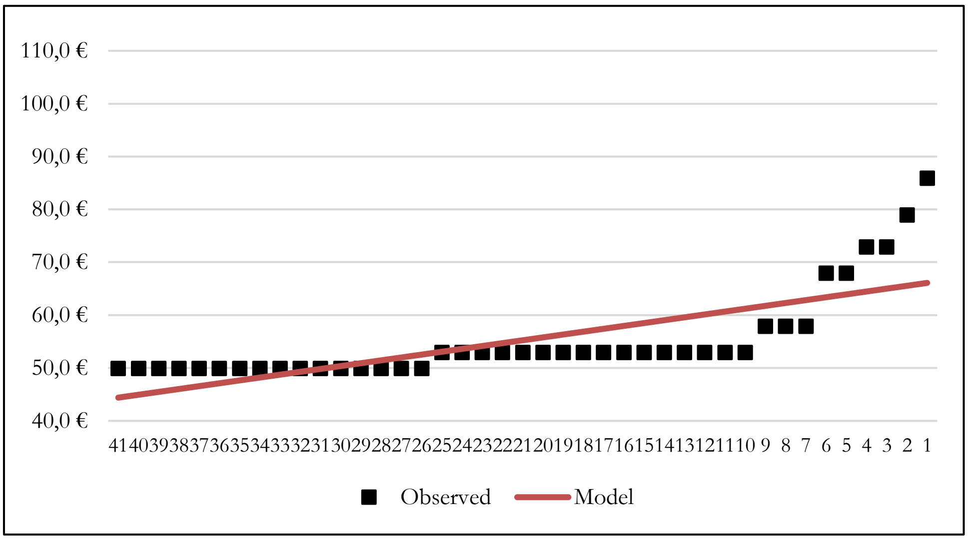


Fig. 5.9. Calibrated parameters of model 2

Tab 5.12 reports the calibrated parameters of model 3. The values of the calibrated parameter α_0 is between 46.9 € and 57.6 €, and the value of α_3 lie between 2.36E-04 and 5.95E-04. The value of the coefficient of determination (ρ^2) oscillates from 0.25 to 0.83. Fig. 5.10 presents the plotted values of the observed fares and of the fare estimated by model 3.

Tab.5.12. Calibrated parameters of model 3

Fig	run	Arrival time	Company	t	|K|	α_0	α_3	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	48.3	3.85E-04	0.83
5.2.II	9634	15:58		3/9/22	38	54.6	5.95E-04	0.76
5.2.III	6770	16:35	NTV	6/9/22	42	49.9	2.36E-04	0.61
5.2.IV	6770	16:35		3/9/22	38	48.9	5.06E-04	0.78
5.2.V	9631	16:15	Trenitalia	6/9/22	42	47.2	3.97E-04	0.82
5.2.VI	9631	16:15		3/9/22	38	57.6	2.46E-04	0.25
5.2.VII	9981	15:25	NTV	6/9/22	42	47.7	2.73E-04	0.68
5.2.VIII	9981	15:25		3/9/22	38	46.9	4.72E-04	0.66

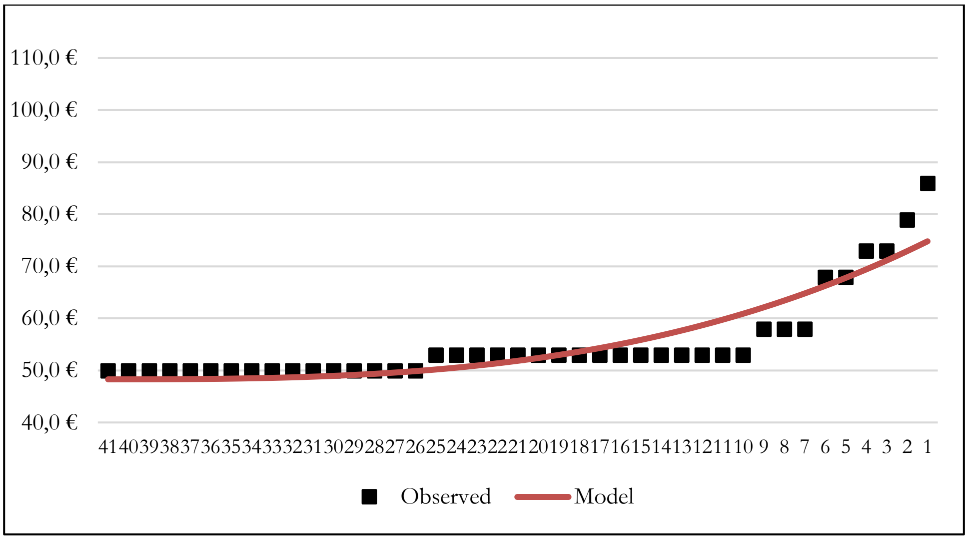


Fig. 5.10. Calibrated parameters of model 3

Tab 5.13 reports the calibrated parameters of model 4. The values of the calibrated parameter α_0 lie between 47.6 € and 57.6 €, and the value of α_4 lie between 6.49E-06 and 1.69E-05. The value of the coefficient of determination (ρ^2) oscillates from 0.35 to 0.89. Fig. 5.11 presents the plotted values of the observed fares and of the fare estimated by model 4.

Tab.5.13. Calibrated parameters of model 4

Fig	run	Arrival time	Company	t	|K|	α_0	α_4	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	49.1	1.02E-05	0.89
5.2.II	9634	15:58		3/9/22	38	55.6	1.69E-05	0.80
5.2.III	6770	16:35	NTV	6/9/22	42	50.3	6.49E-06	0.71
5.2.IV	6770	16:35		3/9/22	38	49.7	1.47E-05	0.87
5.2.V	9631	16:15	Trenitalia	6/9/22	42	47.9	1.06E-05	0.89
5.2.VI	9631	16:15		3/9/22	38	57.6	7.92E-06	0.35
5.2.VII	9981	15:25	NTV	6/9/22	42	48.2	7.46E-06	0.78
5.2.VIII	9981	15:25		3/9/22	38	47.6	1.39E-05	0.75

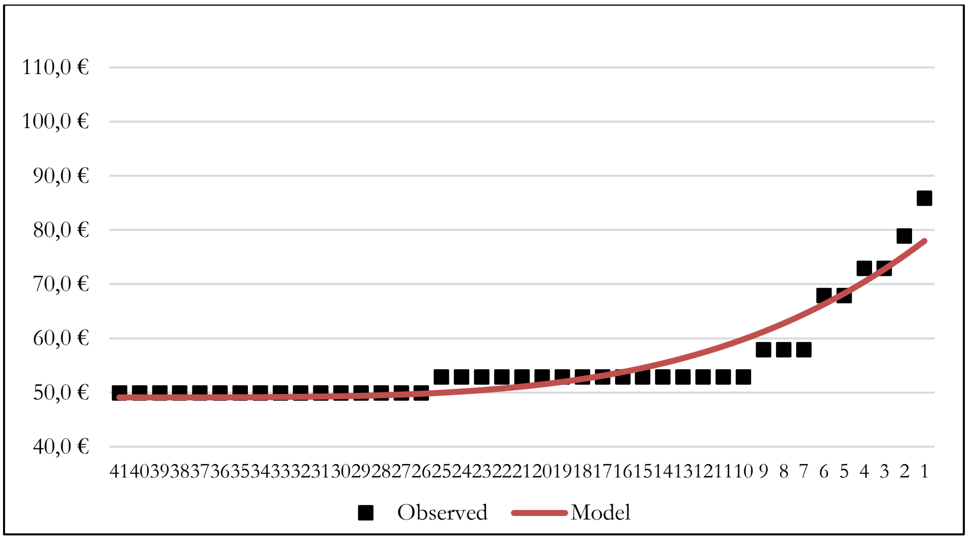


Fig. 5.11. Calibrated parameters of model 4

Tab 5.14 reports the calibrated parameters of model 5. The values of the calibrated parameter α_0 lie between 48.1 € and 57.7 €, and the value of α_5 lie between 1.75E-07 and 4.76E-07. The value of the coefficient of determination (ρ^2) oscillates from 0.43 to 0.94. Fig. 5.12 presents the plotted values of the observed fares and of the fare estimated by model 5.

Tab.5.14. Calibrated parameters of model 5

Fig	run	Arrival time	Company	t	|K|	α_0	α_5	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	49.7	2.67E-07	0.93
5.2.II	9634	15:58		3/9/22	38	56.4	4.76E-07	0.84
5.2.III	6770	16:35	NTV	6/9/22	42	50.5	1.75E-07	0.78
5.2.IV	6770	16:35		3/9/22	38	50.3	4.19E-07	0.92
5.2.V	9631	16:15	Trenitalia	6/9/22	42	48.6	2.79E-07	0.94
5.2.VI	9631	16:15		3/9/22	38	57.7	2.42E-07	0.43
5.2.VII	9981	15:25	NTV	6/9/22	42	48.5	1.99E-07	0.85
5.2.VIII	9981	15:25		3/9/22	38	48.1	4.01E-07	0.82

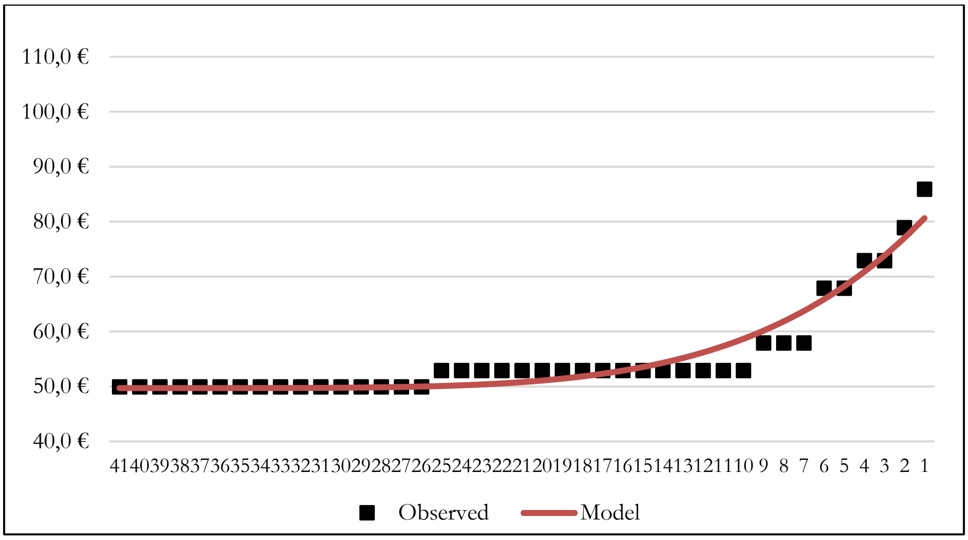


Fig. 5.12. Calibrated parameters of model 5

b. Calibration phase 2

Phase 2 concerns the calibration of models 6-9, which have been specified in the previous paragraph, as reported in Tab. 5.8.

Tab 5.15 reports the calibrated parameters of model 6. The values of the calibrated parameter α_0 lie between 45.9 € and 57.6 €, the value of α_1 lie between 7.05E-03 and 2.08E-02 and the value of α_2 results in each case 0. The value of the coefficient of determination (ρ^2) oscillates from 0.16 to 0.69. Fig. 5.13 presents the plotted values of the observed fares and of the fare estimated by model 6.

Tab.5.15. Calibrated parameters of model 6

Fig	run	Arrival time	Company	t	|K|	α_0	α_1	α_2	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	46.9	1.43E-02	0	0.73
5.2.II	9634	15:58		3/9/22	38	52.7	2.08E-02	0	0.68
5.2.III	6770	16:35	NTV	6/9/22	42	49.3	8.31E-03	0	0.49
5.2.IV	6770	16:35		3/9/22	38	47.8	1.69E-02	0	0.65
5.2.V	9631	16:15	Trenitalia	6/9/22	42	45.9	1.45E-02	0	0.69
5.2.VI	9631	16:15		3/9/22	38	57.6	7.05E-03	0	0.16
5.2.VII	9981	15:25	NTV	6/9/22	42	47.0	9.72E-03	0	0.55
5.2.VIII	9981	15:25		3/9/22	38	45.9	1.56E-02	0	0.53

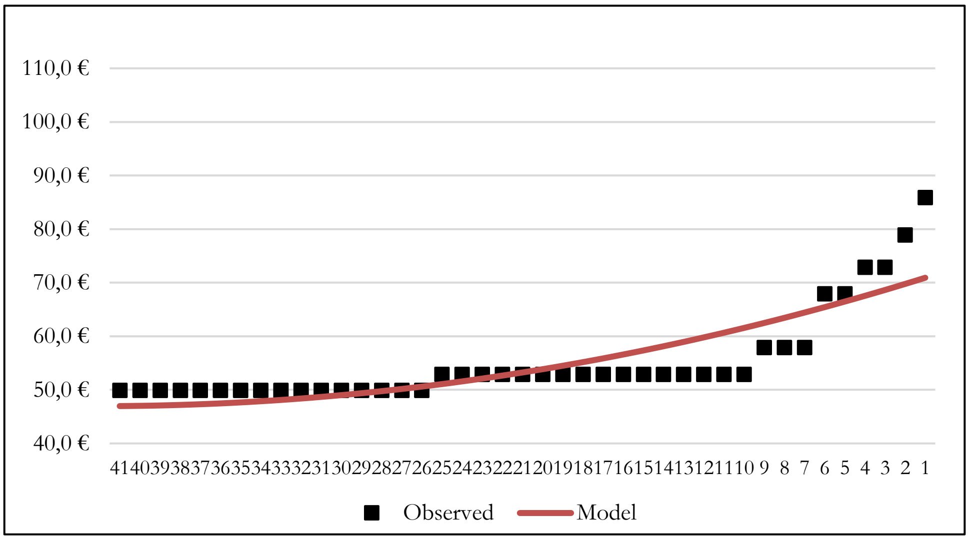


Fig. 5.13. Calibrated parameters of model 6

Tab 5.16 reports the calibrated parameters of model 7. The values of the calibrated parameter α_0 lie between 46.8 € and 57.6 €, the value of α_2 lie between 2.19E-04 and 5.95E-04, the value of α_3 lie between 9.99E-05 and 0. The value of the coefficient of determination (ρ^2) oscillates from 0.24 to 0.82. Fig. 5.14 presents the plotted values of the observed fares and of the fare estimated by model 7.

Tab.5.16. Calibrated parameters of model 7

Fig	run	Arrival time	Company	t	|K|	α_0	α_2	α_3	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	48.2	3.60E-04	1.00E-03	0.82
5.2.II	9634	15:58		3/9/22	38	54.5	5.95E-04	0.00E+00	0.76
5.2.III	6770	16:35	NTV	6/9/22	42	49.9	2.33E-04	9.82E-05	0.61
5.2.IV	6770	16:35		3/9/22	38	48.9	5.03E-04	9.99E-05	0.78
5.2.V	9631	16:15	Trenitalia	6/9/22	42	47.1	3.72E-04	1.01E-03	0.81
5.2.VI	9631	16:15		3/9/22	38	57.4	2.19E-04	1.00E-03	0.24
5.2.VII	9981	15:25	NTV	6/9/22	42	47.7	2.70E-04	9.82E-05	0.68
5.2.VIII	9981	15:25		3/9/22	38	46.8	4.45E-04	1.00E-03	0.65

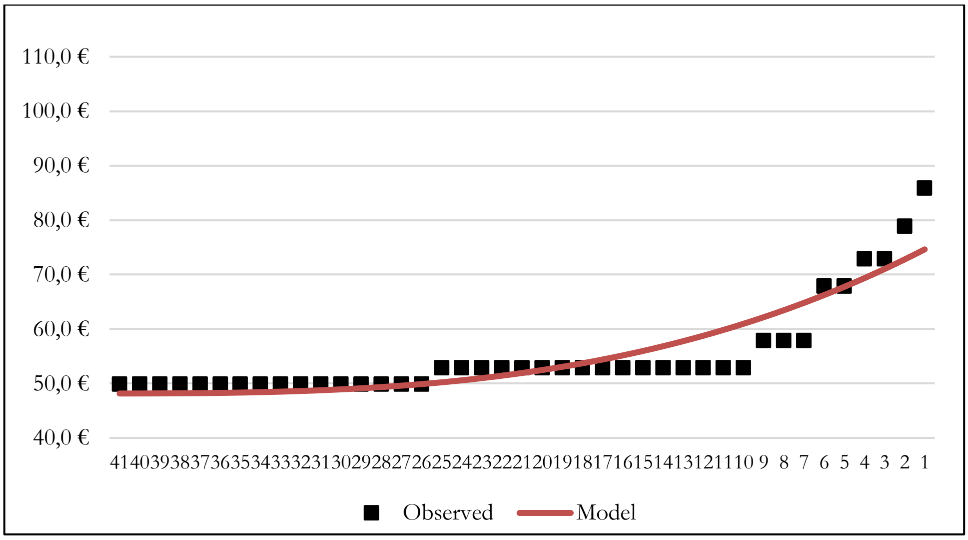


Fig. 5.14. Calibrated parameters of model 7

Tab 5.17 reports the calibrated parameters of model 8. The value of the calibrated parameter α_0 lie between 48.2 € and 57.8 €, the value of α_3 lie between 2.63E-06 and 1.75E-05, the value of α_4 lie between 3.90E-09 and 2.99E-04. The value of the coefficient of determination (ρ^2) oscillates from 0.35 to 0.88. Fig. 5.15 presents the plotted values of the observed fares and of the fare estimated by model 8.

Tab.5.17. Calibrated parameters of model 8

Fig	run	Arrival time	Company	t	|K|	α_0	α_3	α_4	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	48.3	2.63E-06	2.99E-04	0.85
5.2.II	9634	15:58		3/9/22	38	54.9	1.75E-05	6.41E-07	0.80
5.2.III	6770	16:35	NTV	6/9/22	42	51.3	5.88E-06	1.92E-07	0.69
5.2.IV	6770	16:35		3/9/22	38	50.1	1.44E-05	5.17E-07	0.86
5.2.V	9631	16:15	Trenitalia	6/9/22	42	49.3	9.82E-06	3.27E-07	0.88
5.2.VI	9631	16:15		3/9/22	38	57.8	7.72E-06	2.62E-07	0.35
5.2.VII	9981	15:25	NTV	6/9/22	42	48.3	7.35E-06	1.93E-07	0.78
5.2.VIII	9981	15:25		3/9/22	38	48.2	1.34E-05	3.90E-09	0.75

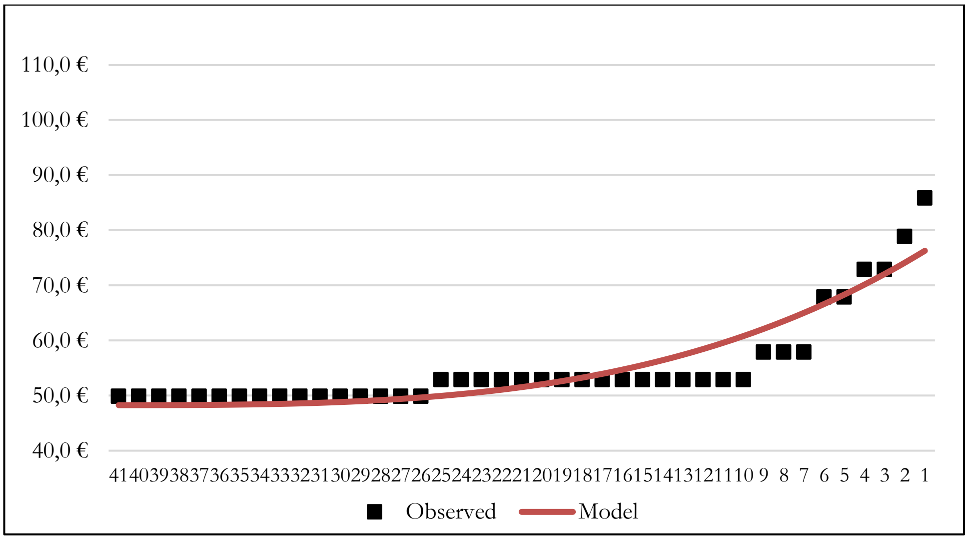


Fig. 5.15. Calibrated parameters of model 8

Tab 5.18 reports the calibrated parameters of model 9. The value of the calibrated parameter α_0 lie between 44.9 € and 57.6 €, the value of α_4 lie between 5.85E-09 and 4.69E-07, the value of α_5 lie between 5.43E-12 and 1.69E-05. The value of the coefficient of determination (ρ^2) oscillates from 0.35 to 0.91. Fig. 5.16 presents the plotted values of the observed fares and of the fare estimated by model 9.

Tab.5.18. Calibrated parameters of model 9

Fig	run	Arrival time	Company	t	|K|	α_0	α_4	α_5	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	48.1	2.91E-07	1.23E-09	0.91
5.2.II	9634	15:58		3/9/22	38	55.6	8.21E-09	1.69E-05	0.80
5.2.III	6770	16:35	NTV	6/9/22	42	50.4	1.77E-07	7.32E-10	0.78
5.2.IV	6770	16:35		3/9/22	38	49.7	1.04E-08	1.47E-05	0.87
5.2.V	9631	16:15	Trenitalia	6/9/22	42	47.9	5.85E-09	1.06E-05	0.89
5.2.VI	9631	16:15		3/9/22	38	57.6	1.68E-08	7.92E-06	0.35
5.2.VII	9981	15:25	NTV	6/9/22	42	46.9	2.22E-07	3.65E-10	0.81
5.2.VIII	9981	15:25		3/9/22	38	44.9	4.69E-07	5.43E-12	0.75

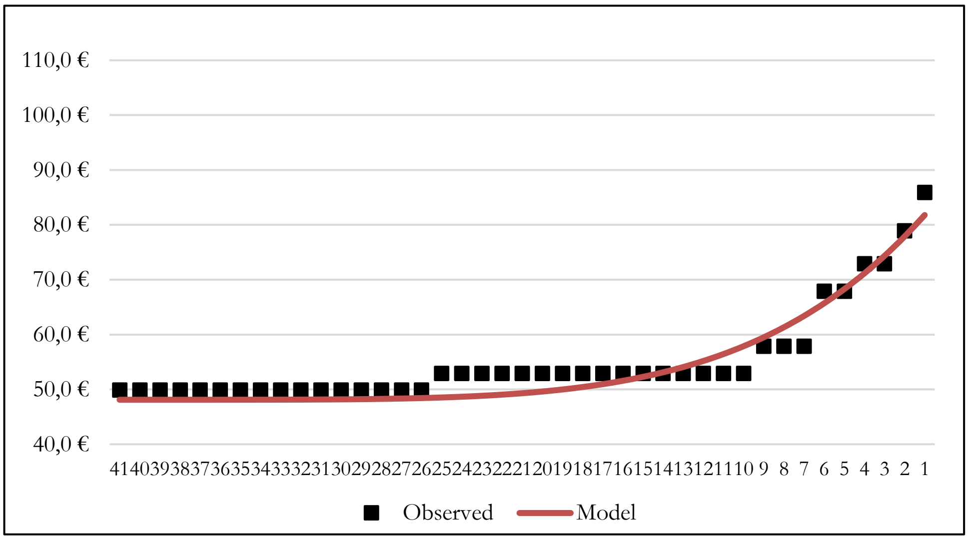


Fig. 5.16. Calibrated parameters of model 9

c. Calibration phase 3

Phase 3 is the last phase for the single run case and concerns the calibration of models 10-11, which have been specified in Tab. 5.7.

Tab 5.19 reports the calibrated parameters of model 10. The value of the calibrated parameter α_0 lies between 48.5 € and 57.7 €, the value of α' lies between 1.25E-08 and 3.17E-05 and the value of α_1 between 3.807 and 5.709. The value of the coefficient of determination (ρ^2) oscillates from 0.47 to 0.95. Fig. 5.17 presents the plotted values of the observed fares and of the fare estimated by model 10.

Tab.5.19. Calibrated parameters of model 10

Fig	run	Arrival time	Company	t	|K|	α_0	α'	α_1	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	49.9	3.26E-08	5.571	0.95
5.2.II	9634	15:58		3/9/22	38	56.1	3.17E-05	3.807	0.79
5.2.III	6770	16:35	NTV	6/9/22	42	50.6	1.25E-08	5.709	0.82
5.2.IV	6770	16:35		3/9/22	38	50.5	3.33E-08	5.694	0.94
5.2.V	9631	16:15	Trenitalia	6/9/22	42	48.8	4.01E-08	5.517	0.96
5.2.VI	9631	16:15		3/9/22	38	57.7	2.06E-08	5.698	0.47
5.2.VII	9981	15:25	NTV	6/9/22	42	48.6	3.59E-08	5.466	0.87
5.2.VIII	9981	15:25		3/9/22	38	48.5	9.42E-08	5.395	0.84

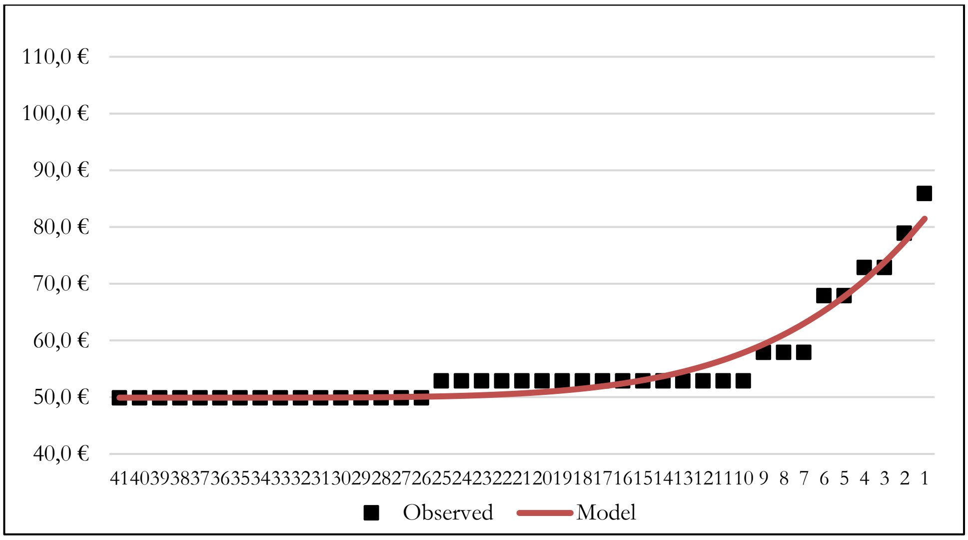


Fig. 5.17. Calibrated parameters of model 10

Tab 5.20a and Tab 5.20b report the five calibrated parameters of model 11. The value of the coefficient of determination (ρ^2) oscillates from 0.45 to 0.96. Fig. 5.18 presents the plotted values of the observed fares and of the fare estimated by model 11.

Tab.5.20a. Calibrated parameters of model 11 (part a)

Fig	run	Arrival time	Company	t	|K|	α_0	α'	α''
5.2.I	9634	15:58	Trenitalia	6/9/22	42	50.0	2.68E-08	9.25E-05
5.2.II	9634	15:58		3/9/22	38	56.9	1.89E-09	8.00E-05
5.2.III	6770	16:35	NTV	6/9/22	42	50.8	3.46E-08	6.08E-05
5.2.IV	6770	16:35		3/9/22	38	50.7	7.70E-09	2.99E-05
5.2.V	9631	16:15	Trenitalia	6/9/22	42	49.2	3.71E-08	3.00E-05
5.2.VI	9631	16:15		3/9/22	38	57.9	5.41E-08	7.99E-05
5.2.VII	9981	15:25	NTV	6/9/22	42	48.5	5.01E-08	8.00E-06
5.2.VIII	9981	15:25		3/9/22	38	48.6	4.33E-08	8.00E-06

Tab.5.20b. Calibrated parameters of model 11 (part b)

Fig	run	Arrival time	Company	t	|K|	α_1	α_2	ρ^2
5.2.I	9634	15:58	Trenitalia	6/9/22	42	5.615	0.932	0.94
5.2.II	9634	15:58		3/9/22	38	6.535	0.996	0.87
5.2.III	6770	16:35	NTV	6/9/22	42	5.431	1.041	0.81
5.2.IV	6770	16:35		3/9/22	38	6.108	0.891	0.96
5.2.V	9631	16:15	Trenitalia	6/9/22	42	5.531	0.895	0.95
5.2.VI	9631	16:15		3/9/22	38	5.418	0.997	0.45
5.2.VII	9981	15:25	NTV	6/9/22	42	5.371	0.979	0.86
5.2.VIII	9981	15:25		3/9/22	38	5.596	0.949	0.85

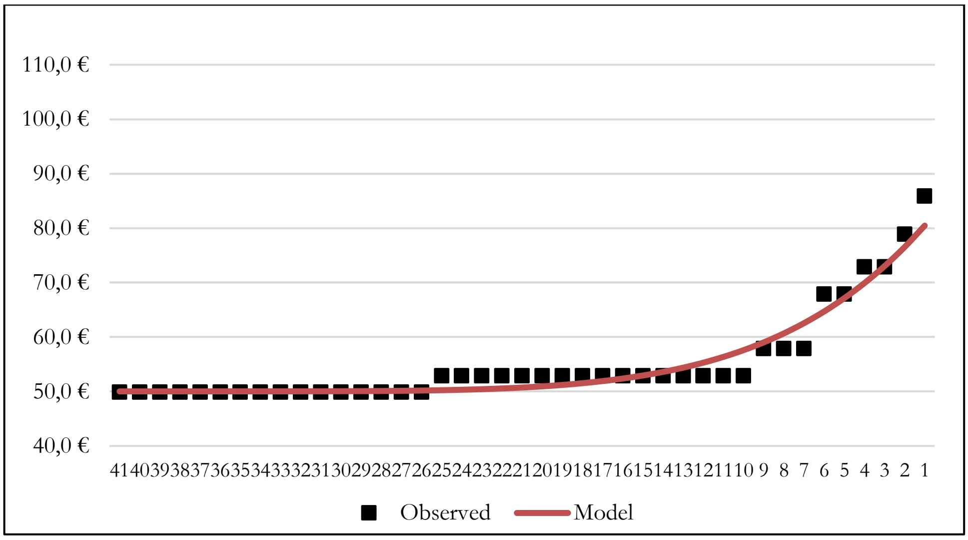


Fig. 5.18. Calibrated parameters of model 11

5.6.3. Calibrated parameters: set of runs

The paragraph contains the calibration of the models from number 12 to number 20 (Tab. 5.8) and clustered in 2 phases of calibration. The latter models are specified in the above paragraph and referred to the case of set of runs analysis.

a. Calibration phase 4

Phase 4 concerns the calibration of models 12-15, that have been specified in the previous paragraph, as reported in Tab. 5.8.

Tab 5.21 reports the calibrated parameters of model 12. The value of the calibrated parameter α_0 lie between 46.4 € and 52.9 and the value of α_2 lie between 8.38E-03 and 1.88E-02. The value of the coefficient of determination (ρ^2) oscillates from 0.42 to 0.72. Fig. 5.19 presents the plotted values of the observed fares are plotted and of the fare estimated by model 12.

Tab.5.21. Calibrated parameters of model 12

Fig	run	Arrival time	Company	t	|K|	α_0	α_2	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	46.4	1.44E-02	0.72
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	48.6	1.88E-02	0.64
5.2.XI	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}	NTV	6/9/22	42	49.3	8.38E-03	0.49
5.2.XII	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}		3/9/22	38	48.7	1.36E-02	0.58
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	47.1	1.37E-02	0.66
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	52.9	1.28E-02	0.42
5.2.XV	8111,9981, 9963	10:19, 15:25, 00:19^{+1}	NTV	6/9/22	42	48.6	8.84E-03	0.51
5.2.XVI	8111,9981, 9963	10:19, 15:25, 00:19^{+1}		3/9/22	38	48.2	1.20E-02	0.49

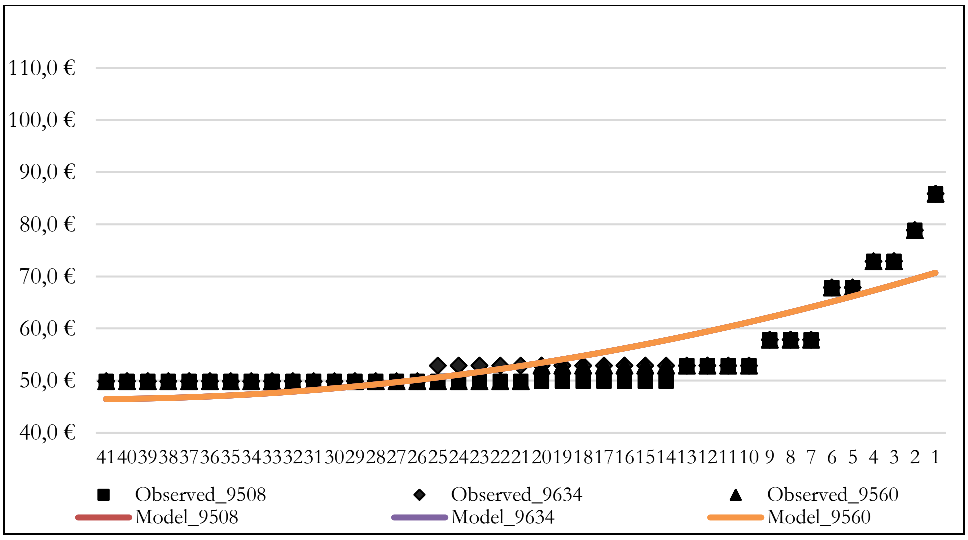


Fig. 5.19. Calibrated parameters of model 12

Tab 5.22 reports the calibrated parameters of model 13. The value of the calibrated parameter α_0 lie between 47.7 € and 53.8 and the value of α_3 lie between 2.37E-04 and 5.42E-04. The value of the coefficient of determination (ρ^2) oscillates from 0.51 to 0.83. Fig. 5.20 presents the plotted values of the observed fares and of the fare estimated by model 13.

Tab.5.22. Calibrated parameters of model 13

Fig	run	Arrival time	Company	t	|K|	α_0	α_3	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	47.7	3.92E-04	0.83
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	50.2	5.42E-04	0.72
5.2.XI	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}	NTV	6/9/22	42	49.9	2.37E-04	0.61
5.2.XII	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}		3/9/22	38	49.6	4.09E-04	0.70
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	48.2	3.78E-04	0.79
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	53.8	3.86E-04	0.51
5.2.XV	8111,9981, 9963	10:19, 15:25, 00:19^{+1}	NTV	6/9/22	42	49.2	2.49E-04	0.63
5.2.XVI	8111,9981, 9963	10:19, 15:25, 00:19^{+1}		3/9/22	38	48.9	3.65E-04	0.62

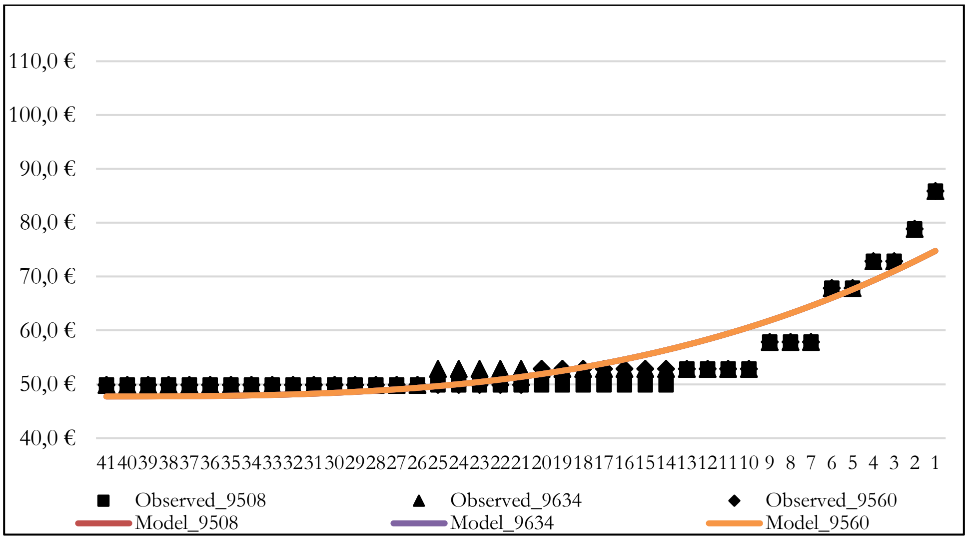


Fig. 5.20. Calibrated parameters of model 13

Tab 5.23 reports the calibrated parameters of model 14. The value of the calibrated parameter α_0 lie between 48.6 € and 54.4 and the value of α_4 lie between 6.52E-06 and 1.54E-05. The value of the coefficient of determination (ρ^2) oscillates from 0.58 to 0.89. Fig. 5.21 presents the plotted values of the observed fares and of the fare estimated by model 14.

Tab.5.23. Calibrated parameters of model 14

Fig	run	Arrival time	Company	t	|K|	α_0	α_4	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	48.6	1.04E-05	0.89
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	51.2	1.54E-05	0.76
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$	NTV	6/9/22	42	50.2	6.52E-06	0.71
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$		3/9/22	38	50.2	1.19E-05	0.79
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	48.9	1.02E-05	0.87
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	54.4	1.13E-05	0.58
5.2.XV	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$	NTV	6/9/22	42	49.6	6.83E-06	0.72
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$		3/9/22	38	49.4	1.08E-05	0.71

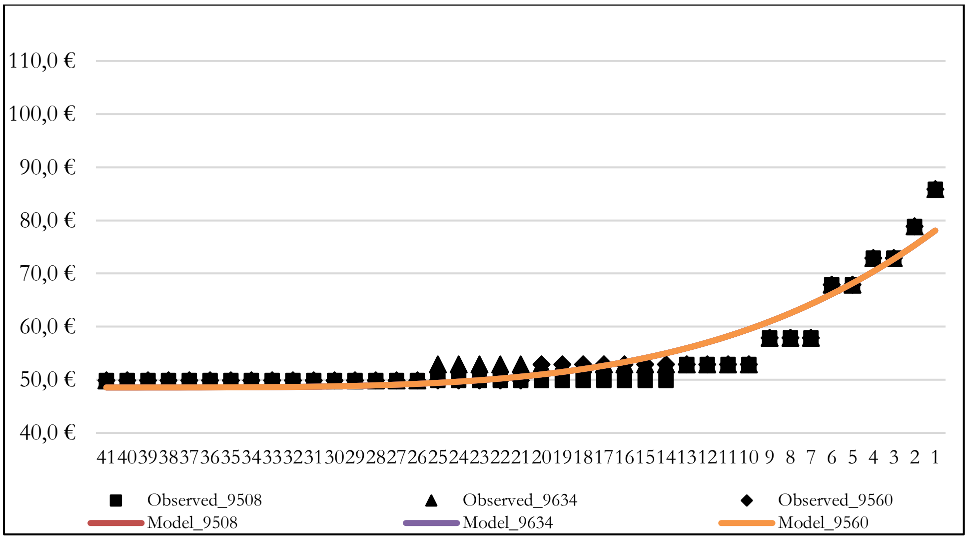


Fig. 5.21. Calibrated parameters of model 14

Tab 5.24 reports the calibrated parameters of model 15. The value of the calibrated parameter α_0 lie between 49.5 € and 54.9 and the value of α_5 lie between 1.67E-07 and 4.52E-07. The value of the coefficient of determination (ρ^2) oscillates from 0.62 to 0.93. Fig. 5.22 presents the plotted values of the observed fares and of the fare estimated by model 15.

Tab.5.24. Calibrated parameters of model 15

Fig	run	Arrival time	Company	t	|K|	α_0	α_5	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	49.9	2.61E-07	0.93
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	50.9	4.52E-07	0.79
5.2.XI	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}	NTV	6/9/22	42	50.9	1.69E-07	0.78
5.2.XII	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}		3/9/22	38	50.9	3.33E-07	0.84
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	49.5	2.67E-07	0.92
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	54.9	3.17E-07	0.62
5.2.XV	8111,9981, 9963	10:19, 15:25, 00:19^{+1}	NTV	6/9/22	42	50.9	1.67E-07	0.77
5.2.XVI	8111,9981, 9963	10:19, 15:25, 00:19^{+1}		3/9/22	38	50.9	2.84E-07	0.76

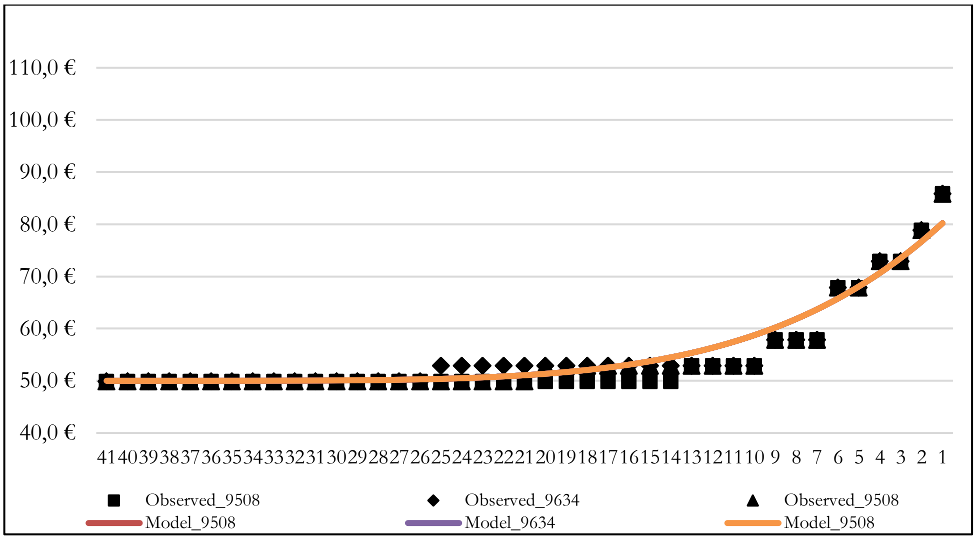


Fig. 5.22. Calibrated parameters of model 15

b. Calibration phase 5

Phase 5 concerns the calibration of models 16-20, that have been specified in the previous paragraph, as reported in Tab. 5.8.

Tab. 5.25a and Tab. 5.25b reports the calibrated parameters of model 16. The value of the calibrated parameter α_0 lie between 46.2 € and 52.9, the value of $\alpha_{2,r1}$ lie between 8.38E-03 and 1.59E-02, the value of $\alpha_{2,r2}$ lie between 8.17E-03 and 2.54E-02, the value of $\alpha_{2,r3}$ lie between 8.38E-03 and 1.49E-02. The value of the coefficient of determination (ρ^2) oscillates from 0.48 to 0.73. Fig. 5.25 presents the plotted values of the observed fares are plotted and of the fare estimated by model 16.

Tab.5.25a. Calibrated parameters of model 16 (part a)

Fig	run	Arrival time	Company	t	|K|	α_0	$\alpha_{2,r1}$	$\alpha_{2,r2}$
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	46.4	1.40E-02	1.47E-02
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	48.6	1.59E-02	2.54E-02
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$	NTV	6/9/22	42	49.3	8.38E-03	8.38E-03
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$		3/9/22	38	48.7	1.34E-02	1.58E-02
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	47.1	1.45E-02	1.34E-02
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	52.9	1.68E-02	1.22E-02
5.2.XV	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$	NTV	6/9/22	42	48.6	9.02E-03	8.17E-03
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$		3/9/22	38	48.2	1.14E-02	1.31E-02

Tab.5.25b. Calibrated parameters of model 16 (part b)

Fig	run	Arrival time	Company	t	|K|	$\alpha_{2,r3}$	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	1.45E-02	0.72
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	1.49E-02	0.73
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$	NTV	6/9/22	42	8.38E-03	0.49
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$		3/9/22	38	1.16E-02	0.60
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	1.34E-02	0.66
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	9.45E-03	0.48
5.2.XV	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$	NTV	6/9/22	42	9.33E-03	0.51
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$		3/9/22	38	1.14E-02	0.50

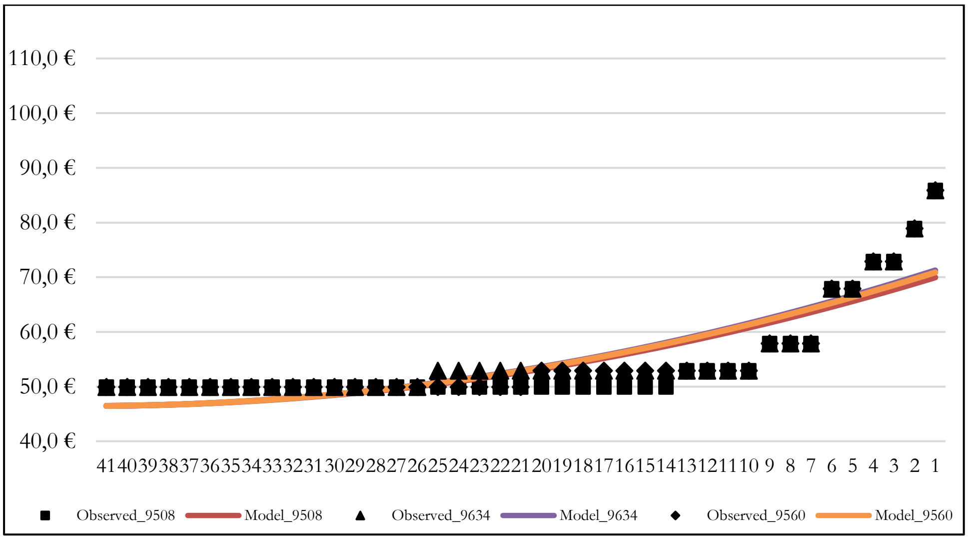


Fig. 5.23. Calibrated parameters of model 16

Tab. 5.26a and Tab. 5.26b reports the calibrated parameters of model 17. The value of the calibrated parameter α_0 lie between 47.7 € and 53.8, the value of $\alpha_{3,r1}$ lie between 2.37E-04 and 4.86E-04, the value of $\alpha_{3,r2}$ lie between 2.37E-04 and 7.28E-04, the value of $\alpha_{3,r3}$ lie between 2.37E-04 and 4.37E-04. The value of the coefficient of determination (ρ^2) oscillates from 0.55 to 0.83. Fig. 5.24 presents the plotted values of the observed fares and of the fare estimated by model 17.

Tab.5.26a. Calibrated parameters of model 17 (part a)

Fig	run	Arrival time	Company	t	|K|	α_0	$\alpha_{3,r1}$	$\alpha_{3,r2}$
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	47.7	3.84E-04	3.98E-04
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	50.2	4.60E-04	7.28E-04
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$	NTV	6/9/22	42	49.9	2.37E-04	2.37E-04
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$		3/9/22	38	49.6	3.98E-04	4.84E-04
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	48.2	3.93E-04	3.71E-04
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	53.8	4.86E-04	3.61E-04
5.2.XV	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$	NTV	6/9/22	42	49.2	2.52E-04	2.36E-04
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$		3/9/22	38	48.9	3.41E-04	4.12E-04

Tab.5.26b. Calibrated parameters of model 17 (part b)

Fig	run	Arrival time	Company	t	|K|	$\alpha_{3,r3}$	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	3.95E-04	0.83
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	4.37E-04	0.79
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$	NTV	6/9/22	42	2.37E-04	0.61
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00:20^{+1}$		3/9/22	38	3.43E-04	0.73
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	3.71E-04	0.79
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	3.11E-04	0.55
5.2.XV	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$	NTV	6/9/22	42	2.60E-04	0.63
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00:19^{+1}$		3/9/22	38	3.41E-04	0.63

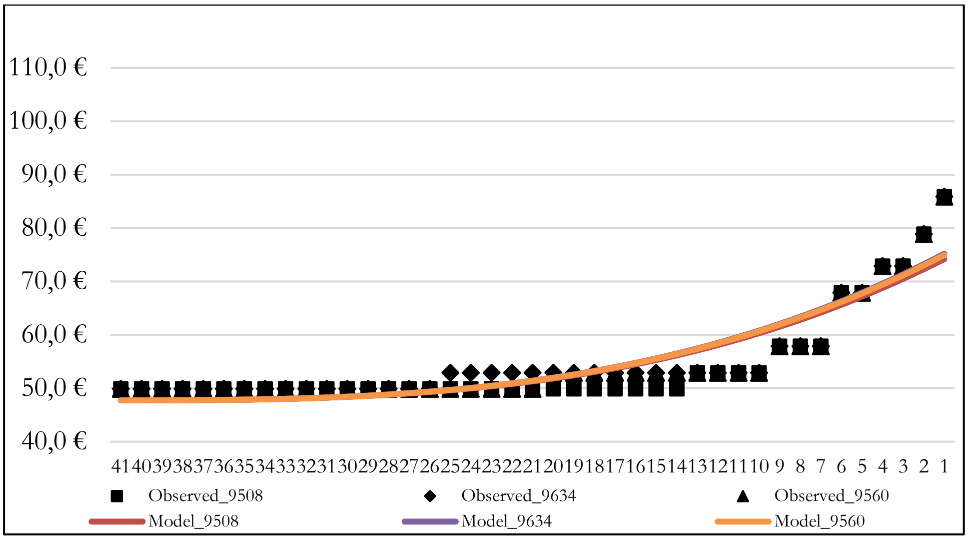


Fig. 5.24. Calibrated parameters of model 17

Tab. 5.27a and Tab. 5.27b reports the calibrated parameters of model 18. The value of the calibrated parameter α_0 lie between 48.6 € and 54.4, the value of $\alpha_{4,r1}$ lie between 6.52E-06 and 1.54E-05, the value of $\alpha_{4,r2}$ lie between 6.52E-06 and 0, the value of $\alpha_{4,r3}$ lie between 6.52E-06 and 0. The value of the coefficient of determination (ρ^2) oscillates from 0.59 to 0.89. Fig. 5.25 presents the plotted values of the observed fares and of the fare estimated by model 18.

Tab.5.27a. Calibrated parameters of model 18 (part a)

Fig	run	Arrival time	Company	t	|K|	α_0	$\alpha_{4,r1}$	$\alpha_{4,r2}$
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	48.6	1.03E-05	1.06E-05
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	51.2	1.54E-05	0.00E+00
5.2.XI	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}	NTV	6/9/22	42	50.2	6.52E-06	6.52E-06
5.2.XII	6636, 6770, 9962	10:10, 16:35, 00:20^{+1}		3/9/22	38	50.2	1.15E-05	1.43E-05
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	48.9	1.04E-05	1.00E-05
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	54.4	1.36E-05	1.06E-05
5.2.XV	8111,9981, 9963	10:19, 15:25, 00:19^{+1}	NTV	6/9/22	42	49.6	6.87E-06	6.57E-06
5.2.XVI	8111,9981, 9963	10:19, 15:25, 00:19^{+1}		3/9/22	38	49.4	9.95E-06	1.24E-05

Tab.5.27b. Calibrated parameters of model 18 (part b)

Fig	run	Arrival time	Company	t	|K|	$\alpha_{4,r3}$	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	1.05E-05	0.89
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	0.00E+00	0.76
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$	NTV	6/9/22	42	6.52E-06	0.71
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$		3/9/22	38	9.98E-06	0.82
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	1.00E-05	0.87
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	9.65E-06	0.59
5.2.XV	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$	NTV	6/9/22	42	7.06E-06	0.72
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$		3/9/22	38	9.95E-06	0.73

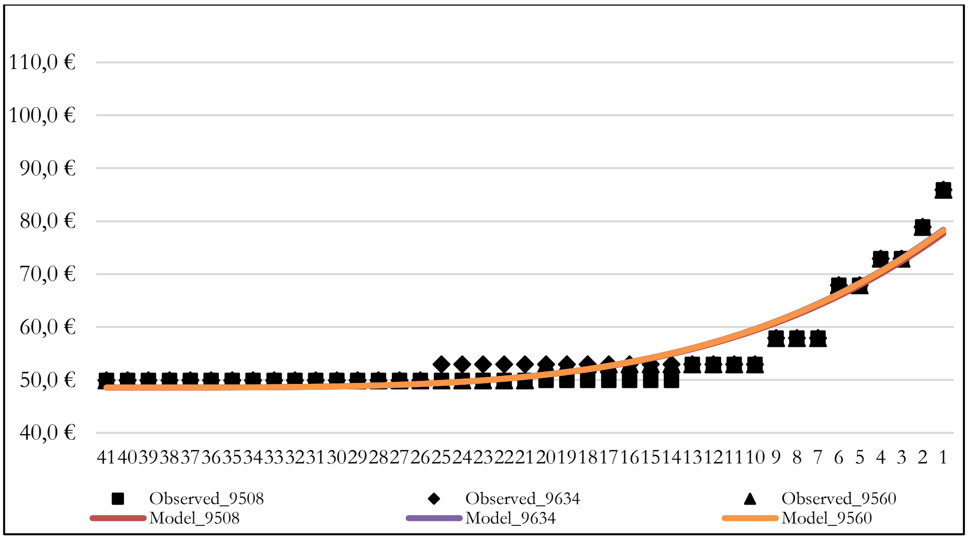


Fig. 5.25. Calibrated parameters of model 18

Tab. 5.28a and Tab. 5.28b reports the calibrated parameters of model 19. The value of the calibrated parameter α_0 lie between 48.9 € and 50.9, the value of $\alpha_{5,r1}$ lie between 1.75E-07 and 4.59E-07, the value of $\alpha_{5,r2}$ lie between 1.75E-07 and 6.92E-07, the value of $\alpha_{5,r3}$ lie between 1.75E-07 and 4.25E-07. The value of the coefficient of determination (ρ^2) oscillates from 0.50 to 0.94. Fig. 5.26 presents the plotted values of the observed fares and of the fare estimated by model 19.

Tab.5.28a. Calibrated parameters of model 19 (part a)

Fig	run	Arrival time	Company	t	|K|	α_0	$\alpha_{5,r1}$	$\alpha_{5,r2}$
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	48.9	2.75E-07	2.94E-07
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	48.9	3.43E-07	6.92E-07
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$	NTV	6/9/22	42	50.5	1.75E-07	1.75E-07
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$		3/9/22	38	50.9	3.20E-07	4.03E-07
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	50.9	2.52E-07	2.53E-07
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	50.9	4.59E-07	3.87E-07
5.2.XV	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$	NTV	6/9/22	42	49.9	1.83E-07	1.77E-07
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$		3/9/22	38	50.9	2.58E-07	3.37E-07

Tab.5.28b. Calibrated parameters of model 19 (part b)

Fig	run	Arrival time	Company	t	|K|	$\alpha_{5,r3}$	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	2.79E-07	0.94
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	4.25E-07	0.78
5.2.XI	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$	NTV	6/9/22	42	1.75E-07	0.79
5.2.XII	6636, 6770, 9962	10:10, 16:35, $00{:}20^{+1}$		3/9/22	38	2.78E-07	0.87
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	2.33E-07	0.89
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	3.69E-07	0.50
5.2.XV	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$	NTV	6/9/22	42	1.87E-07	0.79
5.2.XVI	8111,9981, 9963	10:19, 15:25, $00{:}19^{+1}$		3/9/22	38	2.58E-07	0.78

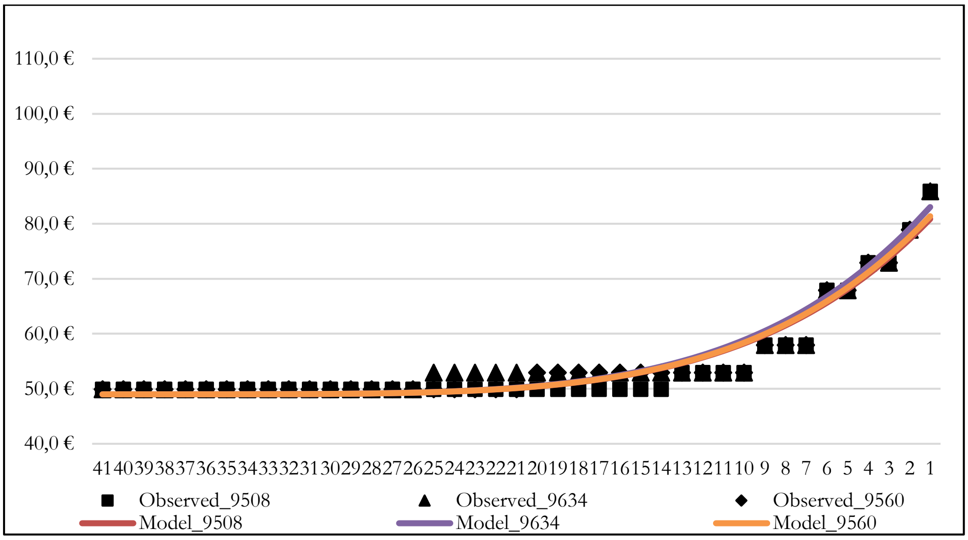


Fig. 5.26. Calibrated parameters of model 19

Tab. 5.29 reports the calibrated parameters of model 20. The value of the calibrated parameter α_0 lie between 49.7 € and 55.3, the value of α' lie between 9.79E-12 and 3.65E-08, the value of α lie between 5.69 and 7.69. The value of the coefficient of determination (ρ^2) oscillates from 0.65 to 0.96. Fig. 5.27 presents the plotted values of the observed fares are plotted and of the fare estimated by model 20.

Tab.5.29. Calibrated parameters of model 20

Fig	run	Arrival time	Company	t	|K|	α_0	α'	α	ρ^2
5.2.IX	9508, 9634, 9560	9:50, 15:58, 22:50	Trenitalia	6/9/22	42	49.7	1.26E-08	5.84	0.96
5.2.X	9508, 9634, 9560	9:50, 15:58, 22:50		3/9/22	38	52.0	3.65E-08	5.69	0.80
5.2.XI	6636, 6770, 9962	10:10, 16:35, 00:20+1	NTV	6/9/22	42	50.9	9.79E-12	7.69	0.91
5.2.XII	6636, 6770, 9962	10:10, 16:35, 00:20+1		3/9/22	38	51.4	9.49E-10	6.64	0.89
5.2.XIII	9515, 9631, 9961	10:49, 16:15, 22:10	Trenitalia	6/9/22	42	50.1	9.71E-10	6.54	0.95
5.2.XIV	9515, 9631, 9961	10:49, 16:15, 22:10		3/9/22	38	55.3	9.38E-10	6.63	0.65
5.2.XV	8111,9981, 9963	10:19, 15:25, 00:19+1	NTV	6/9/22	42	50.3	9.19E-10	6.45	0.86
5.2.XVI	8111,9981, 9963	10:19, 15:25, 00:19+1		3/9/22	38	50.2	9.43E-10	6.63	0.86

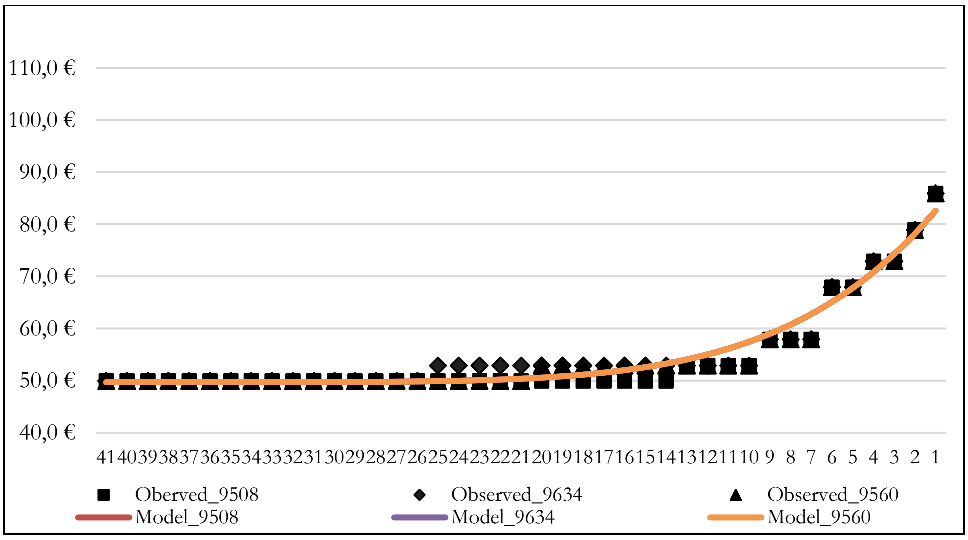


Fig. 5.27. Calibrated parameters of model 20

5.7. Discussion

The paragraph presents a discussion about the calibrated parameters of the fare models related to the single run and the models of set of runs.

Tab. 5.30 below shows the calibration results of the models specified in Tab. 5.7 and Tab. 5.8, in terms of maximum and minimum values of the coefficient of determination (ρ^2) with the corresponding figure.

Tab.5.30. Comparison of calibrated models: values of ρ^2

Model	R^2 max	Figure	ρ^2 min	Figure
		PHASE 1		
0	0.000	-	0.000	-
1	0.572	5.2.II	0.068	5.2.VI
2	0.727	5.2.I	0.155	5.2.VI
3	0.828	5.2.I	0.254	5.2.VI
4	0.896	5.2.V	0.347	5.2.VI
5	0.944	5.2.V	0.425	5.2.VI
		PHASE 2		
6	0.727	5.2.I	0.155	5.2.VI
7	0.823	5.2.I	0.243	5.2.VI
8	0.882	5.2.V	0.347	5.2.VI
9	0.908	5.2.I	0.350	5.2.VI
		PHASE 3		
10	0.956	5.2.V	0.470	5.2.VI
11	0.959	5.2.IV	0.452	5.2.VI
		PHASE 4		
12	0.715	5.2.IX	0.421	5.2.XIV
13	0.827	5.2.IX	0.514	5.2.XIV
14	0.897	5.2.IX	0.578	5.2.XIV
15	0.932	5.2.IX	0.619	5.2.XIV
		PHASE 5		
16	0.729	5.2.X	0.475	5.2.XIV
17	0.827	5.2.IX	0.547	5.2.XIV
18	0.897	5.2.IX	0.599	5.2.XIV
19	0.937	5.2.IX	0.500	5.2.XIV
20	0.959	5.2.IX	0.654	5.2.XIV

Tab. 5.31 contains the comparison of the calibrated parameters of the fare models. It has been calibrated the parameters of the models of the full sample of the year 2022. It is worth noting that the models that better replicate the observed fares are the specified model with fixed exponent 5. The Model

5, model 9 and model 15 result to have similar coefficient α_0 but different parameter α_5, with same mathematical order. The model 10 and 20 have analogous specification, the difference is that is calibrated the exponent, the latter result to be near the models with fixed exponent 5. Model 11 is specified with two exponents to be calibrated, and the results show that the latter two balance themselves towards the value of 5.

Tab.5.31. Comparison of calibrated parameters: fare modelling

Sample	Model	α_0	α_5	α'	α''	α_1	α_2
2022	5	48.6	2.79E-07	-	-	-	-
2022	9	48.1	2.91E-07	-	-	-	-
2022	10	48.8	-	4.01E-08	-	5.52	-
2022	11	57.9	-	7.71E-09	2.99E-05	6.11	0.89
2022	15	49.9	2.61E-07	-	-	-	-
2022	20	49.7	-	1.26E-08	-	5.84	-

5.A. Appendix

Appendix B contains some additional model specifications.

<u>Case 1.</u>

The general specification of the cylinder is given by eq. (B.1):

$$x^2+y^2=r^2 \tag{B.1}$$

The 2D function generated by the intersection of the three-dimensional geometrical surface with a plane (w-k) for $t=\bar{t}$ is a circle. An example is depicted in Fig. 5.28. This model allows to specify eq. (B.1) as an arc of a circle characterized by an initial value of fare. w_0. and a maximum finite value of fare. w_{MAX}. or final fare correspondent to k=0. The intersection with the plane $w=w(t)$. $k=\bar{k}$ (7) produces two horizontal lines. which reveals that the average value of fare is constant along t.

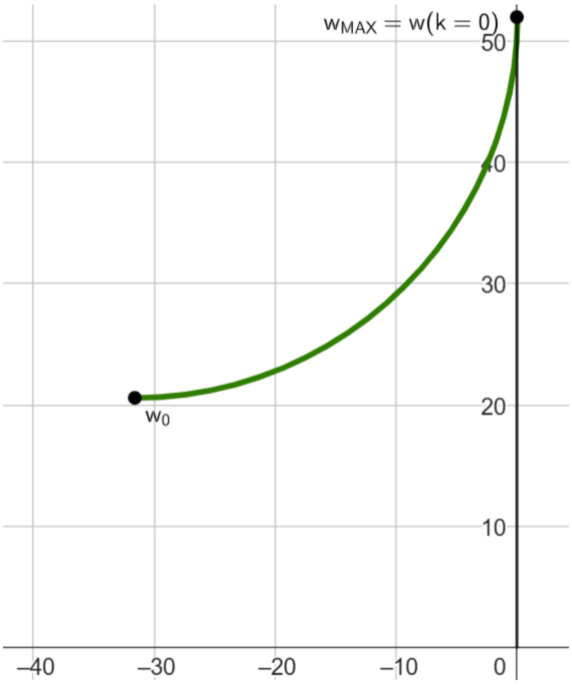


Fig. 5.28. Bidimensional model (w-k): arc of the circle (Case 1)

<u>Case 2.</u>

The specification is an elliptical cylinder (eq. B.2):

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0 \tag{B.2}$$

(with a. b > 0)

The 2D function generated by the intersection of the three-dimensional geometrical surface with the plane (w-k) for $t=\bar{t}$ is an ellipse. An example is represented in Fig. 5.29. This model allows to have the arc of an ellipse characterized by an initial value of fare, w_0, and a maximum finite value of fare. w_{MAX}. or final fare correspondent to k=0.

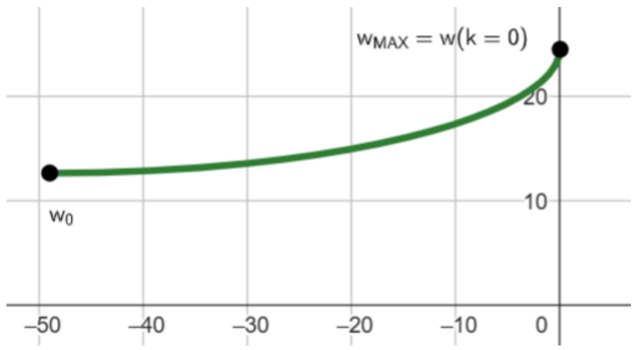


Fig. 5.29. Arc of the ellipse (Case 2)

<u>Case 3.</u>

The specification is an equilateral hyperbole (eq. B.3):

$$x\,y = k \tag{B.3}$$

(with: $k \neq 0$)

The 2D function provided by the intersection of the three-dimensional geometrical surface. Thus, the hyperboloid. with the plane (w-k) for $t=\bar{t}$ is the hyperbole. An example is represented in Fig. 5.30. This model allows to obtain the hyperbole with perpendicular asymptotes coinciding with axes. by specifying the domain of the function characterized by an initial fare $w_0=-\infty$. and the maximum result to be $w=+\infty$ with an asymptote to k=0.

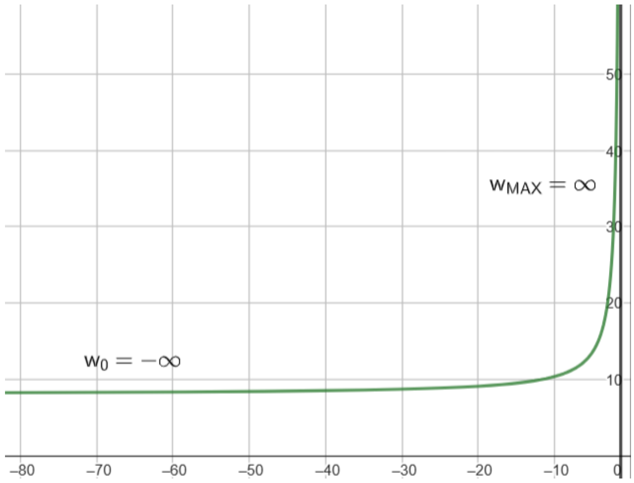


Fig. 5.30. Hyperbole (Case 3)

Case 4.

The specification is a parabolic cylinder (eq. B.4):

$$\frac{x^2}{a^2} - 2y = 0 \tag{B.4}$$

with: a > 0

The 2D function provided by the intersection of the three-dimensional (3D) geometrical surface with the plane (w-k) for $t=\bar{t}$ is the parable. An example is represented in Fig. 5.31. This model allows to obtain the arc of a parable characterized by an initial fare w_0. and the maximum value of fare w_{MAX} or final fare correspondent to k=0.

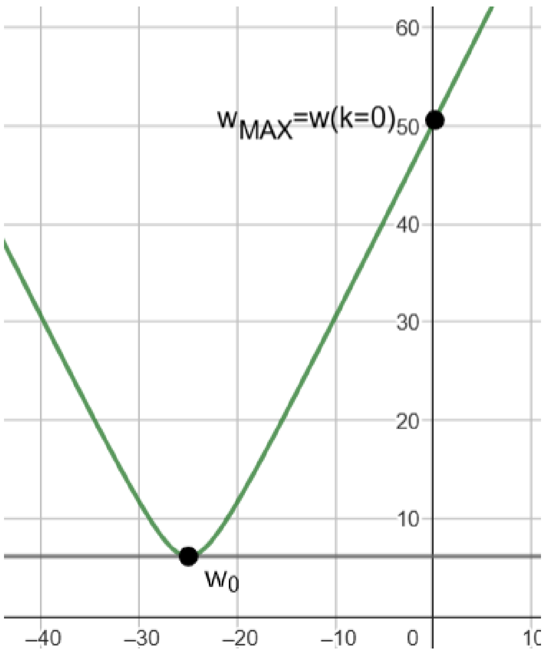


Fig. 5.31. Parable

6.
Conclusions and research perspectives

This chapter contains the conclusions of the thesis highlighting the main achievements of the proposed research and providing indications about future perspectives.

High Speed Rail (HSR) results to be the transport mode-service that allows to achieve sustainable mobility goals due to the environmental, economic and social benefits. Among the seventeen SDG Goals defined by UN, HSR generates direct impacts on Goals 9 and 11 and indirect impacts on Goals 7, 8, 10 and 13.

On the other hand, HSR has experienced rapid growth on both the supply side, in terms of services and infrastructures, and demand side, in terms of transport flows of passengers transported. On the supply side, there are more than 110,000 kilometers of HSR lines planned, in construction and in operation, with a primacy of China with about 45,000 kilometers in operation. On the demand side, there are almost three billion and half of passengers that have travelled worldwide in 2022 with a global value of annual average growth of 2.79.

The growth of HSR infrastructures and services caused competition at inter-modal level (e.g., mode choice) and at intra-modal level (e.g., service, company and run choices), hence on different travelers' choice levels in the context of intercity trips. Consequently, it is important to investigate the phenomenon of travelers' mobility to better understand the users' behavior in relation to the supply of HSR services.

As far as concerns the intra-modal competition, two relevant lines of research are present in the current scientific literature: the run choice modelling and the modelling of the dynamic structure of fares. Both run choice and fares models are commonly calibrated by means of data collected by traditional surveys of passenger mobility, operated via direct interviews to a sample of travelers in case of run choice models; and via the direct acquisition of information about a sample of fares supplied by companies in case of fares models.

The traditional surveys used to obtain data for models' calibration have been recently accompanied by new forms of surveys that use emerging Information and Communication Technologies (e-ICT). The existing e-ICT for mobility applications may be classified into five categories: Internet of things; Big Data; Artificial intelligence; Blockchain; Digital twin.

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

According to the analysis of the literature of the two lines of research described and the two RQs identified in the introduction, the thesis contribution concerns the combination of the two above mentioned lines of research by means of an innovative method for data collection. The method proposed is based on the acquisition of big data containing a great volume and variety of historical information about the characteristics of the transport supply (fares).

The research contribution concerns an original method for the identification of users' behavior in the choice dimension of run from the observation of the day-to-day evolution of ticket costs. The method is composed of two main phases. The former deals with the identification of user's choices analyzing the evolution of the tickets costs. The latter phase deals 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 discrete run choice model. The aim of the validation has been to verify the possibility to ground the proposed method into a specified and calibrated choice model.

The thesis presents an extended investigation of the dynamic structure of HSR fares and a calibration exercise of the fare function: w=w(k).

A calibration exercise of eq. w = *w* (k) was conducted on a portion of available data observed. The available data-base built during the surveys of 2022, 2023 and 2025 will allow to calibrate the more general equation: w = *w* (t, k, g).

The observation of fare structure along the plane w-k led to the identification of two main regions:

- a region located more distant from the day of trip and characterized by a sub-horizontal evolution of fares; in other words, where the prices are almost stable;
- a region located closer to the day of trip and characterized by an increasing positive gradient due to the variation in ticket fares.

It has been possible, in general, to identify a critical value of k, corresponding to one or a group of days of ticket purchase, which separate the two above described regions.

The previous considerations allow to define formally two branches of the fare diagram:

- a stable branch is the portion of the fare diagram such that the variations Δw do not repeat between two consecutive values of k;
- an unstable branch is the portion of the fare diagram such that the variations Δw are repeated at least between two consecutive k;
- the critical value of k is the threshold such that a relevant change in the evolution of the fare from constant to variable is observed.

The observation of fare structure along the plane w-t led to the identification of two main elements. The observed curves show a trend with higher values of fares in proximity to the weekend days in some cases and show higher values of fares as the day of trip approaching. These elements need to be further investigated during the fare modelling step.

The results of the research present some advantages and limitations which are reported below.

On the side of advantages, the framework addresses a significant and complex issue in transportation science: the estimation of traveler's choice models utilizing passively collected, large-scale datasets, thereby reducing or eliminating the necessity for costly and time-intensive surveys. The development of the framework represents a commendable effort for the transition from classical engineering and survey-based methodologies towards data-driven approaches for the comprehension and simulation of mobility patterns and of the supply of fares. Moreover, the study is based upon assumptions never considered before in existing literature. The assumptions of the study assume that a change in the fare of a specific run and fare class between two consecutive days is interpreted as a discrete choice event attributable to at least one user. This assumption is the cornerstone that allows to convert a signal that reveals a change in the fare level into a signal that reveals travel behavior (run choice).

On the side of limitations, the above assumption represents a simplification of the mechanism underlying transportation fare dynamics. Two aspects are of concern: revenue management systems, which play a role in the pricing strategies of High Speed Rail, and airlines. These systems do not necessarily adjust fares in a straightforward manner based on individual transactions: a single ticket purchase may not be sufficient to deplete a fare bucket and prompt the display of a higher fare. Conversely, a company may alter fare availability on a run in response to a competitor's promotional sale, even in the absence of any immediate purchases. The strength of the signal is critical consideration and needs to be further investigated.

By presuming a single user per mutation, the method generates a dataset of choices, and more evidence is necessary to demonstrate that the observed changes of fares are predominantly influenced by the immediate demand. The estimation of intensity of the "signal-to-noise ratio" will require a thorough knowledge of the company's strategies. These strategies, due to highly competitive context in which the HSR companies operate, only in few cases are public, whereas in many other cases are latent. In the case of Italian HSR companies, exogenous factors (not demand-driven factors) are mainly connected to the runs, to the hour of the day and the day of the week, to the availability of the fare in the last days before the trip day.

It is important to recognize that data-driven approaches have some limitations compared to traditional survey-based methods. In particular, such collected datasets generally do not contain detailed information on users' socioeconomic attributes or the motivations behind travel choices. As a result, some behavioral dimensions remain latent and can only be inferred indirectly from observed variables such as the chosen fare or the selected run.

The research perspectives may concern the development of several points described in the following.

As far as concerns the run choice modelling, it will be possible to explore different opportunities in order to develop this side of research. For instance, it is possible to extend the sample of users (e.g., up to millions of users), in order to obtain more statistically grounded calibrated parameters. Wang et al. (2024), in this context, present an empirical benchmark work comparing Machine Learning (ML) models and discrete choice models for travel demand modeling. The results show that the models could provide good predictive performance when large datasets are available, whereas discrete choice models remain essential for behavioral interpretation and market analysis. In addition, further modelling structures may be tested with different configurations of the models' specifications both on the side of run choice set and on the side of run choice model. Another advancement may be the extension of the method to other modes of transport and/or services (e.g., airline services), as well as the application of the proposed framework to HSR services operating in other regional and national contexts.

Finally, in the context of the research activities carried out at the National Center for Sustainable Mobility (MOST), future developments could focus on increasing the Technology Readiness Level (TRL), in reference to the framework proposed in the thesis. This could be achieved by applying the methodology to larger databases, additional rail relationships and different transport markets, as well as other modes of transport (e.g., air), as well as through collaborations with rail operators. From this perspective, run choice and structure models could form the basis for the development of decision support systems (DSS) intended for operators of HSR or air transport services, with possible applications in operational service planning, demand analysis and the definition of pricing strategies. Then, the application of the framework developed in the thesis to operational case studies, in collaboration with railway operators and other stakeholders in the mobility sector. Such developments would foster the shift from research results to operational applications, helping to increase the level of technological maturity of the developed methodologies and strengthen the potential impact of research on decision-making processes in the transport sector.

List of Tables

List of Figures

7. References

7.1. References cited in the thesis

Anda, C., Erath, A., & Fourie, P. J. (2017). Transport modelling in the age of big data. International Journal of Urban Sciences, 21(sup1), 19–42. https://doi.org/10.1080/12265934.2017.1281150

Anderson, C. K., & Wilson, J. G. (2003). Wait or buy? The strategic consumer: Pricing and profit implications. Journal of the Operational Research Society, 54(3), 299–306. https://doi.org/10.1057/palgrave.jors.2601497

Astarita, V., Guido, G., Haghshenas, S. S., Haghshenas, S. S., & Martino, G. (2025). Risk reduction in transportation systems: A Bibliometric-Based Literature Review on Artificial Intelligence in Disaster Response. Transportation Research Procedia, 90, 599–606. https://doi.org/10.1016/j.trpro.2025.06.102

Atzori, L., Iera, A., & Morabito, G. (2017). Understanding the Internet of Things: Definition, potentials, and societal role of a fast evolving paradigm. Ad Hoc Networks, 56, 122–140. https://doi.org/10.1016/j.adhoc.2016.12.004

Behrens, C., & Pels, E. (2012). Intermodal competition in the London–Paris passenger market: High-Speed Rail and air transport. Journal of Urban Economics, 71(3), 278–288. https://doi.org/10.1016/j.jue.2011.12.005

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete choice analysis: Theory and application to travel demand. MIT Press.

Ben-Akiva, M., Cascetta, E., Coppola, P., Papola, P., & Velardi, V. (2010). High speed rail demand forecasting in a competitive market: The Italian case study. 1–9.

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

Ben-Akiva, M., Mcfadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., Daly, A., De Palma, A., Gopinath, D., Karlstrom, A., & Munizaga, M. A. (2002). Hybrid Choice Models: Progress and Challenges. Marketing Letters, 13(3), 163–175. https://doi.org/10.1023/A:1020254301302.

Beria, P., Grimaldi, R., Albalate, D., & Bel, G. (2018). Delusions of success: Costs and demand of high-speed rail in Italy and Spain. Transport Policy, 68, 63–79. https://doi.org/10.1016/j.tranpol.2018.03.011

Borsati, M., & Albalate, D. (2020). On the modal shift from motorway to high-speed rail: Evidence from Italy. Transportation Research Part A: Policy and Practice, 137, 145–164. https://doi.org/10.1016/j.tra.2020.04.006

Campisi, T., Russo, A., Bouhouras, E., Tesoriere, G., & Basbas, S. (2023). The Increase in E-commerce Purchases and the Impact on the Newest European City Logistics Development. The Open Transportation Journal, 17(1), e26671212279047. https://doi.org/10.2174/0126671212279047231128105715

Campos, J., & De Rus, G. (2009). Some stylized facts about high-speed rail: A review of HSR experiences around the world. Transport Policy, 16(1), 19–28. https://doi.org/10.1016/j.tranpol.2009.02.008

Carlan, V., Sys, C., & Vanelslander, T. (2016). How port community systems can contribute to port competitiveness: Developing a cost–benefit framework. Research in Transportation Business & Management, 19, 51–64. https://doi.org/10.1016/j.rtbm.2016.03.009

Carlos Martín, J., Román, C., & Espino, R. (2008). Willingness to Pay for Airline Service Quality. Transport Reviews, 28(2), 199–217. https://doi.org/10.1080/01441640701577007

Cascetta, E. (2009). Transportation Systems Analysis: Models and Applications. Springer US. https://doi.org/10.1007/978-0-387-75857-2

Cascetta, E. (2013). Transportation systems engineering: Theory and methods (Vol. 49). Springer Science & Business Media.

Cascetta, E., & Coppola, P. (2012). An elastic demand schedule-based multimodal assignment model for the simulation of high speed rail (HSR) systems. EURO Journal on Transportation and Logistics, 1(1–2), 3–27. https://doi.org/10.1007/s13676-012-0002-0

Cascetta, E., & Coppola, P. (2014). High Speed Rail (HSR) Induced Demand

Models. Procedia - Social and Behavioral Sciences, 111, 147–156. https://doi.org/10.1016/j.sbspro.2014.01.047

Cascetta, E., & Coppola, P. (2017). Evidence from the Italian High-Speed Rail Market: Competition between Modes and Between HSR Operators. In High-Speed Rail and Sustainability: Decision-making and the Political Economy of Investment.

Cascetta, E., & Russo, F. (1997). Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance. Transportation, 24(3), 271–293. https://doi.org/10.1023/A:1004968411792

Cascetta, E., Cartenì, A., Henke, I., Pagliara, F. (2020). Economic growth, transport accessibility and regional equity impacts of high-speed railways in Italy: Ten years ex post evaluation and future perspectives. Transportation Research Part A: Policy and Practice, 139, 412–428. DOI: 10.1016/j.tra.2020.07.008

Cascetta, E., Nuzzolo, A., Biggiero, L., & Russo, F. (1995). Passenger and freight demand models for the Italian transportation system. Proceedings of the 7th World Conference on Transport Research, Volume 2.

Cascetta, E., Papola, A., Pagliara, F., & Marzano, V. (2011). Analysis of mobility impacts of the high speed Rome–Naples rail link using withinday dynamic mode service choice models. Journal of Transport Geography, 19(4), 635–643. https://doi.org/10.1016/j.jtrangeo.2010.07.001

Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002a). A model of route perception in urban road networks. Transportation Research Part B: Methodological, 36(7), 577–592. https://doi.org/10.1016/S0191-2615(00)00041-2

Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002b). A model of route perception in urban road networks. Transportation Research Part B: Methodological, 36(7), 577–592. https://doi.org/10.1016/S0191-2615(00)00041-2

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation Research Part C: Emerging Technologies, 68, 285–299. https://doi.org/10.1016/j.trc.2016.04.005

Cheng, Y.-H. (2010). High-speed rail in Taiwan: New experience and issues for future development. Transport Policy, 17(2), 51–63. https://doi.org/10.1016/j.tranpol.2009.10.009

Cirianni, F. M. M., Comi, A., & Quattrone, A. (2023). Mobility Control Centre and Artificial Intelligence for Sustainable Urban Districts. Information, 14(10), 581. https://doi.org/10.3390/info14100581

Clever, R., & Hansen, M. M. (2008). Interaction of Air and High-Speed Rail in Japan. Transportation Research Record: Journal of the Transportation Research Board, 2043(1), 1–12. https://doi.org/10.3141/2043-01

D'Alfonso, T., Jiang, C., & Bracaglia, V. (2016a). Air transport and high-speed rail competition: Environmental implications and mitigation strategies. Transportation Research Part A: Policy and Practice, 92, 261–276. https://doi.org/10.1016/j.tra.2016.06.009

D'Alfonso, T., Jiang, C., & Bracaglia, V. (2016b). Air transport and high-speed rail competition: Environmental implications and mitigation strategies. Transportation Research Part A: Policy and Practice, 92, 261–276. https://doi.org/10.1016/j.tra.2016.06.009

Di Gangi, M., & Russo, F. (2022). Potentiality of rail networks: Integrated services on conventional and high-speed lines. Computers in Railways XVIII, 101–112.

Di Gangi, M., & Vitetta, A. (2021). Quantum utility and random utility model for path choice modelling: Specification and aggregate calibration from traffic counts. Journal of Choice Modelling, 40, 100290. https://doi.org/10.1016/j.jocm.2021.100290

Dobruszkes, F. (2011). High-speed rail and air transport competition in Western Europe: A supply-oriented perspective. Transport Policy, S0967070X11000837. https://doi.org/10.1016/j.tranpol.2011.06.002

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

EIM. (2008). European Railway Technical Strategy—Technical Vision to guide the development of TSIs. https://transport.ec.europa.eu/document/download/60e067f5-1e7d-43a1-bce1-e7d6964dd84b_en?filename=20090305_eim.pdf&prefLang=es

EU and Council of the European Union. (2024). Regulation (EU) 2024/1679 on Union guidelines for the development of the trans-European transport network, amending Regulations (EU) 2021/1153 and (EU) No 913/2010 and repealing Regulation (EU) No 1315/2013. https://eur-lex.europa.eu/eli/reg/2024/1679/oj/eng

EU, (1996). Directive (UE) 96/48/EC of 23 july 1996 on the interoperability of the trans-European high-speed rail system. L235/6. 1-3. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32002D0735

EU. (2016). Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016L0797

EU. (2025). Connecting Europe through High-Speed Rail (COM(2025) 903 final). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52025DC0903

Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital Twin: Enabling Technologies, Challenges and Open Research. IEEE Access, 8, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358

Givoni, M. (2006). Development and Impact of the Modern High‐speed Train: A Review. Transport Reviews, 26(5), 593–611. https://doi.org/10.1080/01441640600589319

Givoni, M., & Dobruszkes, F. (2013). A Review of Ex-Post Evidence for Mode Substitution and Induced Demand Following the Introduction of High-Speed Rail. Transport Reviews, 33(6), 720–742. https://doi.org/10.1080/01441647.2013.853707

Goodman, L. A. (1961). Snowball Sampling. The Annals of Mathematical Statistics, 32(1), 148–170. https://doi.org/10.1214/aoms/1177705148.

Hensher, D. A. (1997). A practical approach to identifying the market potential for high speed rail: A case study in the Sydney-Canberra corridor. Transportation Research Part A: Policy and Practice, 31(6), 431–446. https://doi.org/10.1016/S0965-8564(97)00001-3

Hetrakul, P., & Cirillo, C. (2014). A latent class choice based model system for railway optimal pricing and seat allocation. Transportation Research Part E: Logistics and Transportation Review, 61, 68–83.

https://doi.org/10.1016/j.tre.2013.10.005

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002

Jung, S.-Y., & Yoo, K.-E. (2014). Passenger airline choice behavior for domestic short-haul travel in South Korea. Journal of Air Transport Management, 38, 43–47. https://doi.org/10.1016/j.jairtraman.2013.12.017

Khwanpruk, S., U-tapao, C., Khwanpruk, K., Laokhongthavorn, L., & Moryadee, S. (2021). A Study on High-Speed Rail Pricing Strategy for Thailand Based on Dynamic Optimal Pricing Model.

Koenigsberg, O., Muller, E., & Vilcassim, N. J. (2008). easyJet® pricing strategy: Should low-fare airlines offer last-minute deals? Quantitative Marketing and Economics, 6(3), 279–297. https://doi.org/10.1007/s11129-007-9036-2

Li, Z.-C., & Sheng, D. (2016). Forecasting passenger travel demand for air and high-speed rail integration service: A case study of Beijing-Guangzhou corridor, China. Transportation Research Part A: Policy and Practice, 94, 397–410. https://doi.org/10.1016/j.tra.2016.10.002

Malighetti, P., Paleari, S., & Redondi, R. (2009). Pricing strategies of low-cost airlines: The Ryanair case study. Journal of Air Transport Management, 15(4), 195–203. https://doi.org/10.1016/j.jairtraman.2008.09.017

Malighetti, P., Paleari, S., & Redondi, R. (2015). EasyJet pricing strategy: Determinants and developments. Transportmetrica A: Transport Science, 11(8), 686–701. https://doi.org/10.1080/23249935.2015.1063021

Manski, C. F. (1977). The structure of random utility models. Theory and Decision, 8(3), 229–254. https://doi.org/10.1007/BF00133443

Manski, C. F., & McFadden, D. (1981). Alternative estimators and sample designs for discrete choice analysis. Structural analysis of discrete data with econometric applications, 2–50.

McAfee, R. P., & te Velde, V. (2006). Dynamic Pricing in the Airline Industry. California Institute of Technology.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior.

McFadden, D. (1972). Conditional Logit Analysis of Qualitative Choice Behavior. UC Berkeley: Institute of Urban and Regional Development. https://escholarship.org/uc/item/61s3q2xr

McFadden, D. (1981). Econometric Models of Probabilistic Choice. Structural analysis of discrete data with econometric applications. MIT Press.

McFadden, D. (2001). Economic Choices. American Economic Review, 91(3), 351–378. https://doi.org/10.1257/aer.91.3.351

Meng, H., Yan, Z., Wang, Y., & Xu, Y. (2022). Optimizing Joint Decisions of Dynamic Pricing and Ticket Allocation for High-Speed Railway with Operators' Risk Preference. Journal of Advanced Transportation, 2022, 1–16. https://doi.org/10.1155/2022/7383396

Morlotti, C., Cattaneo, M., Malighetti, P., & Redondi, R. (2017). Multi-dimensional price elasticity for leisure and business destinations in the low-cost air transport market: Evidence from easyJet. Tourism Management, 61, 23–34. https://doi.org/10.1016/j.tourman.2017.01.009

Musolino, G. (2022). Sustainable Mobility as a Service: Demand Analysis and Case Studies. Information, 13(8), 376. https://doi.org/10.3390/info13080376

Musolino, G., Pedà, G., & Russo, F. (2022). Emerging ICT and port community systems: a survey of scientific literature. 123–135. Https://doi.org/10.2495/umt220111

Nurhidayat, A. Y., Widyastuti, H., & Utomo, D. P. (2018). Model of transportation mode choice between aircraft and high speed train of Jakarta-Surabaya route. IOP Conference Series: Earth and Environmental Science, 202, 012002. https://doi.org/10.1088/1755-1315/202/1/012002

Nuzzolo, A., & Russo, F. (1993). Un modello di rete diacronica per l'assegnazione dinamica al trasporto collettivo extraurbano. Ricerca operativa, (1993/67).

Nuzzolo, A., & Russo, F. (1997). Modelli e metodi per l'analisi dei sistemi di trasporto. FrancoAngeli Editore.

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., & Gangemi, F. (2000). A behavioural choice model for the evaluation of railway supply and pricing policies. Transportation Research Part A: Policy and Practice, 34(5), 395–404. https://doi.org/10.1016/S0965-8564(99)00041-5

Nuzzolo, A., Russo, F., & Crisalli, U. (2003). Transit network modelling: the schedule-based dynamic approach. Milano: Franco Angeli.

Ortúzar S., J. de D., & Willumsen, L. G. (2001). Modelling transport (3. ed., reprint. August 2001). Wiley.

Ottomanelli, M., Caggiani, L., & Sassanelli, D. (2007). Simultaneous path choice model calibration and O-D matrix estimation using traffic counts: A GLS estimator for congested network. Proceedings of the 11th World Conference on Transport Research. World Conference on Transport Research Society.

Oxford Learner's Dictionaries. (2025). Oxford Learner's Dictionaries. https://www.oxfordlearnersdictionaries.com/

Pels, E., & Rietveld, P. (2004). Airline pricing behaviour in the London–Paris market. Journal of Air Transport Management, 10(4), 277–281. https://doi.org/10.1016/j.jairtraman.2004.03.005

Quattrone, A., & Vitetta, A. (2011). Random and fuzzy utility models for road route choice. Transportation Research Part E: Logistics and Transportation Review, 47(6), 1126–1139. https://doi.org/10.1016/j.tre.2011.04.007

Ren, X., Chen, Z., Wang, F., Wang, J., Wang, C., Dan, T., & Du, Z. (2019). Impact of high-speed rail on intercity travel behavior change: The evidence from the Chengdu-Chongqing Passenger Dedicated Line. Journal of Transport and Land Use, 12(1). https://doi.org/10.5198/jtlu.2019.1302

Rindone, C. (2022). Sustainable Mobility as a Service: Supply Analysis and Test Cases. Information, 13(7), 351. https://doi.org/10.3390/info13070351

Román, C., & Martín, J. C. (2014). Integration of HSR and air transport: Understanding passengers' preferences. Transportation Research Part E: Logistics and Transportation Review, 71, 129–141. https://doi.org/10.1016/j.tre.2014.09.001

Román, C., Espino, R., & Martín, J. C. (2010). Analyzing Competition between the High Speed Train and Alternative Modes. The Case of the Madrid-Zaragoza-Barcelona Corridor. Journal of Choice Modelling, 3(1), 84–108.

https://doi.org/10.1016/S1755-5345(13)70030-7

Russo, F. (1997). Fuzzy theory in transportation field: Fuzzy sets for simulating path choice behaviour. Advances in intelligent systems, 279–283.

Russo, F. (2022). Sustainable Mobility as a Service: Dynamic Models for Agenda 2030 Policies. Information, 13(8), 355. https://doi.org/10.3390/info13080355

Russo, F., & Comi, A. (2022). Providing dynamic route advice for urban goods vehicles: The learning process enhanced by the emerging technologies. Transportation Research Procedia, 62, 632–639. https://doi.org/10.1016/j.trpro.2022.02.078

Russo, F., & Musolino, G. (2012). A unifying modelling framework to simulate the Spatial Economic Transport Interaction process at urban and national scales. Journal of Transport Geography, 24, 189–197. https://doi.org/10.1016/j.jtrangeo.2012.02.003

Russo, F., & Vitetta, A. (2003). An assignment model with modified Logit, which obviates enumeration and overlapping problems. Transportation, 30(2), 177–201. https://doi.org/10.1023/A:1022598404823

Russo, F., & Vitetta, A. (2011). Reverse assignment: Calibrating link cost functions and updating demand from traffic counts and time measurements. Inverse Problems in Science and Engineering, 19(7), 921–950. https://doi.org/10.1080/17415977.2011.565339

Russo, F., Comi, A., & Rindone, C. (2025). The Emerging ICT for Exposure Reduction in Urban Mobility: Transport Models for Risk Cycle. In A. Razminia & D. H. Nguyen (Eds), Intelligent Technology for Future Transportation (Vol. 2378, pp. 173–187). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-84148-4_14

Russo, F., Moschella, M., & Musolino, G. (2024). Railway Demand Evaluation: HSR Induced Component. In O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, & M. N. Faginas Lago (Eds), Computational Science and Its Applications – ICCSA 2024 Workshops (Vol. 14822, pp. 173–187). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65318-6_12

Russo, F., Musolino, G., & Sgro, D. (2025). Estimating run choice models with innovative data collection: Day-to-day tickets evolution in High Speed Rail. Transport Policy, 176, 103909. https://doi.org/10.1016/j.tranpol.2025.103909

Russo, F., Sgro, D., & Musolino, G. (2023). Sustainable Development of Railway Corridors: Methods and Models for High Speed Rail (HSR) Demand Analysis. In O. Gervasi, B. Murgante, A. M. A. C. Rocha, C. Garau, F. Scorza, Y. Karaca, & C. M. Torre (Eds), Computational Science and Its Applications – ICCSA 2023 Workshops (Vol. 14110, pp. 527–538). Springer Nature Switzerland. 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 O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, & M. N. Faginas Lago (Eds), Computational Science and Its Applications – ICCSA 2024 Workshops (Vol. 14822, pp. 188–201). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65318-6_13

Russo, F., Sgro, D., & Musolino, G. (2025). Modelling dynamic structures of fares in high-speed rail market. Transportation Research Procedia, 90, 734–741. https://doi.org/10.1016/j.trpro.2025.06.085

Selcuk, A. M., & Avsar, Z. M. (2019). Dynamic pricing in airline revenue management. Journal of Mathematical Analysis and Applications, 478(2), 1191–1217. https://doi.org/10.1016/j.jmaa.2019.06.012

The Britannica Dictionary. (2025). The Britannica Dictionary. https://www.britannica.com/dictionary

Train, K. E. (2009). Discrete choice methods with simulation (2nd ed). Cambridge university press.

UIC. (2018). High Speed Rail fast track to sustainable mobility. https://uic.org/passenger/highspeed/

UIC. (2024). RAILISA (RAIL Information System and Analyses). https://uic.org/support-activities/statistics/

UIC. (2025). High Speed Traffic in the World – Passenger-km by Country, Updated 1st February 2025. https://uic.org/IMG/pdf/20250618_high-speed_traffic_passkm.pdf

United Nations. (1987). Our Common Future. Report of the World Commission on Environmental and Development.

United Nations. (2015). Transforming our world: The 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda

United Nations. (2018). The sustainable development goals report 2018. https://unstats.un.org/sdgs/report/2018

Vitetta, A. (2016). A quantum utility model for route choice in transport systems. Travel Behaviour and Society, 3, 29–37. https://doi.org/10.1016/j.tbs.2015.07.003

Vitetta, A. (2022). Sustainable Mobility as a Service: Framework and Transport System Models. Information, 13(7), 346. https://doi.org/10.3390/info13070346

Wan, Y., Ha, H.-K., Yoshida, Y., & Zhang, A. (2016). Airlines' reaction to high-speed rail entries: Empirical study of the Northeast Asian market. Transportation Research Part A: Policy and Practice, 94, 532–557. https://doi.org/10.1016/j.tra.2016.10.014

Wang, H., Mo B., Zheng, Y., Hess, S. & Zhao J. (2024). Comparing hundreds of machine learning and discrete choice models for travel demand modeling: An empirical benchmark. Transportation Research Part B: Methodological, Volume 190, 2024. https://doi.org/10.1016/j.trb.2024.103061

Wang, X., Wang, H., & Zhang, X. (2016). Stochastic seat allocation models for passenger rail transportation under customer choice. Transportation Research Part E: Logistics and Transportation Review, 96, 95–112. https://doi.org/10.1016/j.tre.2016.10.003

Wu, D., & Martín, J. C. (2022). Research on Passengers' Preference for High-Speed Railways (HSRs) and High-Speed Trains (HSTs). Sustainability, 14(3), 1473. https://doi.org/10.3390/su14031473

Xiao, Y., Chen, J., & Liu, X. (2008). Joint Dynamic Pricing for Two Parallel Flights based on Passenger Choice Behavior. Systems Engineering - Theory & Practice, 28(1), 46–55. https://doi.org/10.1016/S1874-8651(09)60007-7

Yan, Z., Li, X., Zhang, Q., & Han, B. (2020). Seat allocation model for high-speed railway passenger transportation based on flexible train composition. Computers & Industrial Engineering, 142, 106383. https://doi.org/10.1016/j.cie.2020.106383

Yao, E., & Morikawa, T. (2005). A study of on integrated intercity travel demand model. Transportation Research Part A: Policy and Practice, 39(4), 367–381. https://doi.org/10.1016/j.tra.2004.12.003

Yao, E., Yang, Q., Zhang, Y., & Sun, X. (2013). A Study on High-Speed Rail Pricing Strategy in the Context of Modes Competition. Discrete Dynamics in Nature and Society, 2013, 1–6. https://doi.org/10.1155/2013/715256

Zembri, P. (2010). New objectives of the French high-speed rail system within the framework of a highly centralized network: A substitute for the domestic air transport market?

Zhang, Q., Yang, H., & Wang, Q. (2017). Impact of high-speed rail on China's Big Three airlines. Transportation Research Part A: Policy and Practice, 98, 77–85. https://doi.org/10.1016/j.tra.2017.02.005

Zhang, X., Lang, M., & Jin, Z. (2017). Dynamic pricing for passenger groups of high-speed rail transportation. Journal of Rail Transport Planning & Management, 6(4), 346–356. https://doi.org/10.1016/j.jrtpm.2017.01.001

Zheng, J., & Liu, J. (2016). The Research on Ticket Fare Optimization for China's High-Speed Train. Mathematical Problems in Engineering, 1–8. https://doi.org/10.1155/2016/5073053

Zheng, J., Liu, J., & Clarke, D. B. (2017). Ticket Fare Optimization for China's High-Speed Railway Based on Passenger Choice Behavior. Discrete Dynamics in Nature and Society, 2017, 1–6. https://doi.org/10.1155/2017/6237642

7.2. Candidate's publications

Musolino, G., Sgro, D. (2025). Estimating the Run Choice Behavior in High Speed Rail (HSR) Mode-Service: A Revealed Preference Survey. In: Gervasi, O., et al. Computational Science and Its Applications – ICCSA 2025 Workshops. ICCSA 2025. Lecture Notes in Computer Science, vol 15894. Springer, Cham. https://doi.org/10.1007/978-3-031-97648-3_28

Rindone, C., Panuccio, P., Sgro, D. (2023). Methodologies for Sustainable Development of TEN-T/RFC Corridors and Core Ports: Workers Mobility between Urban and Port-Related Areas. In: Gervasi, O., et al. Computational Science and Its Applications – ICCSA 2023 Workshops. ICCSA 2023. Lecture Notes in Computer Science, vol 14110. Springer, Cham. https://doi.org/10.1007/978-3-031-37123-3_42

Rindone, C., Sgro, D., Vitetta, A. (2026). Risk Reduction by Evacuation in

Transportation System: Pilot Survey with Technology Acceptance Model. In: Gervasi, O., et al. Computational Science and Its Applications – ICCSA 2025 Workshops. ICCSA 2025. Lecture Notes in Computer Science, vol 15892. Springer, Cham. https://doi.org/10.1007/978-3-031-97638-4_9

Russo F., Moschella M., Musolino G., Rindone C., Sgro D. and Vitetta A. (2025). Disaster risk reduction in urban area: survey design for mobility in evacuation condition. 2025 11th International Conference on Control, Decision and Information Technologies (CoDIT), Split, Croatia, 2025, pp. 183-188, doi: 10.1109/CoDIT66093.2025.11321566.

Russo, F., Musolino, G., & Sgro, D. (2025). Estimating run choice models with innovative data collection: Day-to-day tickets evolution in High Speed Rail. Transport Policy, 176, 103909. https://doi.org/10.1016/j.tranpol.2025.103909

Russo, F., Sgro, D., & Musolino, G. (2023). Sustainable Development of Railway Corridors: Methods and Models for High Speed Rail (HSR) Demand Analysis. In O. Gervasi, B. Murgante, A. M. A. C. Rocha, C. Garau, F. Scorza, Y. Karaca, & C. M. Torre (Eds), Computational Science and Its Applications – ICCSA 2023 Workshops (Vol. 14110, pp. 527–538). Springer Nature Switzerland. 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 O. Gervasi, B. Murgante, C. Garau, D. Taniar, A. M. A. C. Rocha, & M. N. Faginas Lago (Eds), Computational Science and Its Applications – ICCSA 2024 Workshops (Vol. 14822, pp. 188–201). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65318-6_13

Russo, F., Sgro, D., & Musolino, G. (2025). Modelling dynamic structures of fares in high-speed rail market. Transportation Research Procedia, 90, 734–741. https://doi.org/10.1016/j.trpro.2025.06.085

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. Computational Science and Its Applications – ICCSA 2023 Workshops. ICCSA 2023. Lecture Notes in Computer Science. vol 14110. Springer. Cham

7.3. References of MOST project

Di Gangi. M., & Russo. F. (2022). Potentiality of rail networks: Integrated services on conventional and high-speed lines. Computers in Railways XVIII. WIT Transactions on The Built Environment. 213. 101–112.

Di Gangi. M., & Russo. F. (2023). Design of hybrid rail services on conventional and high-speed lines. International Journal of Transport Development and Integration. 7(2). 113–121.

Moschella, M. (2025). Diodos 2: Approfondimenti sulla mobilità terrestre. Fondazione Filippo Caracciolo. https://fondazionecaracciolo.aci.it//app/uploads/2025/01/Diodos_2_Approfondimenti-sulla-mobilita-terrestre.pdf

Giunta. M. (2023). Trends and challenges in railway sustainability: The state of the art regarding measures. strategies. and assessment tools. Sustainability. 15(24). 16632.

Giunta. M. (2025). Data-driven track geometry defects localization and strategies for preventive maintenance: A case study. Transportation Research Procedia. 90. 742–749.

Giunta. M., & Leonardi. G. (2024). Framework for life cycle railway sustainability assessment: A methodological approach based on advanced methods and tools. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2024 Workshops (Lecture Notes in Computer Science. vol. 14822. pp. 233–244). Springer.

Giunta. M., Barrile. V., Leonardi. G., & Genovese. E. (2025). Comprehensive railway track monitoring using unmanned aerial systems (UASs) and building information modeling (BIM). In Lecture Notes in Computer Science: Computational Science and Its Applications – ICCSA 2025 Workshops (Vol. 15894. pp. 407–419). Springer Nature Switzerland.

Giunta. M., Frontera. P., & Er Rouisse. M. (2025). Circular economy in rail track: Estimating the environmental burdens of the use of recycled material and EoL processing through life cycle assessment. In Lecture Notes in Computer Science: Computational Science and Its Applications – ICCSA 2025 Workshops (Vol. 15894. pp. 391–406). Springer Nature Switzerland.

Musolino. G., Cartisano. A., & Fortugno. G. (2023). Methodologies for sustainable development of TEN-T/RFC corridors and core ports: Estimation of time-series economic impact. In International Conference on Computational Science and Its Applications – ICCSA 2023 Workshops (Lecture Notes in Computer Science. vol. 14110. pp.551–562).

Panuccio. P. (2024). State of the Art of Sustainable Development of Railway Nodes: The High Speed Rail (HSR). In: Gervasi. O., Murgante. B., Garau. C., Taniar. D., C. Rocha. A.M.A. Faginas Lago. M.N. (eds) Computational Science and Its Applications – ICCSA 2024 Workshops. ICCSA 2024. Lecture Notes in Computer Science. vol 14822. Springer. Cham.

Panuccio. P. (2025). High-speed rail stations in territorial areas: Sustainable development and quality of new landscape. In Lecture Notes in Computer Science: Computational Science and Its Applications – ICCSA 2025 Workshops (Vol. 15894. pp. 434–451). Springer Nature Switzerland.

Pellicanò. D. S., & Trecozzi. M. R. (2023). Methodologies for sustainable development of TEN-T/RFC corridors and core ports: Public incentives for industrial activities location in port related areas. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2023 Workshops (Lecture Notes in Computer Science. vol. 14110. pp. 635–646). Springer.

Praticò. F. G., & Fedele. R. (2023). Economic Sustainability of High–Speed and High–Capacity Railways. Sustainability. 15(1). 725.

Rindone. C., & Russo. A. (2024). A network analysis for HSR services in the South of Italy. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2024 Workshops (Lecture Notes in Computer Science. vol. 14822. pp. 217–232). Springer.

Rindone. C., & Russo. A. (2025). Centrality analysis for railway network planning: Constrained measures for High Speed. Transportation Engineering, 22, 100393. https://doi.org/10.1016/j.treng.2025.100393

Russo. A., Tesoriere. G., Rindone. C., & Campisi. T. (2025). Inter-urban transit connections: A network analysis in an Italian region. In Lecture Notes in Computer Science: Computational Science and Its Applications – ICCSA 2025 Workshops (Vol. 15894. pp. 375–390). Springer Nature Switzerland.

Russo. F., & Moschella. M. (2025). Mode choice at national level: Revealed and stated preference survey for high-speed rail realization. In Lecture Notes in Computer Science: Computational Science and Its Applications – ICCSA 2025 Workshops (pp. 359–374). Springer Nature Switzerland.

Russo. F., & Musolino. G. (2023). Methodologies for sustainable development of TEN-T/RFC corridors and core ports: Economic impacts generated in port-related areas. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2023 Workshops (Lecture Notes in Computer Science. vol. 14110. pp. 515–526). Springer.

Russo. F., & Pellicanò. D. S. (2023). Methodologies for Sustainable Development of TEN-T/RFC Corridors and Core Ports: The Role of Governance in the Export Time Optimization. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2023 Workshops (Lecture Notes in Computer Science. vol. 14110. pp. 622–634). Springer.

Russo. F., Chilà. G., & Zito. C. (2023). Methodologies for sustainable development of TEN-T/RFC corridors and core ports: Settlement capacity of industrial firms in port related areas. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2023 Workshops (Lecture Notes in Computer Science. Vol. 14110. pp. 539–550). Springer.

Russo. F., Moschella. M., & Musolino. G. (2024). Railway demand evaluation: HSR induced component. In O. Gervasi et al. (Eds.). Computational Science and Its Applications – ICCSA 2024 Workshops (Lecture Notes in Computer Science. vol. 14822. pp. 173–187). Springer.

Russo. F., Musolino. G., & Moschella. M. (2025). High-Speed Rail travel demand analysis: Pilot experimentation on generation model. Transportation Research Procedia. 90. 822–829.

Russo. F., Rindone. C., & Maiolo. G. A. (2025). High-speed railway planning for sustainable development: The role of length between conventional line and straight length. Future Transportation.