CORVINUS UNIVERSITY OF BUDAPEST

DOCTORAL SCHOOL OF BUSINESS AND MANAGEMENT

The impacts of Service Quality and Substitutes on Demand in the Hungarian rail passenger market

Ph.D. dissertation

Balázs Édes

INSTITUTE OF BUSINESS ECONOMICS
DEPARTMENT OF LOGISTICS AND SUPPLY CHAIN MANAGEMENT

Supervisor:

Attila Chikán,

Full Member of Hungarian Academy of Sciences

CONTENTS

Contents *3*

Figures *5*

Tables *6*

INTRODUCTION *8*

1 Context *12*

1.1 The rail sector in crisis: more questions than answers 12

1.2 The transformation of the railways: a clearer picture 15

1.3 The rail sector in Hungary 18

1.4 Origin-destination ticketing and clock-face scheduling in Hungary 21

1.5 The environmental relevance of the rail sector 27

2 Literature summary *30*

2.1 Sources of demand and the role of subjective factors 33

2.2 Modelling transport demand 38

2.3 Modal choice-focused models 46

2.4 Aggregated demand models 55

2.5 Models focusing on quality measurement 65

2.6 Conclusions 80

3 Methodology *81*

3.1 Research questions 81

3.2 Structure of the model 82

3.3 Scope of the data used 94

3.4 Building the model 119

4 Results 122

4.1 The basic model 122

4.2 Substitutes and other structural effects 127

4.3 Quality effects 130

4.4 Estimation results of the full model 135

5 Conclusions 138

5.1 Interpretation of the results 138

5.2 Evaluation of the research questions 141

6 List of references 144

FIGURES

Figure 1. Planning hierarchy in rail passenger transport ... 24

Figure 2. Example of an integrated timetable structure ... 26

Figure 3. Average external costs in passenger transport, EU, 2016 ... 29

Figure 4. Purpose and prioritisation of trips ... 34

Figure 5. Summary of results on price elasticity of demand ... 39

Figure 6. Routing scheme between competing stations ... 46

Figure 7. Mode-selection decision process ... 47

Figure 8. Logic framework of the model for the US long-distance market ... 49

Figure 9. Integrated modelling concept in high-speed rail forecasting ... 50

Figure 10. Modelling framework for high-speed rail forecasting in California ... 51

Figure 11. The planned state of the European high-speed TEN rail network by 2020 . 52

Figure 12. Fit of the check-in based model ... 54

Figure 13. Accessibility of stations in the evaluation criteria ... 69

Figure 14. Service quality tolerance zones and observed data ... 70

Figure 15. Service quality factors ... 71

Figure 16. The decision-making process for elderly passengers ... 73

Figure 17. Quality loop of public transportation ... 74

Figure 18. Results of the quality assessments ... 76

Figure 19. Sample question from the stated preference research questionnaire ... 79

Figure 20. Customer experience-focused customer journey schema for a travel chain 88

Figure 21. Logical framework of the model ... 93

Figure 22. Types of rail tickets before 2007: bark tickets and computerised tickets ... 94

Figure 23. Printed computer ticket after 2007 ... 95

Figure 24. Route planning discrepancies in destination data ... 108

Figure 25. Example of a traditional rail timetable format ... 109

Figure 26. Train and vehicle quality assessment form in the public service contract 118

Figure 27. Results of the estimation in integrated logical framework ... 139

TABLES

Table 1.	Environmental impacts of high-speed rail and air transport	28
Table 2.	Background on the factors shaping transport demand	35
Table 3.	Summary of time assessment studies	36
Table 4.	Subjective time assessment results (relative to in-vehicle time = 1.0)	37
Table 5.	Assessment of factors describing the service quality	40
Table 6.	Cross-price elasticities in long-distance transport	41
Table 7.	Variables of the Spanish high-speed rail model	59
Table 8.	Results of the Spanish high-speed rail gravity model	60
Table 9.	Elasticities of the Spanish high-speed rail gravity model	60
Table 10.	Variables of the model applied to European air transport	63
Table 11.	Estimating the competitive impact of rail based on time spent on board	64
Table 12.	Estimating the competitive impact of rail based on the presence of rail	65
Table 13.	Evaluation criteria for the quality questionnaire	66
Table 14.	Impact of journey times on rail demand, Spain, 1988-1991	67
Table 15.	Results of regression on the importance of evaluation criteria	69
Table 16.	Importance and satisfaction scores for quality factors	72
Table 17.	Evaluation criteria and passenger service indicators of a railway operator	75
Table 18.	Subjective consumer evaluation of the benefits of improvements	77
Table 19.	Evaluation of congestion levels by trip type	79
Table 20.	Proportionality of bus and car journey times	85
Table 21.	Comparison of bus and car journey times in relation to rail	86
Table 22.	Characteristics of the demand database	99
Table 23.	Variables used in the demand database	100

Table 24.	**The municipal background variables**	**101**
Table 25.	**Mixed community routes problems in Google DM API data**	**104**
Table 26.	**Access data of the MTA GEO database compared to the total sample**	**106**
Table 27.	**Comparison of GDM API and MTA GEO database data**	**107**
Table 28.	**Summary of the measurement data of the MÁF system**	**115**
Table 29.	**Quality evaluation criteria of the Public Service Contract**	**117**
Table 30.	**Availability and average cleanliness of Stadler EMU trainsets**	**119**
Table 31.	**Compilation of the demand data – structure, sources and links**	**120**
Table 32.	**The basic model with cross-sectional estimation**	**122**
Table 33.	**The basic model with fixed-effects panel estimation**	**124**
Table 34.	**Adding an annual lag variable to the base model**	**125**
Table 35.	**Adding economic variables to the basic model**	**126**
Table 36.	**Results of the substitution model, fixed-effects panel**	**127**
Table 37.	**Comparison of estimates of substitution variables**	**128**
Table 38.	**Impacts of Balaton and GYSEV dummies, random effects estimation**	**129**
Table 39.	**Balance of departure and arrival stations in the database**	**131**
Table 40.	**Statistics of different delay indicators in the model**	**132**
Table 41.	**Statistics on the number of daily services in the model**	**133**
Table 42.	**Statistics of the clock-face schedule in the model**	**133**
Table 43.	**Statistics of EMU trainsets in the model**	**134**
Table 44.	**Quality statistics of the MÁF system in the model**	**134**
Table 45.	**Weights of quality statistics of the MÁF system**	**135**
Table 46.	**Estimation results of the complex demand model**	**136**
Table 47.	**Comparison of the characteristics of quality impacts**	**142**

INTRODUCTION

For decades, the rail passenger transport sector played a special role among public services, which itself is generally an exciting issue from an economic point of view, both globally and locally. In most countries, the industry affects a wide range of people, with strong social, economic and political impact and is generally a highly critical topic in public discourse. It is in the centre of debate and criticism, an important user of community resources, while also generally a prominent employer.

Beyond the economic and social context, its role is also important from an environmental perspective, with increasing attention being paid to this mode of transport, which could alleviate many of these difficulties, in the context of the growing pollution and congestion trends of recent decades.

While this industry is in need of renewal and is struggling with countless difficulties like burdens and losses due to outdated technical and organisational solutions, it is becoming clear in more and more areas that what is actually needed at a societal level is to operate efficiently and to a high standard. It cannot be said that two decades ago the public was convinced that a discussion about the need for railways is of relevance; public opinion and politics alike tended to treat it as a problem that was tolerated and accepted as a necessity. Although railways have traditionally been a priority area for development in the European Union, at local level this policy direction is often met with a lack of understanding. The emergence of high-speed railways is a clear deviation from this trend, but relevant only for a minority of Member States.

Economic development and the motorisation that goes with it, suburbanisation trends, road infrastructure that is often still inadequate even after significant expansion, the problems caused by increasing congestion and pollution, particularly in large cities, are now making it increasingly clear that it is in the community's interest that more people use the different modes of rail transport. For such a change to take place, an attractive public rail service is essential: as passengers no longer choose this mode out of necessity, it must be able to offer a genuinely competitive service to meet growing demand. To achieve this, improvements are needed.

Since my university studies I have followed the development of this process in Hungary, but I have also come across a lot of interesting information from Europe and other parts of the world – and experiences from my travels – over the years. In the early 2000s, I often had to justify my choice of topic, but today the situation has changed noticeably in academic and everyday communication, as well as in professional and public policy discourse. Many people have personal experience of more advanced, modern, user-friendly rail services in other countries, and understand the importance of how a region or country can use this powerful tool to develop its own economy, solve social problems and reduce the impacts of environmentally damaging activities.

The progressive approach of transport is also becoming increasingly prominent among public policy issues. The dominance of a car-only concept is still prominent in the debate, but the importance of ecological aspects are strengthening. Urban cycling is perhaps the area where this shift is most visible in Hungary. While urban cycling is back in the spotlight after a century, more importantly, it is a subject of constant debate, with more and more people becoming aware of their transport choices, while an active, movement-like civil attitude does not allow these issues to be ignored. In some areas, the rail sector has been revived from its previous hibernation, where the growing role of European Funds is the most important factor. At the same time, elites who only know public transport by reputation continue to do significant damage to decision-making by ignoring rail, which is not a Hungarian or Eastern European specificity, but an absolutely global problem, often resulting in significant deviations from the social optimum (Walker, 2017).

In recent years, I have personally had the opportunity to see the sector up close at times. I worked for two years at MÁV-Csoport – one year in the CEO's office – and then for more than three years at BKK, where in a few years the biggest development in Budapest transport since the 1980s took place (Vitézy, 2014). These processes have been partly interrupted and transformed, but even with many difficulties, it has been an extremely positive experience to see the impacts that well-established improvements can generate; that even small but well thought-out steps can make the daily lives of hundreds of thousands of people easier.

Modern business and technical knowledge has only an isolated presence in public transport services in Hungary, where outdated practices, inefficient procedures and non business-driven logic still dominate the industry. At the same time, more and more developments and up-to-date service components are being created, which are often isolated and unconnected elements at the state railways.

Data-driven operations, decision making and the use of business intelligence can be considered as such a hot area. These are categories that even the large enterprises are just learning, so it is no wonder that this is not a natural way of doing things within the state railways. For both MÁV and BKK, I have spent a lot of time identifying potential data sources and laying the foundations for an integrated analytical framework. Building such a data system requires much more time and resources than are typically available in the peaceful 2-3 year periods between hectic management changes, so in the case of MÁV the very limited time frame was more for exploring the possibilities, and according to my knowledge the task has not been fully completed even to date. However, it has become clear how much missed potential there is in the data that has been collected but not linked and used for meaningful analysis. In the case of BKK, I also had the opportunity to be involved in the development of the data warehouse, the first phase of which was largely completed by 2014, and then the process slowed down after the change of management. The missed opportunity here is perhaps even sadder, as the data collection and linking was partially completed, and real-time data sources were already available, but the decision support and use of these data is still 10 years away. In general, it is typical of these large companies that a single department collects data and information on its own operations in a more or less digitized way, but the use of this data is limited to internal auditing, controlling and providing data for the management. In the case of MÁV, for example, data on the stock of real estate, vehicle allocation, delays and ticket sales were already available in 2012, but these were presented in the reports in an unlinked, juxtaposed way, i.e. only at the level of showing different distributions and averages – the need for, and the opportunity for any kind of deeper analysis did not arise. The lag is illustrated by the fact that the railway company only launched a call for tenders for such a system in 2018 (MÁV, 2018).

In my view these topics are very likely to be strongly connected: transport development, encouraging the use of socially useful rail transport, is a task that requires a conscious, measured and planned improvement in the quality of services. The importance of the subject for the Community, decisions with an impact over many decades, and the extremely high development costs all justify the need to collect, systematise and analyse the otherwise available data to a high professional standard, as this is a much more focused and effective way of developing transport.

The aim of my research is to investigate how the demand for passenger transport in Hungary is influenced by service quality and the impact of substitutes on demand. On the one hand, my aim is to show that by combining existing available data with appropriate methodology, it is possible to create an analytical tool that can be of interest from both a scientific and a decision-support perspective. On the other hand, the aim of the research is to investigate what can be learned about the demand for rail services through the use of this tool. What is the impact of quality factors and what improvements might be needed to make railways more competitive. By comparing the factors and examining their relationships we can gain new perspectives for choosing between development options and making decisions to use resources as efficiently as possible.

The initial data source of the model is the ticket sales database of MÁV-Start, with which I try to model the development of demand as closely as possible by supplementing it with available data on service quality, or data on other economic background variables that can be derived from some indirect source.

The thesis is structured in the following way: in the first chapter I provide an overview of the Hungarian rail passenger transport sector, highlighting the factors relevant to the research topic and the most relevant background information related to the data sources. This is followed by a literature review in which I summarise measurements and their results with a broader perspective, covering the range of methodological tools used specifically on models similar to the solution I will apply.

During the presentation of the model, I will formulate the research questions, the theoretical structure of the model, and the data sources used. After the introduction of the data sources, I will describe the compilation of the database, highlighting the modifications and corrections relevant to the results. I will then present the outcomes of the model and compare them with previous literature, and summarise the main conclusions of the research.

1 CONTEXT

1.1 The rail sector in crisis: more questions than answers

The rail sector is often an important symbol for a country: in developed economies it is often seen as an object of pride representing the technical and economic development of a nation, or as a symbol of its crisis (Connolly, 2018). The railways have made a huge difference since their emergence, and in many developing economies they are still seen as both a depository of civilisation and a force that holds the nation together, where they are figuratively a true national institution (Wolmar, 2018).

For the first hundred years, the technology replacing animal power was one of the most advanced industries of the age, a driving force, and its role in our time is most comparable to that of information technology. Railways attracted attention, capital, large-scale speculation and the potential for excellent returns, often built and operated by large private operators. Railways were the most efficient mode of land transport and the most modern form of passenger transport. It may seem surprising to today's eyes that a Romantic poet concerned with the great affairs of the nation would use a poem about the railways that pathetically deifies technical innovation and radiates technological optimism, as was the case in Hungarian culture, but there is reason to believe that this poem is representative of the spirit of the age (Petőfi, 1847).

After another century, the development of the internal combustion engine and road technology has brought a huge change to transport. Road surfaces and the running gear of vehicles on them together have evolved so rapidly that, with increasingly affordable vehicle production, motorisation has become a real alternative and an increasingly powerful competitor to rail. From a marginal role before the Second World War, road transport became dominant in both long-distance and local traffic within a few decades, a change that affected both passenger and freight transport.

The importance of the railways steadily declined towards the end of the twentieth century, falling further and further behind as road and then air transport became more and more important. In the meantime, however, state-of-the-art super-expresses still enjoy the same privileged status in the eyes of the ordinary traveller as they did in the first railway era

(Fleischer, 2006). This new rail renaissance, of course, mainly affects those – from China to France, Poland to Japan – who are in the fortunate position of having access to rail services 2-3 times faster than car (Bode *et al.*, 2018). For the majority, however, rail is now much more about a somewhat disparaged service of precarious quality, often operating with outdated equipment, rust-belt metropolitan areas, untidy properties and crowded, uncomfortable journeys (Duranton *et al.*, 2017). This dichotomy creates tensions as well as specific situations; in Hungary, for example, the preparation of a high-speed rail project is taking place while the question is whether the journey times already achieved in the 1970s can be achieved today on a large part of the main rail network (Hörcher, 2021).

This dichotomy can be observed all over the world, including in European rail networks, and in the last 2-3 decades, especially with the spread of high-speed rail networks, there is a clear trend that countries and regions that have neglected this sector are starting to rediscover its importance or at least to reflect on the issues involved (Luk, 2017). In metropolitan and suburban regions, road congestion, excessive motorisation and environmental damages are leading to a growing recognition of the importance of rail transport, and in urban regions where it has been marginalised for decades, rail transport is increasingly coming to the fore. One important consequence of this change is that the position of infrastructures and services that have long been in the undecided category in public policy is beginning to change, as decision-makers have not dared to take any decisions about them (neither their removal nor the necessary major upgrading has often been taking place for decades). Thus, these systems and sectors are stuck in this hibernated state, gradually amortizing, conserving the transport policy and urban planning concepts of the past, waiting for a better fate, and in more and more places the balance is tipping towards preservation and development.

A striking example of these trends is New York: by the 1940s, the tram network in Manhattan had virtually disappeared, thanks to the intervention of the car industry[1]. Over the past decade, the city has attracted international attention for its use of underutilised rail

---

[1] It was not a consequence of economic necessity per se, since it was a well-functioning form of transport: the automobile and rubber industries, organised by General Motors, actively facilitated the process (which was later condemned by a court for economic conspiracy). This conscious strategy led to the rapid construction of urban and suburban tram networks in America (Kwitny, 1981).

infrastructure, with the New York High Line project taking a model approach to reclaiming former rail infrastructure as a green urban interface (The High Line, 2018). As a sign of this new approach, a current plan is to start construction of the first new line, which could be called a tram in European terminology, in the near future (Wright, 2016).

Depending on when a country or region has abandoned its railways, and to what extent it has neglected them, very different situations can arise. The central regions of the western half of Europe, Japan and the highly populated areas of the world have generally preserved their main railways better during the 20th century, and today they use the latest modern technologies on their networks. In the case of less modernised countries, on the other hand, development policy that recognises the importance of rail must often start from single-track, low-speed, non-electrified rail lines, but such situations also occur for branch lines in rail-developed countries such as Germany (Allianz pro Schiene, 2007). The most extreme example might be the rebuilding of lines that have been closed down, and such ideas are also regularly considered in countries and regions that have cut back their networks substantially in the past (Topham, 2017).

A symptomatic feature of rail services is that technical-infrastructural aspects dominate development, while the specific aspects of service level and customer experience, now fundamental development directions of service sectors, are often overshadowed or even left out. Trends that are now absolutely fundamental, such as service design (Holmid, 2005) or the user experience (UX) approach (Norman, Miller and Henderson, 1995), have become embedded in the majority of service business thinking and have radically changed it in a few years, but have left the rail sector and services in Hungary (Közlekedő Tömeg Egyesület, 2018) and in more developed regions almost untouched (Nohe, 2017). However, there are growing signs that this situation will undergo radical change in the coming period (LA CoMotion, 2017).

1.2 The transformation of the railways: a clearer picture[2]

With the expansion of motorisation throughout the 20th century, apart from a few priority projects (such as the Channel Tunnel, now used by the Eurostar super-express, or the development of the ICE and TGV) the development of rail services in Europe in the 1990s was characterised by a kind of stagnation. The technologies have not changed for decades, and the industry's attention has been focused mainly on the European Union's rail regulatory activity. In 1992, the EU drew up the Common Transport Policy, which, by definition, focuses on the TEN corridors, i.e. the routes connecting the Member States instead on domestic transport at national level (European Commission, 1992). The long-term regulatory interventions of the integration policy are linked to a strong vision of an interoperable, competitive European rail core network (European Commission, 1996). In the period under review, the public debate focused on the questions and possibilities of opening up the rail market. These issues about the regulation and market structure were in line with the liberalisation under way in the telecommunications and energy sectors, where the processes were ahead and already had some promising results (de Jong, 2017). While this discourse has been persistent ever since (The Economist, 2018), a whole new external influence has galvanised the more dynamic, business-minded segment of the rail sector.

Aviation has become the industry where, after the uncertainties of the 1990s (Hansson, Ringbeck and Franke, 2003), low-cost airlines completely disrupted the market in a few years (Centre for Aviation, 2011), and this growing competition, with digitalisation, advanced pricing technologies, online sales, innovative business solutions has led to a breakthrough in a few years (The Wall Street Journal, 2017). Among many other factors, this also had a defining impact on the long-distance rail market through four major mechanisms.

First, and most importantly, with a very determining downward trend in prices, air transport has become a real competitor to rail transport, as the previous significant price differentials have disappeared and shorter-haul flights have become not only faster but, partly due to international tax anomalies, often a cheaper commodity, a real substitute. This has created strong competition that poses a serious challenge to the rail sector.

---

[2] This sub-chapter is based, among other things, my previous work (Édes, 2019)

Secondly, the low-cost revolution, and the transformation of aviation in general, has brought innovative service processes that have changed customer preferences. More and more travellers have become accustomed to online ticketing, convenient search, booking and payment systems; much simpler processes than before, and the use of user-friendly interfaces (Diaconu, 2012). While it is now often the low-cost carriers that are beginning to incorporate less customer-friendly practices into their models (Independent.ie, 2018), in the first decade they explicitly led this innovation and raised the level of customer expectations.
Thirdly, at European level, aviation has made the long-distance transport market more international and interoperable than ever before. Although the rail and air networks were already an important part of the international transport system, the greater accessibility of air travel has made the transport markets of EU and beyond more integrated for customers. This has also meant that the more inflexible and outdated way in which rail has handled border crossings and international services has become more problematic than ever before. As a reaction to strong competition, a fairly integrated high-speed rail network has now evolved in the more developed western region of the EU, where rail is now a fully-fledged alternative to air travel for example between Brussels, Paris and London, which was not nearly so obvious to customers in the past.
The fourth factor to be highlighted is the reaction of business operators, as the transformation in aviation has fundamentally shaped the more receptive managers of rail operators. With slower or quicker reactions depending on the area, cultural attitudes, capabilities and opportunities, but for almost all major rail operators, services and model elements that work well as a complex package in aviation have emerged over the last one and a half decade: flexible, capacity-dependent pricing, online ticketing, telephone applications with real-time information and shopping, etc. The take-up is slow but it is also progressing in Hungary, where formal elements rather than content are dominant, but the trend is clear: it is much easier to develop an application than, for example, to create the legal and technical conditions for a flexible, demand-based pricing model.

After a decade and a half, the revolutionary transformation of the air transport sector seems to have had a significant and lasting impact on rail transport, with the industry undergoing significant development and competition – especially in the case of high-speed rail – generating continuous interactions between the two sectors (Cadarso *et al.*, 2017). The most innovative railways are also focusing on expanding into the whole travel chain, and are also trying to cover the urban part of the journey with additional services such as car or bike sharing (Deutsche Bahn, 2018).

Meanwhile another, perhaps even more important change has occurred, which is basically in line with social and environmental impacts. Railways have been rediscovered in many developed economies, coming back into focus as a means to alleviate the problems of congested urban areas, thus presenting new expectations for the sector (Lalive, Luechinger and Schmutzler, 2013). The two effects have pushed the development of the rail sector in a similar direction, to find a truly innovative, customer-centred model, a more prominent participation in the transport division of labour, the development of rail has again become a public interest increasingly recognised by many (Jaffe, 2012).

The issue of participation in the division of modes has become increasingly relevant in recent years, with a new approach to the issue of how to manage the mobility mix resulting from the combination of public and shared transport modes in a coherent way to find a truly competitive alternative to motorisation. The Mobility-as-a-Service (MaaS) concept is a common form of this consideration, approaching the problem in a unified framework. Of course there have been transport systems in the past that have sought to combine as many modes and forms as possible. Typical examples are regional transport systems, such as VOR, the Vienna City Region system (Verkehrsverbund Ostregion), which combines the entire urban public transport system with the suburban rail and bus network (VOR, 2021). The mobility system in Zurich has been offering a shared car service since 1997 (Mobility, 2021), which was already partly integrated in terms of customer management through a common card system (Swisspass, 2021). However, a development in line with the MaaS concept is the creation of a single access point for all these services: the ZüriMobil phone application was launched in 2021, making these services available through a single user interface (ZüriMobil, 2021). A prominent example is the Whim system in Helsinki, where a subscription-based scheme makes five modes of transport available in a highly integrated way: urban public transport, taxis and car rental, e-scooters and bicycles (Whimapp, 2021).

In conclusion, the transformation of rail transport over the past decades is far from complete, and it is one of the most exciting processes in the service industries, where centuries-old practices and habits are mixed with high-tech solutions. In this transition a myriad of business and economic dilemmas are brought to the fore, making it an exciting subject of theoretical and scientific investigation. All of this can be explored with an unprecedented breadth and colour of analytical tools, building on the unprecedented quantity and quality of data resources made available by digitalisation. In addition, the field is still relatively unexplored in many places, as there is often a lack of business motivation due to the passive attitude of the public sector giants. Some measurements, research and data do not exist neither as scientific, public data, nor as company secrets. Often the data that companies do have has not been examined before using deeper, more complex analytical tools. To my knowledge, the data sources of the Hungarian State Railways that I have examined are also a largely unexplored research territory, both from an academic and business perspective.

1.3 The rail sector in Hungary

The rail sector is a key player in the Hungarian economy in several respects. With 38,000 employees (Homolya, 2018), MÁV and the members of the group together are the largest employer in Hungary before the integration of the railway and main private bus operator (only OTP Bank would come first, foreign employees included), and GYSEV is also part of the railway sector with a workforce of around 5,000 (GYSEV, 2018). As of 1 January 2021, Volánbusz also joined the MÁV group, bringing the total group's headcount to around 57,000 at present (wikipedia.org, 2021a).

The two railway companies manage a total network of more than 8,000 km of track, which carries a significant share of suburban and long-distance passenger traffic and freight. Over the last decade, rail has consistently accounted for a share of passenger transport close to 10% (8.6% in 2018), while in the freight modal split it accounts for a quarter of the performance, 27.0% in 2018 (KTI, 2021).

The total consolidated expenditure of the MÁV group amounted to HUF 582.9 billion in 2020 (MÁV-csoport, 2021b). The rail network it operates has deteriorated significantly over the last 30-40 years as a result of neglected maintenance and renovation work, with speed restrictions in place at nearly 2,000 locations, more than half of which are permanent. As a passenger transport company, MÁV-Start has the oldest rolling stock in the region, with an average age of 35 years. Although from 2006 onwards the Hungarian railways received modern EMU trains in several waves through significant acquisitions, the scale of these improvements alone could not bring a breakthrough in the overall picture of the extremely deteriorated fleet, as the average age of EMU trains is still 23 years, but most of the conventional wagons and locomotives have already exceeded 40 years of service (Homolya, 2018).

Since the fall of communism, the history of railways in Hungary has seen countless changes, including the introduction of modern rolling stock in suburban traffic, the complete renovation of several lines with the help of European funds, the increasing availability of online ticketing, and the introduction of a fully-fledged mobile application for Hungarian state railways.

At the same time, much of the rail network operates with an unsatisfactory line speed and is therefore not competitive with car travel, the ticket pricing model is completely inflexible, has no capacity-related, meaningful incentives and, despite full digitalisation, remains on an analogue logic. The customer experience and communication are at least half-hearted, often of rather poor quality, and the handling of disruption and service reliability are unsatisfactory. An electronic ticketing system has still not been introduced. While the breakthrough successes of European integrated station developments with a real estate development approach have been visible for more than a decade, in Hungary, in the case of railway real estate, even operation and maintenance is extremely difficult, rather than utilisation.

In a simplified and summarised way, the situation of the Hungarian railways and the thematisation of thinking about transport in the last thirty years can be characterised by three eras. The change of regime and the first decade was a period of collapse and disorientation for the railways, with a decline in supply and traffic, a worsening financial situation and no way out. The railways were operating on their previous reserves and the quality of service was moving further and further away from the level that would have ensured maintenance; improvements were also lagging behind. Perhaps the tragic great rail strike of 1999[3] and further union action in the period can be seen as the symbolic end of the era.

In the following decade, the focus of thinking on the railways increasingly shifted to EU accession, with the debate on branch lines dominating the professional discourse of the period, alongside issues related to liberalisation and the choice of the ideal market model. Although according to some views – and it is perhaps difficult to argue – this debate was a distraction from the more important issues (Mihályi, 2004), paradoxically it was the subject of recurrent discussion by all actors. In the end, traffic was suspended on some of the lines between 2006 and 2008 – a closure that was not formally accepted by the decision-makers – and then resumed on some of the sections in 2010, but generally only token traffic, with two or three trains a day. In the meantime, the liberalisation process was under way: the first private rail freight services were able to emerge with difficulty, and within a few years the institutional system of rail regulation was created, then, in a peculiar way, immediately dismantled and only restarted in a limited form (Édes, Gerhardt and Micski, 2011). With EU accession, new rolling stock procurement started in this period, and track renewal projects intensified (much of the preparation took place during this period). The process of rail liberalisation in Hungary remained rudimentary, but the similar attitude of other larger Member States meant that the country did not face any major difficulties in the larger EU context because of this.

[3] The stress of the strike has left the CEO unable to cope and died suddenly.

In the third decade, the intensity of these chaotic processes has slightly decreased, although the railway sector is still not operating in an orderly environment: the liberalisation of passenger transport has been – unspoken – postponed by the government, and market competition as a motive for development is thus permanently absent from the operating environment of the railway sector. The pace of permanent changes of leadership seems to be slowing down somewhat: while the average tenure of MÁV presidents was less than 2 years between 1990 and 2012, there has been no change in this position between 2012 and 2018. There is still constant movement within the railway organisation, with the eternal process of transformation, mergers and splits that has been going on for years. At the same time, the institutional environment of railway management has been somewhat more stable and the improvement in financing conditions after the previous indebtedness, with the state as owner allowing a moderate but steady reduction in debt, has been a major positive.

The business development processes that started in the second half of the 2000s, such as the origin-destination ticketing, the clock-face scheduling system, the expansion of internet services, have gradually taken hold, and today we have a fragmented situation with both modern and backward elements. Two of these, relevant to my research, are described in more detail below.

1.4 Origin-destination ticketing and clock-face scheduling in Hungary

One of the neuralgic points in the history of the Hungarian state railways is the lagged development of the ticket system. In the last decade, the world's leading railways have increasingly sought to take the advantages of flexible pricing methods known from low-cost airlines into their traditionally highly regulated and highly constrained rail pricing systems. These flexible pricing systems are using complex mathematical models, are based on demand characteristics, congestion probabilities, knowledge of demand for specific flights, are much more efficient, dynamic and, overall, much better for passengers. At the same time, the challenge for the Hungarian state railways is to draw the right conclusions about passenger traffic from sales data. Of course, trends can be seen, significant segments of traffic can be estimated, but the picture is distorted and incomplete, and is delayed, so the feedback into supply side is not provided.

For centuries, the Hungarian railway network has maintained the practice of tickets containing only the place of departure (stamped, pre-printed in busy areas or printed on the spot) and the valid distance. So if a passenger wanted to travel from Budapest-Déli station to Siófok, it was possible to buy a ticket for 120 km travel, which would show only the departure station in addition to the distance. With this model, despite the development of computerised ticketing in the 1980s, it was not possible to estimate how many people were actually travelling between various stations, based on the data in the computer system at the time. Given that the most populous and heavily trafficked settlements are typically nodes in the Hungarian railway network, it was not even possible to classify the traffic by railway line for the tickets sold there. Demand analyses, and thus the decisions taken, were therefore based primarily on passenger censuses, other questionnaire surveys and often on the empirical experience and intuition of staff.

After these precedents, the introduction of origin-destination ticketing on 1 July 2007 was a significant step forward (iho.hu, 2009). With the new system, the ticket shows both endpoints, so the sales data recorded in the system can now be used to determine the exact point of departure and arrival of a given journey. Since then, the system has been extended in three stages with online ticketing, first in 2008 with user printed online tickets at station ticket machines, then in 2011 with printing at home and finally a fully digital version in 2018. The usability of the interface is still limited, it is more difficult to buy tickets than in the applications of some other rail companies in Europe, but improvements have reduced the gap. A fully electronic ticketing system, integrated with other operators, is still to come.

Clock-face scheduling is a system in which the timetable of a given type of public transport service is structured so that services follow each other at the same interval in both directions of the service (itf.hu, 2008). The time period of repetition in the timetable can vary greatly: while a high density service in a city centre, such as tram or metro, may have a 5-10 minute headway, long-distance services such as railways tend to have a following of half an hour, an hour or two hours. At less frequent intervals, the user benefits of frequency might be lost. The most important of these is the creation of a transparent, predictable and easy to remember timetable, with trains departing at the same time every hour from a stop.

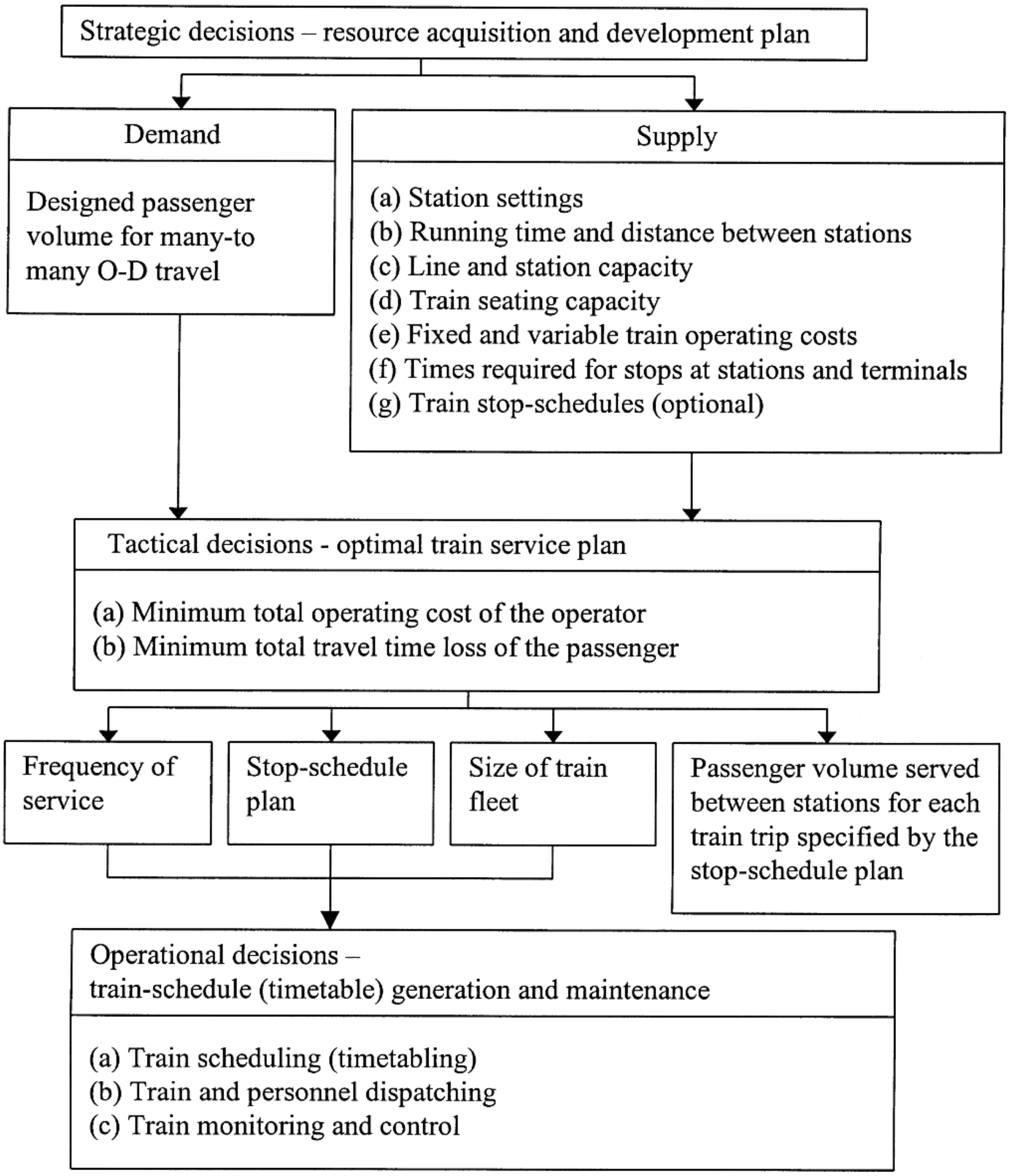
While the concept of the clock-face scheduling could be relevant for any service, its role in the case of railways is especially important in terms of improving quality, because the supply of this typically underfinanced public service has tended to decline over the years during the difficult economic periods after 1990. This works against regularity, and the less frequent the train departures, the less user-friendly the service – the longer the average waiting time, or the more time the passenger has to somehow bridge if he or she does not want to spend time waiting.

In contrast, the concept of clock-face schedule-based transport, which has developed in the German-speaking world and is most widespread in this region, has brought a supply-side approach, based on the recognition that below a certain level of frequency, service is in fact unacceptable to passengers, so that demand decreases not in proportion to the reduction in supply, but much more dynamically below the critical level. According to this approach, rail passengers are not 'forced' passengers, i.e. they use the railways not because they do not have a car or other alternative.

Achieving the transport policy objectives of increasing rail's share in modal split requires improving the quality of supply to make public transport services an attractive alternative, with a dense and memorable timetable that is closer to fully flexible individual transport. The real strength of the concept lies in the economic recognition that a significant proportion of the costs of rail services are fixed costs or variable costs that cannot be changed in the short term, so that the savings from not running services are in fact much smaller than the data from the controlling systems would suggest. Whether a vehicle and its crew spend 16 hours a day or only 6 hours a day in useful service, the costs are almost the same. As the additive amortisation effect in rail technology is negligible, the track access charge remains within the rail holding (unpaid charges for unused services are a loss to the rail organisation). The main additional cost of transport is the energy cost of switching a timetable from a less frequent to a more frequent service to make better use of existing staff and rolling stock. The additional costs are returned for relatively small amounts of additional trips, and the concept is that the attractiveness of a sustained higher frequency offer can lead to a significant passenger surplus. Experience with the introduction of frequent services in Hungary has also been positive (MÁV-START, 2014).

This logic can be easily followed in the figure below by Chang et al. In their work, they have used complex mathematical tools to determine the optimal service offer and timetable concept for a new line (Chang, Yeh and Shen, 2000), and in this context they show the logical framework of the process by designing the services of a line. In Figure 1, the optimization problem between supply and demand needs is summarized in a broader context.

Figure 1. Planning hierarchy in rail passenger transport



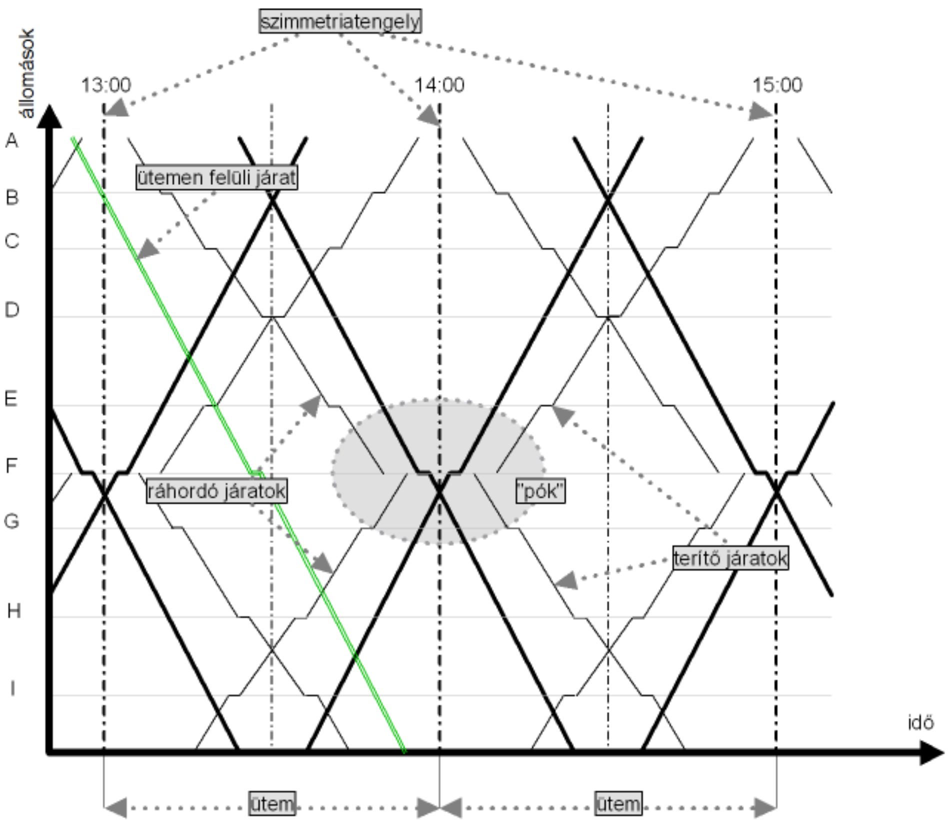
Source: (Chang, Yeh and Shen, 2000)

The more frequent departures are, the less travel time is lost for the passenger, while minimising costs is the limit to optimisation. The optimization of the supply volume, besides the frequency and the planning of stops, is mainly related to the fleet size, while the capacity is already given by the frequency and the fleet size. Operating this resulting capacity with reduced utilisation will obviously result in significantly lower savings due to the presence of fixed costs and variable costs that cannot be changed in the short term, following the logic presented above.

The integrated clock-face scheduling (as abbreviation ITF, from the German expression 'Integrierter Taktfahrplan') is an improved version of the concept of the clock-face scheduling. In the case of an integrated model, 'the timetable in a public transport system covers more than one type of service (e.g. intercity and connecting passenger train) or even several public transport sectors (e.g. train and shuttle bus), so that the timetables of the services are spatially synchronised and planned according to uniform principles (e.g. with the same time symmetry)' (itf.hu, 2008).

The timetable graph shown in Figure 2 is a schematic drawing of an integrated clock-face scheduling structure. It clearly shows that the connecting nodes (called spiders in ITF terminology) are where the faster (in Hungarian railway terminology: zoner trains) and slower services (passenger trains) serving the other stops meet, thus acting as a feeder or spreader. This means that trains do not have to stop at all on the accelerated section, and the transfer without any loss of time ensures the benefits for all concerned.

Figure 2. Example of an integrated timetable structure



Source: (itf.hu, 2008)

In Hungarian railway practice, the timetable structure resembling a clock-face scheduling has a long history, mainly for timetable-editing reasons – a batch structure facilitates the timetable of a railway line with a mix of many types of trains and a large traffic volume – and such a timetable has appeared in several cases. However, there was no consistent concept behind these practices, which typically also meant that regularity always included exceptions, so that there were always one or two trains in a day that were missed or did not depart at the same time as other services managed in a consistent structure, thus preventing the important benefits of the concept from being realised.

The first truly structured, regular clock-faced schedule was introduced in 2004 on the Budapest-Vác-Szob and Budapest-Veresegyház-Vác lines ((itf.hu, 2016). The introduction of hourly zonal trains in the Danube bend has led to a significant reduction in the duration of journeys to and from Vác, which in addition means a journey without transfers (in classical timetable structures, a similar reduction in journey times could be achieved by transferring to a fast train, but this is inherently less convenient and often unsafe due to the transfer).

The Vác clock-face schedule was already integrated at the time of its introduction, and even non-rail services were included (some bus services, as well as the timetable of the Királyrét Forest Railway and the Dömös ferry), a level of integration that still represents a unique level of cooperation in Hungary (itf.hu, 2016).

1.5 The environmental relevance of the rail sector

The environmental impact of transport has been at the forefront of public policy debates for decades, as it is one of the biggest polluters at global level, alongside energy production and industrial production. The particular importance of this issue is that the magnitude of negative externalities is strongly linked to the choice of transport mode: there are significant differences in the levels of environmental impact between different transport technologies, while interventions to reduce pollution are difficult to implement because of the often high costs and other barriers to switching modes. While in the energy sector, building a new power plant represents a commitment and a significant investment over many decades, building transport infrastructure involves decisions with a century-long impact and the costs of different solutions can be many times higher.

The literature on the subject is extremely diverse, the measurement techniques and methodologies are difficult to follow, and the results are usually guaranteed to generate considerable debate, in line with the fragmented interest structures, but a few illustrative examples integrating the results of several studies point to the importance of the subject.

In 2003, Janic compared the two most frequently studied substitutes, high-speed rail and air transport, from an environmental point of view, using data from previous publications (Janic, 2003). His main results are presented in Table 1, where the main motivation for public policy efforts to encourage the use of rail transport is clearly shown, namely that high-speed rail has much more favourable indicators, with only a 13% advantage over air in terms of land use. For the calculation of air pollution, the author has chosen only carbon dioxide emissions from many possible indicators, where the very large difference between the

French and German values is due to the difference in the energy mix between the two countries.[4]

Table 1. Environmental impacts of high-speed rail and air transport

	Rail		Airplane (seat capacity)			Units
	TGV	ICE	(100)	(150)	(400)	
Energy	0.19	0.22	0.38	0.59	1.62	kWh/pkm
Air pollution	4.0	27.5	99.8	153.9	424.9	CO_2 g/pkm
Land-take	2.86			3.23		million pkm/ha/year
Safety	0.00114			0.00580		deaths/billion pkm/year

Source: Author's compilation based on (Janic, 2003)

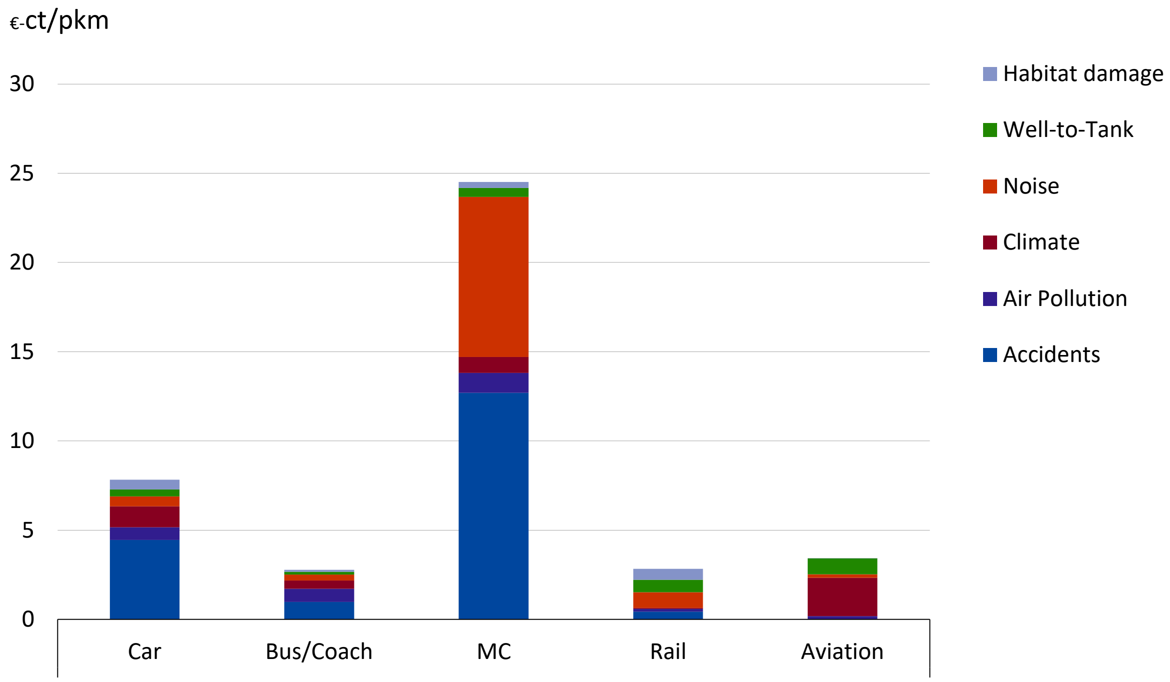
A much more complex methodology was used by Rozycki and co-authors in their analysis, also published in 2003 (Rozycki, Koeser and Schwarz, 2003), where they looked at the whole life cycle and all environmental aspects of the materials used. Thus, they take account of the materials used to build the railway track and the energy used in their production, just as they collect the energy requirements of operation, from the energy used by management offices to the consumption of the individual elements of the line infrastructure concerned, to the actual total consumption. The details of this are beyond the scope of this chapter, but are certainly appropriate to draw attention to the complexity of the subject. The sources of the above comparison published by Janic are also complex measurements that process a lot of data in an appropriate way, but it can be seen that a much deeper investigation of the subject is possible.

The ratios between energy consumption and air pollution data clearly show the advantage of rail transport over air transport in terms of environmental impact.

[4] While Germany had already minimised the use of nuclear power in the period under review, and phased it out completely after the Fukushima accident, in France the nuclear source was close to 90% in the period (Janic, 2003).

The Commission published the preliminary results of a comprehensive recent EU survey in December 2018, summarising the latest measurement results on this topic from 2016, and the externalities are shown in Figure 3.

Figure 3. Average external costs in passenger transport, EU, 2016



Source: (European Commission, 2018)

As can be seen in Figure 3, the choice between road and rail passenger transport has a very significant impact on the externalities caused by transport systems. There are substantial differences in terms of both the costs borne directly by society and the environmental burden.

This leads to the clear conclusion that the competitiveness of rail in a society is an important issue from both an environmental and a climate point of view, as the use of much more polluting modes of transport can be effectively reduced by offering good quality and attractive services. This is complemented by a wide range of local impacts, as rail is the most effective means of tackling local pollution and congestion, especially in metropolitan regions.

2 LITERATURE SUMMARY

Measuring and modelling the demand for transport services, including passenger rail transport, is a field of scientific and business research with a long history. The relevant literature is typically oriented towards economic and development policy. For rail capacity studies, the most critical period in the whole life cycle is planning, particularly because the time-horizon of investments might be measured in a hundred-year time frame. This phase contains among other tasks technical design, preliminary impact assessment, business planning, land use impacts, and other similar forecasting. In the case of rail capacity, the construction or upgrading of rail lines to more modern technology – in fact, more like rebuilding, as there is typically a complete replacement of technology – involves huge volumes of investment, with projects of magnitude greater cost than the average infrastructure development and investment decision. Therefore, estimating and forecasting expected utilisation, demand and capacity needs is of paramount importance in the industry.

While in the classical era of railway construction, development was typically carried out by private operators in the market, over the last 50-150 years, there has been a trend towards public ownership of these projects worldwide, and nowadays non-state actors are typically only involved in the construction. Depending on whether a market consortium or a public institution is the implementer, the forecasts may be aimed at market demand and returns, or, in the latter case, in a more complex approach, at rural development, territorial integration, economic development, expected take-up: whatever the case, it is necessary to forecast demand and estimate expected returns.

Accordingly, the literature on the demand for rail passenger services tends to focus on forecasting, with published papers and research generally analysing historical data for predictive purposes.

Measures of the elasticity of demand for transport modes are summarised and analysed in a review by P.B. Goodwin published in 1992 (Goodwin, 1992). The author reports a peculiar phenomenon: comparing his earlier work with a North American publication on a similar topic (Oum, Waters II and Yong, 1990), he indicates that there is almost no overlap between the total of about 150 sources cited in the two articles. According to Goodwin, this is only partly due to the weak link between the scientific public life of the two continents and the not quite identical focus of the studies – just as importantly, there is a fundamental separation between the traditional scientific professional approach and the policy-oriented, government-commissioned research.

The literature mainly applies two basic theoretical frameworks for these analyses. The first is the set of micro-level mode choice models, where demand and mode choice are predicted and modelled based on individual consumer decisions. For these models, the primary data source is often a questionnaire survey or other individual empirical source based on a sample. In this type of model, the role of economic background variables is fundamental, and the use of associated and substitution costs to influence demand is also common. Substitute services are typically integrated to a logical framework to represent consumer choice.

Another main category includes studies that start from a fundamentally aggregate level, using the most basic spatial theoretical framework, gravity models, often complemented by the effects of substitutes. Here, the role of proxies is not necessarily essential, but economic variables, notably population and economic development, are an inevitable part of the models.

Quality issues are much less frequently addressed in this empirical part of the academic literature. This area is less at the forefront of interest, partly because it is traditionally seen as a purely business issue, which is less relevant in development policy and investment-focused research. The business literature, on the other hand, deals with these types of issues, which are typically more easily and deeply addressed by qualitative research tools, through other data sources rather than large demand databases. While this type of research is becoming more intense in the business world with the emergence of systems that automatically produce large amounts of data (European Union, 2016), it is still a very recent field, typically also driven by business, and has not yet reached the academic community in terms of research, results and publications.

However, the way in which the major technological innovations of a given period have led to new research directions is striking. For example, in the case of the United Kingdom, the introduction of CAPRI (Computer Analysis of Passenger Revenue Information) in 1986 made available fully electronic relative sales data (ORR, 2018), and has led to more robust measurements since the early 1990s.

Similarly, the emergence and rapid expansion of the Chinese superhighway network has had a significant impact in the literature. In China, the planning of a high-speed rail network began in the early 1990s, inspired mainly by the Japanese Shinkansen system. As a first step, increasing lengths of the railway backbone network were upgraded from the previous system of 100-120 km/h maximum speed to 160 km/h, which was already the highest level of conventional technology by European standards, and then, with further improvements, more and more sections were upgraded to speeds above 200 km/h maximum. At the same time, the construction of a superfast rail network dedicated to passenger traffic on completely new tracks started in 2006 and, with lines opened between 2008 and 2018, the network is now almost 30,000 km long (Barrow, 2018). The development process slowed for a while after the Wenzhou rail accident in 2011 (coonan, 2011), but a few years later the lines were built and delivered at the previous pace. This development is astonishing, with an average of approximately the same length of superhighway as the entire Hungarian core rail network built in China every year. Of course, in addition to the size of the country and its unparalleled economic strength, the legal, environmental and authorisation regime is unimaginably supportive of these projects, enabling investors to resolve in a matter of days issues that could take decades to resolve in the Western world. Another feature of the process is that there are countless cases of malpractice that accompany the developments, few of which make the news, and even fewer of which reach the foreign press.

However, the process is undoubtedly unique in that it is a major subject for analysis, both in terms of demand for rail services and competition with air transport. The combination of the incredible pace at which the network has developed and the period of IT developments that preceded it has led to a wealth of articles and analyses on the subject. This latter dichotomy is also a subject of professional reflection in other parts of the world, whether we are looking at the use of high-speed rail or the other most frequently studied topic, competition between high-speed rail and low-cost airlines.

2.1 Sources of demand and the role of subjective factors

Before presenting the main results on the demand for rail services, its measurement, estimation and analysis, it is worth briefly reviewing the basic analytical framework for transport services, which provides a picture of the context and approach of the economic and business research. It is also important to stress that some of this literature is not accessible or visible, as studies, strategic planning and analyses within large companies are often not published or are extremely difficult to research, whereas it is the most relevant in this context[5]. However, sources are still available, and some of the publications on public policy occasionally do cite and summarise relevant details of non-public research (Wardman and Whelan, 2011).

Todd Litman discusses the fundamentals of transport demand in a soft analytical framework, with a consultative approach, in a handbook-like format for public policy purposes in his study on transport demand elasticities (Litman, 2013). In his work, he summarises some of the key considerations about the background to demand and transport planning in a way that is typical of business thinking in the industry. These considerations not based on detailed research, but rather on simple logic and experience, can be useful for an intuitive reflection on the mechanisms of demand.

It identifies the following as key demand drivers:

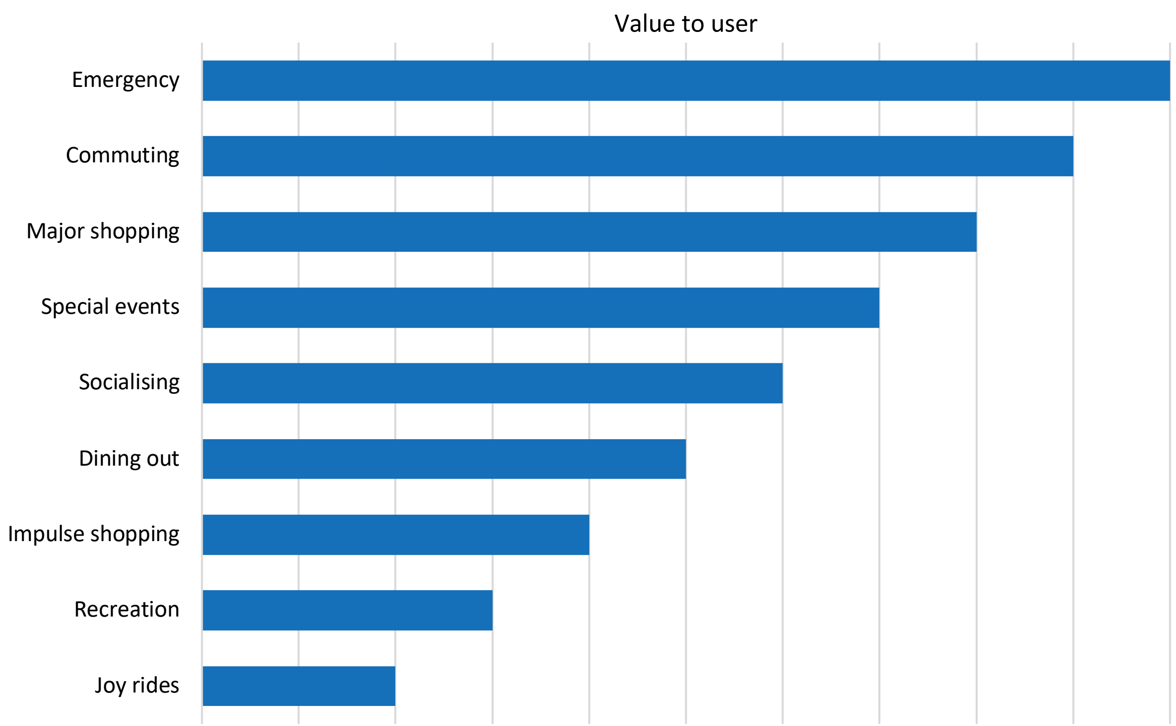
- Demographics and tastes
- Geographic and land use patterns
- Economic activity
- Knowledge of available opportunities
- Quality of available opportunities
- Demand management strategies (in particular active transport policies)
- Costs
- Income

[5] In the case of MÁV, for example, the MÁV Documentation Centre and Library is a source of evidence for researchers to access, but the archive contains very limited material on business processes and decisions, and for the last 2-3 decades it is almost not a suitable source for this kind of research.

In this logic, it is noteworthy that in addition to the aspects that are well covered by models and discussed in detail in the economic literature, such as costs or income, it also includes subjective elements that are rarely focused on but may play an important role, such as taste (individual subjective preferences), the effects of active transport policies (e.g. environmental campaigns in favour of cycling or public transport), the existence, quality and content of knowledge about the available options.

The demand for transport is a classic case of derived demand, since travel is almost exclusively for some purpose, not only for its own sake. In transport policy and in theoretical literature, the motivation behind travel can be an interesting question, as it has a major influence on the behaviour of the people concerned, the elasticity of demand and mode choice decisions. In Figure 4, the author presents such motivational aspects, where he lists some basic types.

Figure 4. Purpose and prioritisation of trips



Source: Author's compilation based on (Litman, 2013)

In Table 2, the content of the factors influencing demand is presented in more detail, which may also help to identify which data sources can be used to measure a particular aspect.

Table 2. Background on the factors shaping transport demand

Demographics	Commercial activity	Transport options	Land use	Demand management	Prices
- Number of people - Employment rate - Wealth/ incomes - Age/ lifecycle - Lifestyles - Preferences	- Number of jobs - Business activity - Freight transport - Tourist activity	- Walking - Cycling - Public transport - Car-sharing - Automobile - Taxi services - Telework - Delivery services	- Density - Mix - Walkability - Connectivity - Public transport service proximity - Roadway design	- Road use prioritisation - Pricing reforms - Parking management - User information - Promotion campaigns	- Fuel prices and taxes - Vehicle taxes & fees - Road tolls - Parking fees - Vehicle insurance - Public transport fares

Source: Author's compilation based on (Litman, 2013)

The above logical framework provides a good opportunity to consider which of the factors and categories of factors influencing the demand for the service might be relevant for the Hungarian rail market. Basically, four logical cases can be distinguished from the above:

- Structural factors that are the basic parameters of the model (e.g. population of settlements)
- Structural factors that are uniformly present throughout the country and therefore not relevant (cultural background, travel habits[6])
- Relevant explanatory variables (quality, service level, prices, etc.)
- Other, irrelevant factors

[6] Differences in cultural and travel patterns vary, of course, by region and by type of settlement, but differences of such subtlety are difficult to define and therefore very limited in their meaningfulness.

In terms of available data sources, Litman's work also provides a good framework to build a comprehensive picture of how much of the relevant factors can be covered by the overall model.

Wardman summarises British research on the problem of valuing time (Wardman, 2001). By reviewing these, we can get an idea of the variability and trends in people's subjective evaluation of time spent on travelling. Table 3 shows the range of studies and measurements that Wardman has summarised for each 'type of time'. It is striking that even relatively identical logical categories are often approached differently, with some factors not necessarily included in all studies. This is mainly because these factors can be assessed more subjectively. This makes comparisons difficult, but also shows that there is no general standard on the subject to which all researchers would adhere.

Table 3. Summary of time assessment studies

Attribute	Studies	Valuations
In-vehicle time	132	539
Walk time	43	142
Access time	19	53
Wait time	13	35
Walk and wait time	20	64
Search time	6	11
Late time	5	18
Adjustment time	13	97
Delay time	7	21
Headway	49	149
Interchange	23	51

Source: Author's compilation based on (Wardman, 2001)

Table 4 summarises the time evaluation results of the above measurements. The values show relative levels, where the observed data are expressed as a proportion of the time spent in the vehicle (time spent in the vehicle is equal to one). It can be clearly seen that the range of time estimates is very wide, for example, waiting for equalisation, to ensure a transfer or to keep tracking time – as necessary 'useful times' – show values below one, all other values are higher. The excess is in the range of 30-50% for most of the factors, but the excess time due to delays is rated by passengers as a particularly extreme multiplier of 7.4, based on the survey data used. The subjective assessment of time spent on transfers has a level even higher, with multipliers above 30. Here, several measurement methods are distinguished. In the case of pure transit, only the time spent in transit is shown, without waiting and other setments. The other two indicators (with slightly different methodologies) give an assessment of the total transfer, where the beginning and end of the period are directly related to the time spent in the vehicle.

Table 4. Subjective time assessment results (relative to in-vehicle time = 1.0)

	Mean	S.D.	S.E.	10%	50%	90%	Obs
Walk time	1.66	0.71	0.06	0.90	1.52	2.67	140
Access time	1.81	0.75	0.10	0.88	1.88	2.70	52
Walk and wait time	1.46	0.79	0.10	0.61	1.31	2.43	63
Wait time	1.47	0.52	0.09	0.94	1.33	2.19	34
Adjustment	0.72	0.64	0.09	0.30	0.50	1.30	56
Headway	0.80	0.46	0.04	0.27	0.70	1.41	145
Search time	1.38	0.52	0.17	0.79	1.22	2.26	10
Late time	7.40	3.86	1.16	1.94	8.00	14.00	11
Delay time	1.48	0.32	0.07	1.04	1.43	2.01	21
Interchange (pure)	17.61	10.93	4.13	3.91	13.52	31.70	8
Interchange (full v1)	33.08	22.73	4.64	10.60	28.41	70.47	23
Interchange (full v2)	34.59	25.88	6.46	9.50	27.53	66.70	16

Source: Author's compilation based on (Wardman, 2001)

2.2 Modelling transport demand

The wide-ranging literature on the demand for rail and transport services in general has been summarised in a number of reviews over the past decades. The two works presented below provide an excellent overview of the subject.

Kenneth Button presents results of a 30 years period in his synthesis (Button, 2006), which ranges from the topics of the late 1960s, in particular the problem of optimal pricing, to the computer data-driven analytical systems of today, and takes into account the specific context provided by the EU single market. In Button's view, until the 1960s, the descriptive and institutional perspective on transport economics was essentially dominant. It was really from the 1970s onwards that the deeper analysis of the issues began, with the emergence of new models and, together with impact assessments, higher level public policy debates in transport. The paper summarises the main points of the deregulation wave that has been gaining momentum since the late 1970s, the debates and concerns about the changing market structure, the role of the institutions involved and the regulatory toolkit. This period also saw the emergence of a more detailed analysis of bottlenecks, the literature on the economics of congestion management, and the concepts behind the emergence of congestion charging models.

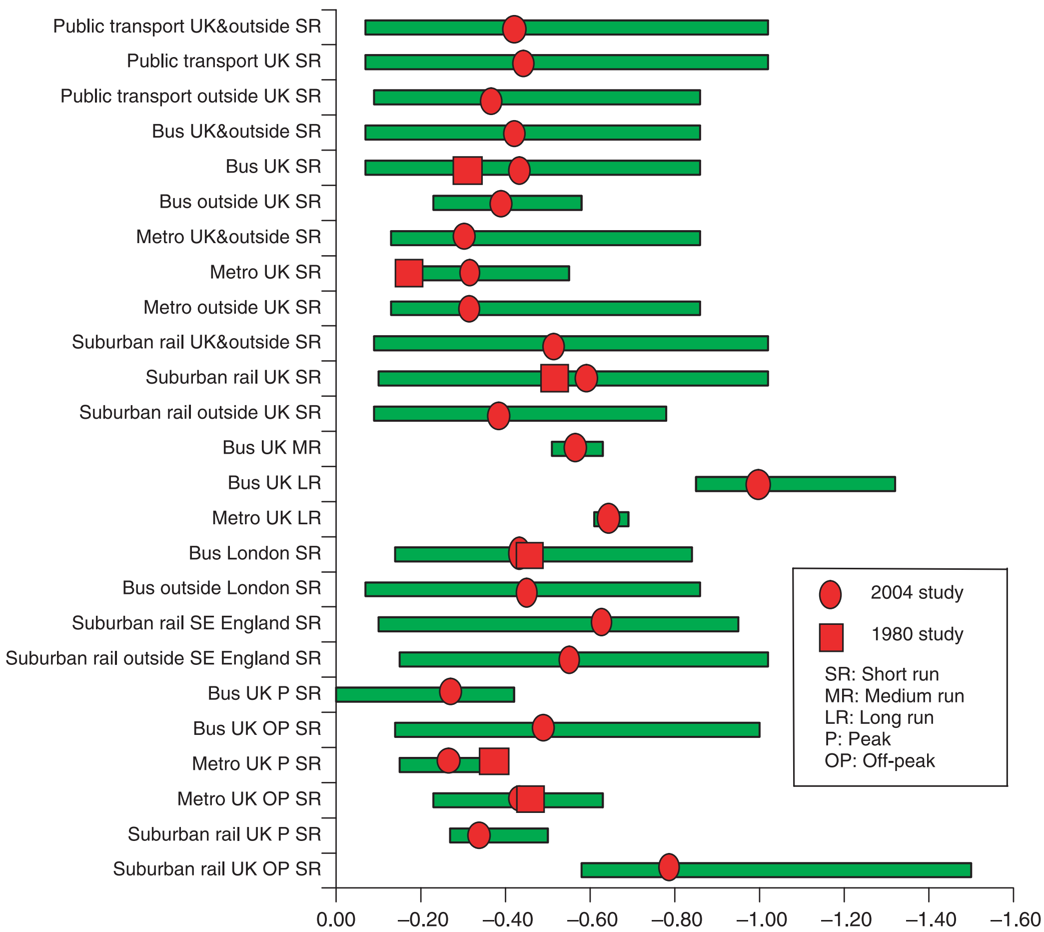
Also important in Button's summary of the evolving literature of the period is the development of transport modelling, including demand models. Earlier analytical practice was constrained both by the range of data available and by limited computational capacity, and with gradually improving capabilities, increasingly complex models were developed for planning transport projects. Aggregate models have been followed by analyses based on household and individual data, i.e. transport behaviour defined at the micro level, where the limitations are usually not the tools but rather the data sources. At the same time, the results of the dynamically developing experimental economics started to percolate into transport studies, models based on stated preference, and other inputs through various data collection methods were increasingly incorporated.

At the end of the review, Button concludes by expressing his expectation that more research on this topic should focus in the coming period on a more detailed understanding of logistics supply chains. Accordingly, a deeper analysis of the supply-side quality characteristics of the Hungarian rail market could also lead to new results in this direction.

In 2006, Neil Paulley and co-authors published the results of a joint study on the impact of fares, service quality, income and car ownership on transport demand (Paulley *et al.*, 2006). The aim of this work was to test the findings of an influential book on methodology published in 1980 (Webster, F.V., Bly, 1980) by building on several data sources, collected from earlier research. The data largely relate to local and suburban travel, so they focus on shorter distance trips than the trips I examined.

The results on price elasticities are shown in Figure 5: the base data is from a 2004 survey in different areas, by mode and time. The green bars show the range defined by the minimum and maximum values, with the average marked. Distinguishing between the earlier data from 1980, it is striking that there is quite a large variation between the different measured elasticities, while the results of measurements on the same subject are often close despite the 24-year time span.

Figure 5. Summary of results on price elasticity of demand



Source: (Paulley *et al.*, 2006)

For analyses of quality factors, a larger number of results from stated preference models are available. The evaluation of factors is expressed in terms of the proportion of time spent in the vehicle, the results are presented in Table 5. Some of the data are presented as demand elasticities, where quality factors were expressed as costs. For low-floor buses, the ratio of benefits to conventional buses is included, in the case of rail vehicle upgrade, the benefits were expressed as a ratio of fares for air-conditioned vehicles. The results are mixed, showing a rather varied and subjective picture in terms of the proportions according to how respondents value a particular amenity. Due to the specificity of the stated preference models, empirical methodological differences may also appear, which can only be intuitively assumed on the basis of the results.

Table 5. Assessment of factors describing the service quality

	Min.	Max.	Obs.	Unit
Access time (walking)	1.4x	2.0x	183	min / min in vehicle time
Access time (all modes)	1.3x	2.1x	53	min / min in vehicle time
Waiting time (bus), short run	– 0.38		27	elasticity
Waiting time (bus), long run	– 0.66		27	elasticity
Waiting time (rail)	– 0.75		3	elasticity
In-vehicle time (general)	– 0.4	– 0.6	3	elasticity
In-vehicle time (regional rail)	– 0.4	– 0.9	5	elasticity
Generalised cost (fare, in-vehicle time, walk and wait time)				
Bus	– 0.4	– 1.7	1	elasticity
London underground	– 0.4	– 1.85	1	elasticity
Rail	– 0.6	– 2.0	1	elasticity
Low-floor bus	5p	14p	1	utility (£0.1)
New rail rolling stock	0.01	0.02	1	min / min in-vehicle time
New rail rolling stock (air conditioned)	2.5%		1	% of the fare
Interchange	21	37	16	min / min in-vehicle time

Source: Author's compilation based on (Paulley *et al.*, 2006)

Pulley and co-authors have also investigated cross-modal elasticities. The results for long-distance traffic are summarised in Table 6, which clearly shows that substitution is present in all relations, both in terms of time and cost. In the case of rail, there is a strong asymmetry towards the car, worryingly for public policy purposes, with a much higher proportion of travellers switching from car to public transport more easily than vice versa. One of the reasons for this seemingly positive data might be the time span of the study and the availability of assets. If rail is used primarily by people without cars, switching is not an option for them in the short term. For car users, however, it is easy to switch whenever the need arises. At first sight, this result may also be surprising, given the well-known 'stickiness' of car use, which is also reflected in the other data, one would expect the greatest inflexibility to be in the case of switching from car to rail. In Hungary, for example, the premise that car users find it difficult to switch to public transport modes is quite persistent in public policy debates.

It is also interesting to note that the elasticities between rail and bus are lower for switching from bus to rail than if the two modes of public transport were separate castes. This result may also highlight the importance of subjective factors as mentioned earlier, but presumably the feasibility of substitution plays a role here, as rail and bus services typically do not overlap completely (in many regions there is no parallel rail and bus offer at all). Such details cannot be deciphered from the summary study.

Table 6. Cross-price elasticities in long-distance transport

	Car use	Rail use	Coach use
Car time	–	0.33	0.60
Car cost	–	0.25	0.34
Rail time	0.057	–	0.20
Rail cost	0.066	–	0.32
Coach time	0.054	0.17	–
Coach cost	0.014	0.17	–

Source: Author’s compilation based on (Paulley *et al.*, 2006)

Among the more recent results on the effect of car ownership and income on transport demand, a general trend is the negative income elasticity and the negative effect of the number of cars. Of course, the two are related, as the first factor has a strong impact on the second. However, this is a very complex problem, which can be reversed over time and as the number of cars increases and the availability of physical space becomes saturated. In principle, it is reasonable to assume that, as average income rises, families and individuals without cars will also buy cars and thus increasingly move away from public transport markets. At the same time, the more actors go through this process, the more congested the situation will become, and in some cases – typically in suburban and metropolitan regions – public transport will be preferable to private cars. Thus, economic development will mean that the use of public infrastructure, especially where access is not properly priced, will increasingly bear the hallmarks of a tragedy of the commons situation, and so may be reversed, with the rise of public transport again. It is particularly difficult to measure this through aggregated data, since in some cases opposite processes might take place simultaneously: some actors entering the market for private transport while others, even with maintaining their car ownership, are in the process of returning to public transport.

It is also important to note that these choices are not universal, with a growing proportion of users mixing modes and making daily choices based on destination and other circumstances. (This kind of flexibility is particularly prevalent in urban transport.) This is what MaaS conceptualises and builds on, and is becoming an increasingly important approach to thinking about urban mobility: this logic interprets travellers as general, multimodal users, making mode choices from day to day, trip to trip. While these developments have gained considerable momentum in the last decade, their effects are not yet to be found in the earlier logical measurements and studies, and so this concept is not prominent in the summary communication by Paulley and his co-authors.

Gines de Rus, in his 1990 measurement (de Rus, 1990), examines the elasticities of demand for public transport services in Spain. The initial data are sales data for transport systems, and data for the largest cities – Madrid and Barcelona – were not included in the database, as they already had highly complex transport systems and associated complex pricing systems that were difficult to model. In the baseline estimation, the dependent variables were the monthly number of single-trip trip journeys, the explanatory variables were the price of a single-trip ticket, the price of a multi-trip ticket, the monthly vehicle-km performance, a time trend, and two dummy variables (for months and for major system outages). For the transport modes in the database, estimates of the own-price elasticity and, where meaningful, the cross-price elasticity were performed by de Rus. One striking feature of the results obtained is that much higher values (up to five times) than the own-price elasticities were obtained for the level of service type elasticities. In this case, the level of service describes the quantitative part of the supply, i.e. the mileage in vehicle-km. However, the causality is not sufficiently clear, as the higher performance may be due to an exogenously high level of demand, so in this case it is not the effect of higher frequency of service that leads to higher demand, but rather higher supply (higher vehicle mileage) due to higher demand.

In 2006, Wardman examined the evolution of demand and the impact of factors affecting it (Wardman, 2006). As a starting point, he presents the forecasting method typically used in the UK during that period, as described in the Passenger Demand Forecasting Handbook (PDFH) published by British Railways in 1989. This methodology derives changes in demand solely from GDP growth. Whereas this model was adequate for forecasting between 1970 and 1990, with a basically unchanged rail network, it no longer worked in the following period. According to Wardman, this was mainly due to external factors that caused demand to grow much more dynamically overall than before:

- GDP grew at a much faster rate than in previous periods
- the size of the main road network stopped growing
- the cost of car ownership has increased significantly

All these factors combined had a strong impact that rendered the previous model inadequate, with real growth figures being several times higher than the values of the results in the previous projections.

For this reason, Wardman used an extended model to investigate which approach could have provided a more accurate forecast. The gravity-based model includes, in addition to population and GDP data, car ownership and rail service quality data (travel time, frequency, transit). The model is based on the CAPRI database of relational data, to which additional explanatory variables have been added. However, due to the large number of background variables in the augmented model, the problem of collinearity appeared, with strong correlation between the background variables. The measurement results basically confirmed the hypothesis of bias due to omitted variables, but the measurement model cannot be considered ideal, as it raised many additional problems.

In 2005, Andrés López-Pita and Francesc Robusté studied the Madrid-Barcelona high-speed rail line, then under construction, from the point of view of its potential impact on air traffic between the two cities (Andrés López-Pita, 2005). This route, which is also a major European route, carried more than 4 million passengers a year in 2003, with more than 60 flights a day. The study describes in detail the evolution of the route, current offers, operators, categories and average fares. In examining the likely effects, the authors show that the Madrid-Barcelona route is expected to fall just within the sensitive range for the changeover.

In forecasting, individual preferences are determined in a price-time model, where individual optimisation is represented as the individual valuing his/her own time and, depending on this, taking a shorter travel time solution in exchange for a premium, thus increasing his/her own free time. The total travel cost determined by the time valuation is incorporated into a gravity model. The results of the analysis show an expected share of 63.5%, compared to Iberia's previous forecast of 52.5%. The forecast does not include a time dimension. More than 10 years later, the share of rail, which opened in 2008, was measured at 63% of the market by 2018 compared to air (Global Railway Review, 2018), which was mainly explained by more favourable door-to-door travel times, among several factors (Bode *et al.*, 2018). Although of course further trends cannot be predicted and the authors did not define a time horizon for the prediction, the result can definitely increase confidence in the usefulness of these type of gravity models.

Wardman and co-authors also address the problem of forecasting in their 2007 paper, reviewing cross-sectional models (Wardman, Lythgoe and Whelan, 2007). The basic model used here is again the gravity model, where the cost of a given train is a proxy for quality and service level. It raises the shortcomings of the time dimension in the context of cross-sectional studies, while highlighting the following potential advantages of such models:

- the possibility of forecasting demand for stations (i.e. the model does not only apply to aggregate demand for a single line or region, but to specific origins and destinations)
- the elasticities of access, interruption and continued travel allow the effectiveness of projects to improve the accessibility of the rail network to be assessed

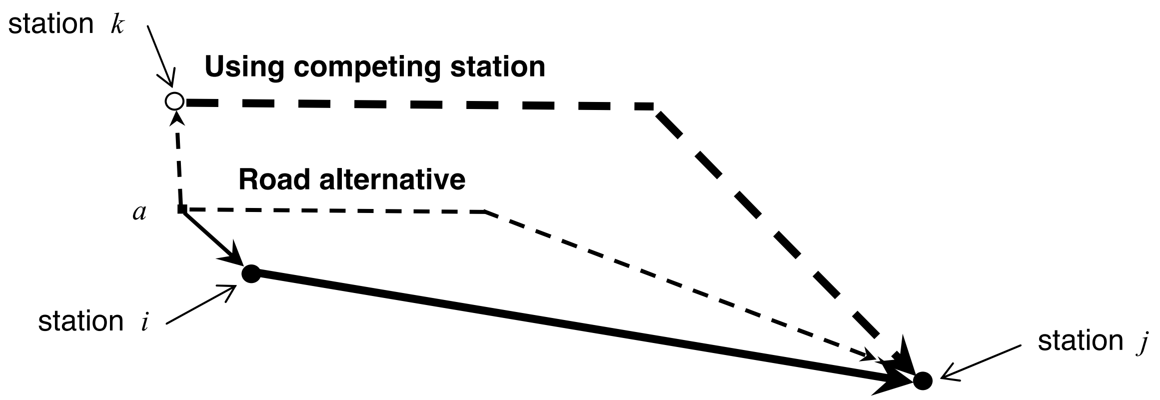
- to assess competition between stations – a comparative analysis of the passenger's choice of stations, which differs from the expected one, will allow the evaluation of improvements; the relationship between different services and solutions can be explored by means of passenger flows which differ from the expected trend
- cross-sectional origin-destination sales databases are also a good way of measuring demand responses to the quality of rail services
- it allows a wide range of socio-economic background data to be linked on a territorial basis and possible effects on earnings to be examined
- generally allow the impact of factors that vary on a spatial basis to be estimated, and are more stable over time

These last three aspects are particularly noteworthy for my study, as I aim to make a similar estimation on the basis of Hungarian data. The issue of quality analysis is relevant, but the examination of background data that can be tested on a spatial basis is also relevant because of the possibility of linking them to relevant spatial statistics. Here, however, the collinearity problem already encountered in the previous Wardman study in 2006 is a strong limitation, i.e. the selection and inclusion of a single more strongly influencing factor may be justified, but the combination of too many explanatory variables moving together may cause more problems.

The approximation data can be further explored using a spatial refinement model. In Wardman's example, the accessibility of a station is broken down into smaller zones, allowing the costs of proximity (transport, information search, uncertainty, local transport costs, parking, etc.) to be measured, and even the preliminary impact assessment of intervention plans to improve access to a station.

In another approach, this problem can also be studied through the choice between stations. In this case, the most interesting areas the ‘border zones’ between two or more stations, and the difference between such indifferent zones and other locations is an important issue. In the case where the actual boundaries of two stations’ access areas deviate from the predicted one, it can be assumed, after taking into account all relevant access costs, that there is a property in the services or other characteristics of the stations that causes this deviation. A spatial schematic drawing of this logical situation is shown in Figure 6, where station *i* would be the shortest, faster route between the origin and destination *j*, but where other characteristics of the real options justify it, station k may be a realistic option, as it does not involve excessive detours or time loss. In identifying these types of situations, it may be appropriate to examine the two stations and their differences, as this will reveal the subjective factors and effects that determine demand.

Figure 6. Routing scheme between competing stations



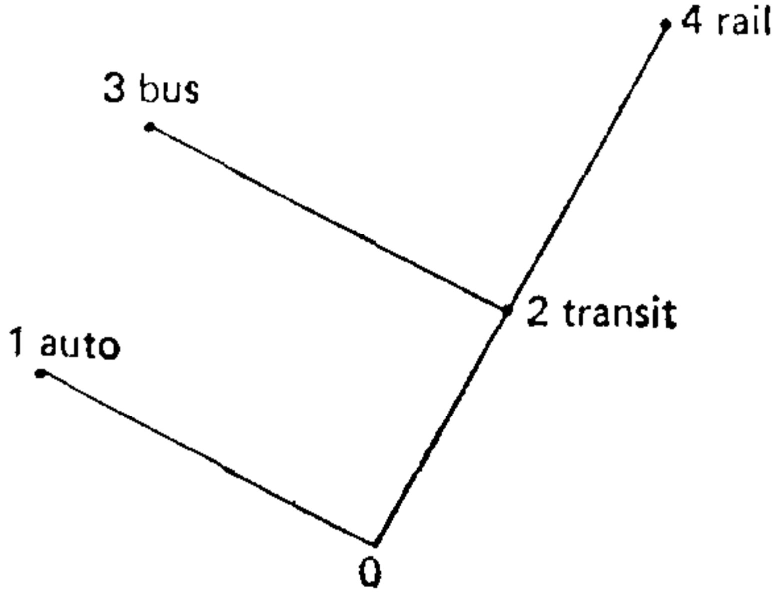
Source: (Wardman, Lythgoe and Whelan, 2007)

2.3 Modal choice-focused models

The choice between transport modes is one of the large slices of the demand literature for transport services. For the purposes of my study, this area is of particular relevance because of the aspects of choice between factors, substitution, and the tangibility of quality issues, which cannot otherwise be directly translated into a relational analysis in the absence of specific data on mode choice.

Daniel McFadden focuses on urban transport demand in his 1974 study (McFadden, 1974), where he constructs demand from individual transport preferences. As can be seen in Figure 7, he discusses mode choice decisions in a sequential approach, where first the choice between individual and community mode is made, with choice between bus and rail as additional alternatives made later.

Figure 7. Mode-selection decision process



Source: (McFadden, 1974)

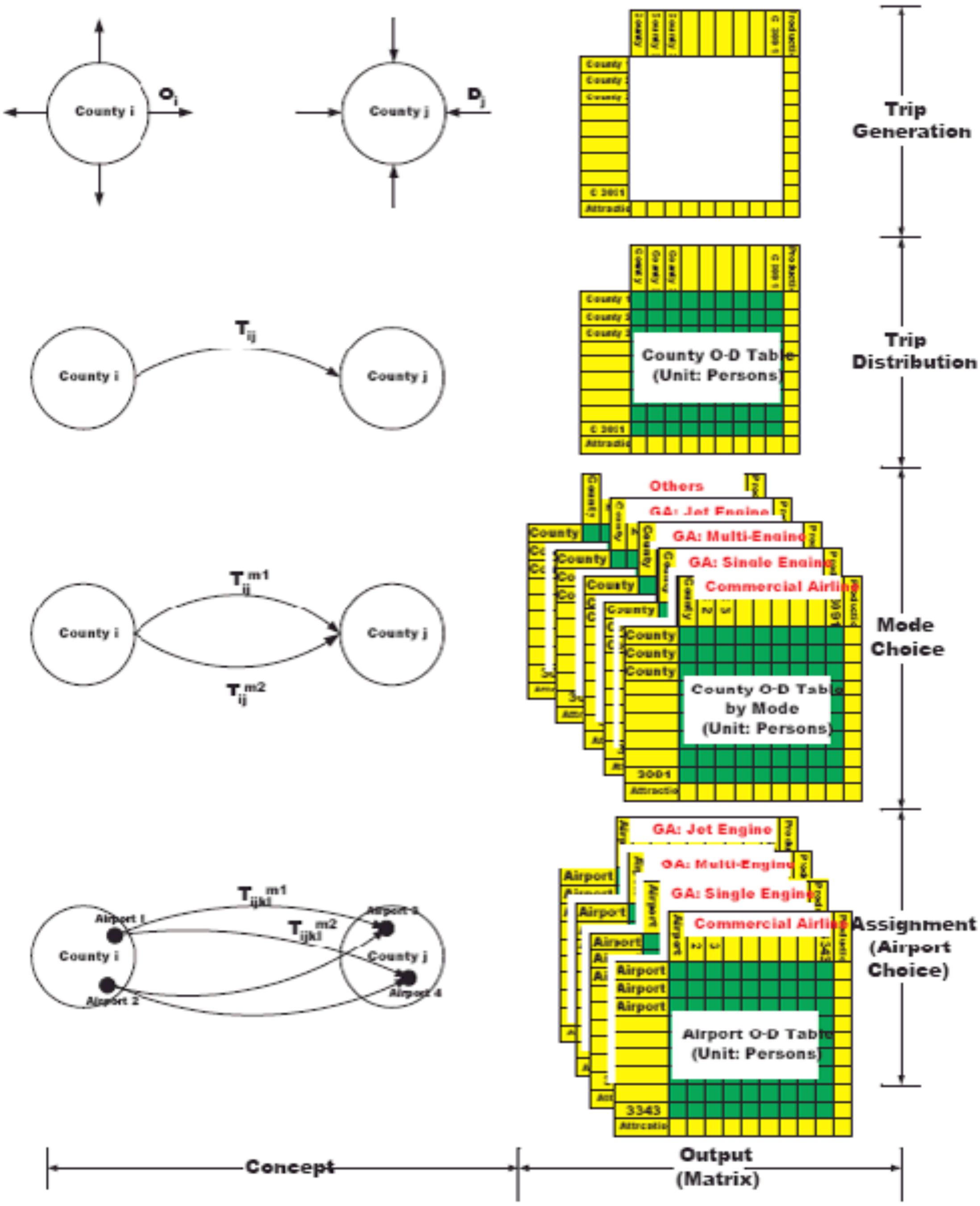
Ahern and Tapley carried out a questionnaire-based analysis in 2008 (Ahern and Tapley, 2008), where they measured preferences for rail and road public transport for a model. Questionnaires were taken on two routes (Dublin-Sligo, Dublin-Galway) and analysed in two separate sections, but there was also an explored preference section. Two types of methodology were used, one version was a ranking questionnaire and the other a stated choice questionnaire, the latter being simpler and the former a more plentiful source of data. These were tested on pilots with 20 participants. The total recruitment reached less than 200 respondents. The explanatory variables used in the study were cost, trip length, reliability (delay minutes), time spent out of vehicle, access cost, mode choice. This questionnaire is a good example of how a non-aggregated approach can be implemented with a small amount of data and a small number of items, and can already produce valuable data. However, it is of course very limited in several dimensions because of the volume: it is difficult to generalise the results spatially and it is easy to find weak points of the research.

In a paper published in 2007, Ashiabor and co-authors present the results of a research project that is in many ways unique (Ashiabor, Baik and Trani, 2007). One of the special values of this US national long-distance travel survey is that this scale is relatively rare in the literature. The study and forecasting of transport demand is a very frequently researched public policy issue, but is usually undertaken at the state level. Here, the client was the US space agency NASA: in the context of a special project, which still seems futuristic today, the vision of a transport system based on small aircraft was examined in detail, a somewhat utopian concept that has been around for at least 50 years, and whose reality is still not really obvious today (NASA, 2001).

For the task of estimating demand at national level, the data from the systematic studies carried out at national level mentioned above provided a good basis. The National Travel Survey system records from 1976 onwards, in several versions, up to 1990, in total four data collection periods, with 2-4000 records each. In addition, the new American Travel Survey database from 1995 onwards, a much more extensive source of some 400,000 records, was used.

The logical framework of the database built for the modelling is shown in Figure 8. The process for building the complete model was carried out in four basic steps. First, relatively narrow geographical territorial units were defined, with the then total of 3091 counties in the fifty Member States forming the basic unit and a trip generation dataset was created for each basic unit. Destinations were then assigned to the needs, which were used to set up origin-destination pairs between counties, so that the generated trips were distributed spatially (trip distribution). The third logical step was mode choice, where a trip was assigned to a mode of transport (mode choice). In a fourth step, the resulting air transport demand was assigned to specific airports within the counties (assignment – airport choice).

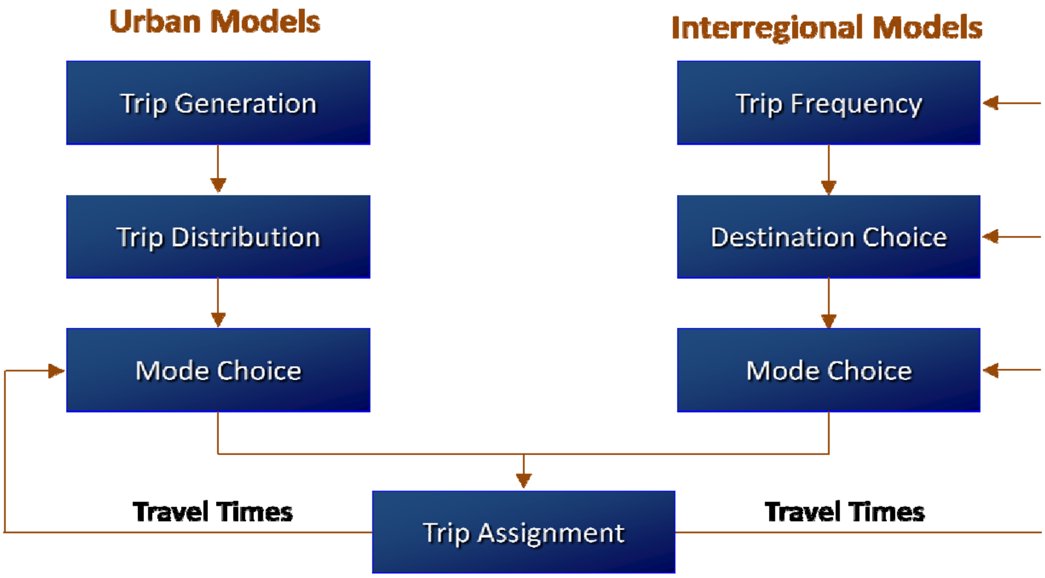
Figure 8. Logic framework of the model for the US long-distance market



Source: (Ashiabor, Baik and Trani, 2007)

A good example of a more complex model is the 2010 study by Outwater and co-authors, where they assessed the potential of high-speed rail in California for a planning commission (Outwater *et al.*, 2010). In the complex model, they distinguish between the logic of modelling urban and interurban traffic. As shown in Figure 9, in the urban model, trips are first generated for the model, then they are split and the resulting trips are split by mode. In contrast, for long-distance models, the initial data is trip frequency, followed by destination choice, and the third logical step is mode choice.

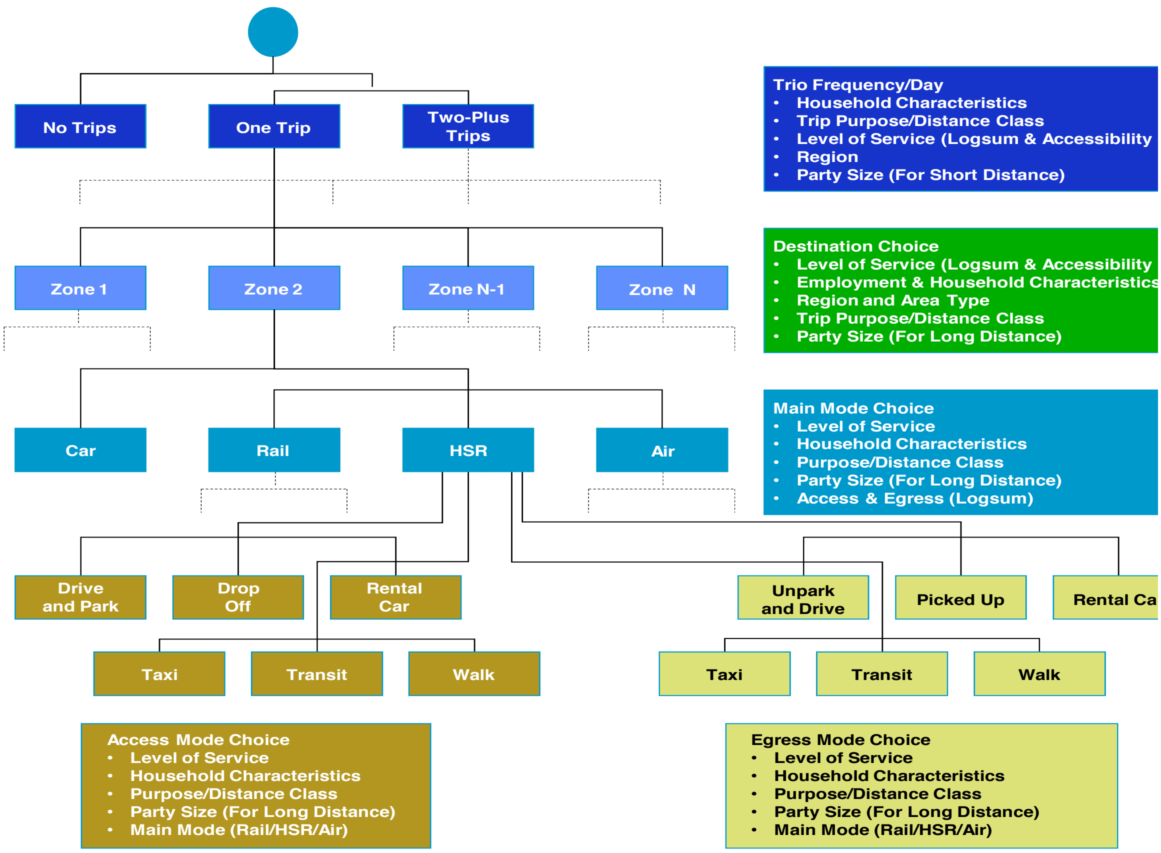
Figure 9. Integrated modelling concept in high-speed rail forecasting



Source: (Outwater *et al.*, 2010)

The overall logical framework of the resulting model is summarised in Figure 10, where its complexity is clearly shown. This analytical framework may at first sight seem overly complex, but describing consumer decision making in today's complex transport chains could not be simpler, as the sequences of interdependent, multi-stage choices can create much more complex situations.

Figure 10. Modelling framework for high-speed rail forecasting in California



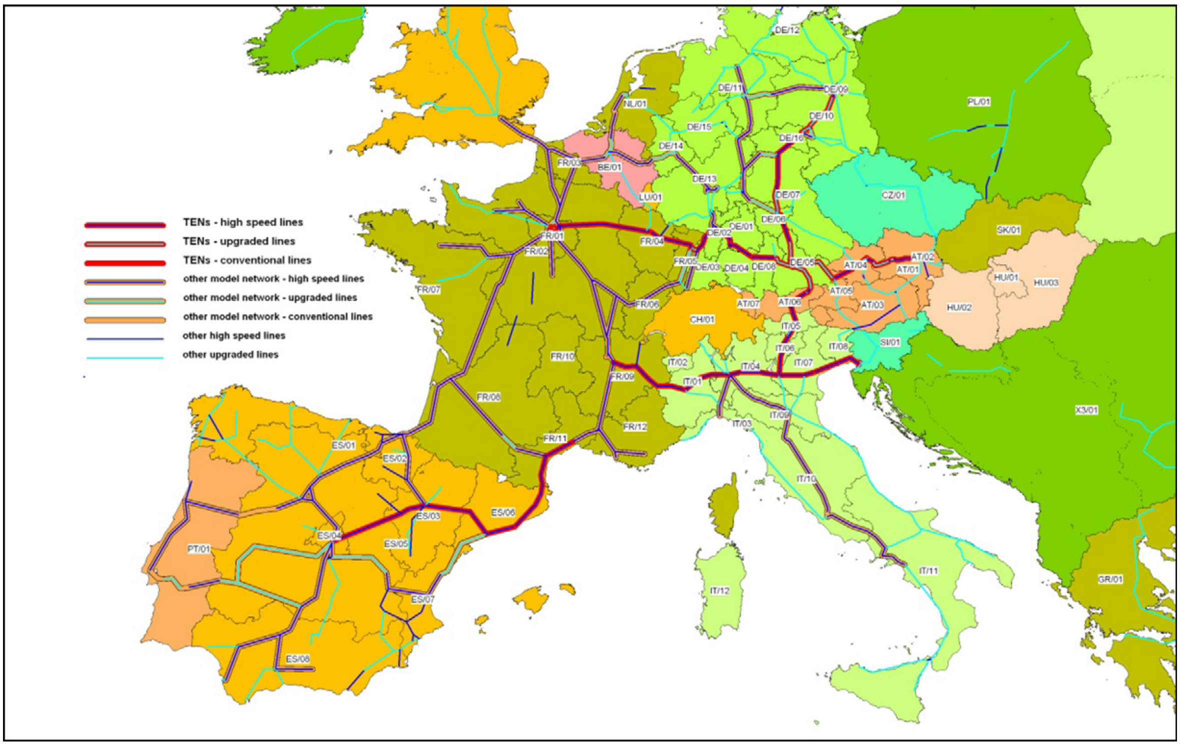
Source: (Outwater *et al.*, 2010)

With today's information technology and the rapid increase in the amount of data available, the possibilities for modelling and forecasting have therefore expanded considerably. At the same time, complex tools for investigating a problem or research question can also carry the risk that it is even more difficult to maintain focus, that different uncertainties can add up due to inconsistent data sources and other potential biases, and a complex model can paradoxically produce even more biased results than an incomplete but simpler study.

In 2010, Adler and co-authors examined the effects of high-speed rail on Europe, focusing on competition between rail and air transport (Adler, Pels and Nash, 2010). Their approach was simplified, they simulated high-speed rail and air networks and markets to form a complete model. Decision variables included prices, capacity and frequency of flights for origin-destination pairs. The supply cost functions of the model reflect industry specificities, distinguishing between conventional and low-cost airlines in the air transport sector, which have more freedom and a larger playing field, but operate at higher cost levels. The rail cost function only takes into account high-speed operators, in line with the characteristics of the model.

Demand zones were defined in the network modelling, over which the simplified air and rail network was defined. For aviation, based on real data, Paris, London and Frankfurt were added as international exits, with three regional hub airports: Prague, Budapest, Warsaw, and London and Berlin as separate low-cost hubs. The rail network, as shown in Figure 11, is based on the current plans for 2020, as of 2010, where only the planned Polish sections in the Eastern region are included.

Figure 11. The planned state of the European high-speed TEN rail network by 2020



Source: (Adler, Pels and Nash, 2010)

The model integrated all known development plans in 2010 according to the expected commissioning dates based on the current schedule. For the market structure and pricing simulations, relatively complex transformation processes were also captured in the results of the runs due to the game theory-based approach. The pricing model for infrastructure and usage, the relationship between the two and the incorporation of environmental costs into prices were also included as parameters in the model.

Finally, the two following publications are characterised by the choice of specific, complex tools to solve a particular problem using existing methodologies. While the first one used a more complex solution to forecast passenger traffic on a planned railway line, thus integrating timetable planning into the process, the second one uses an unconventional data source, i.e. social media logins.

Chang and co-authors studied Taiwan's high-speed rail project, carrying out modelling for the planning of a non-branching high-speed rail line in 1998 (Chang, Yeh and Shen, 2000). The railway line opened in 2007 (Smart Rail World, 2017). The focus of the model was on capacity and timetable planning with fuzzy logic program estimation. The model was based on multidimensional input parameters to determine the optimal schedule according to different references: e.g. the frequency of stops allowed and the number of stations, or output matrices according to multi-stage logic. This methodology is well suited for planning a complex service, where forecast, travel needs, demand information and geographic data serve as inputs to the modelling.

In 2014, Liu and co-authors investigated the question of whether social media logins are suitable for calibrating gravity models and what other patterns can be explained by these data sources in spatial interaction modelling (Liu *et al.*, 2014). The basic gravity model was applied to the estimation in the form corresponding to the original physical gravity equation:

$$I_{ij} = \frac{kP_iP_j}{f(d_{ij})},$$

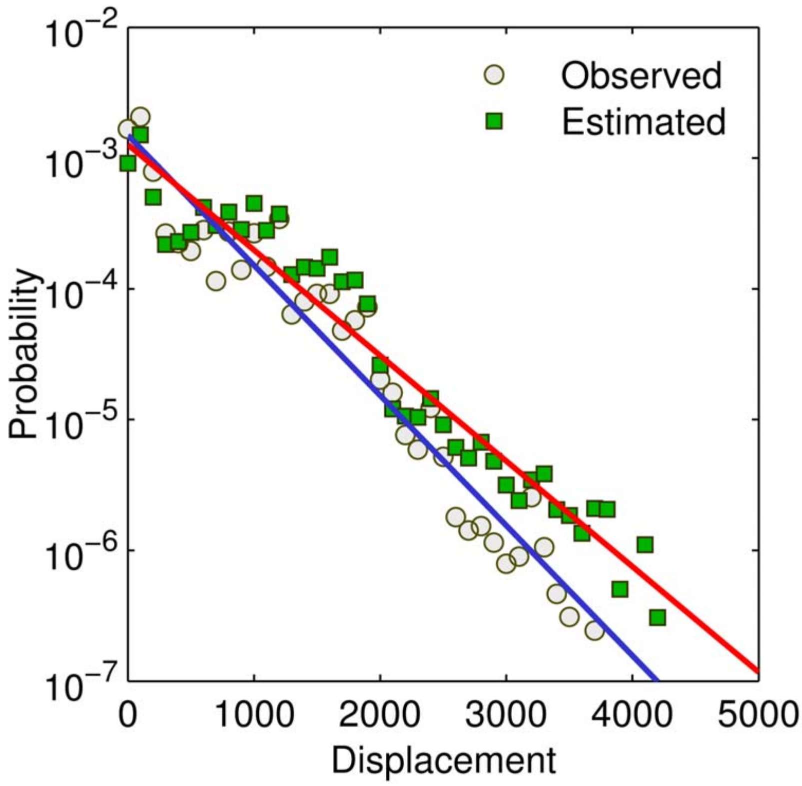
where:

I_{ij}	the interaction from *i* to *j*
P_i, P_j	attraction of place *i* and *j*, respectively
d_{ij}	distance between *i* and *j*

The authors' question was how this model can be calibrated in transport, what are the specificities of this model. The study used a database of check-in data of more than half a million Chinese travellers, including all such data for one year, from an unnamed location-based community service, similar to Foursquare, specifically designed for the Chinese market, where customers can check-in to a location and send short messages and pictures to let their friends know where they are going or what they are doing.

For the comparison with gravity models, the starting point was the check-in data for each city in the database and the movements between a pair of cities (assuming that, when changing location, the consecutive check-in data of a specific user represents the movement between two cities). Comparing the resulting estimate with traffic data, the methodology gave a reasonably good approximation, as shown in Figure 12. The probability of error is exponential as a function of distance, with the authors finding a higher intensity section at a distance of about 1200 km, the reason being that the distance between neighbouring Chinese cities typically falls within this range.

Figure 12. Fit of the check-in based model



Source: (Liu *et al.*, 2014)

2.4 Aggregated demand models

In addition to models built from individual mode choice, the other main logical category is aggregate demand models, which are typically built based on gravity spatial logic and estimates from aggregate datasets.

Jones and Nichols (Jones and Nichols, 1983) investigated the demand for long distance rail travel in the UK. It is a fairly early work, based on ticket sales data from 1970-76 recorded by British Railways for 17 routes from London. The full demand model:

$$Q_t = \beta_0 P_t^{\beta 1} JT_t^{\beta 3} EA_t^{\beta 4} GDP_t^{\beta 5} PP_t^{\beta 6} e^{\left(\beta 7 S_t + \sum_{i=8}^{i=20} \beta_i D_{ii}\right)} \varepsilon_t$$

where:

JT_t	journey time by rail
EA_t	index of cyclical activity
GDP_t	index of real GDP
PP_t	index of petrol prices
S_t	service level of non-rail modes
D_{it}	seasonal dummy variable
ε_t	random error

Owen and Phillips also examined similar research questions on British Rail data (Owen and Phillips, 1987). They used the following model:

$$J_t = \beta_0 J_{t-1}^{\beta_1} G_t^{\beta_2} F_t^{\beta_3} exp\left(\beta_4 T_t + \beta_5 S_t + \beta_6 C_t + \beta_7 A_t + \sum_{i=8}^{i=19} \beta_i D_{it}\right) \varepsilon_t$$

where:

J_t	number of single journeys between two stations
G_t	index of gross domestic product at 1980 factor cost
F_t	average revenue per journey (i.e. average fare)
T_t	linear time trend (i.e. $T_t = t$)
S_t	HST (high speed train) variable

C_t coach competition 0–1 dummy variable

A_t air shuttle 0–1 dummy variable

D_{it} seasonable 0–1 dummy variable (12 in all)

ε_t random error

In 1994, Wardman examined the impact of service quality and made a prediction based on it (Wardman, 1994). The analysis focused on the impact of quality on rail demand. The data used were UK long-distance inter-city travel data for non-London areas, sourced from the CAPRI dataset cited earlier, and covering the period 1985 to 1992. To define quality, the British Railways generalised travel time approach was followed. Its calculation and the demand function were included in the model in the following form:

$$GT = T + \alpha_1 F + \alpha_2 I$$
$$V = GT^{\beta}$$

where:

GT generalised journey time

T journey time (station to station)

F frequency (service headway)

I number of interchanges

V volume of travel

Long-distance rail models were examined by Cohen and co-authors using data from a 1975 survey (Cohen, Erlbaum and Hartgen, 1978). The study and prediction was for the New York City-Buffalo route, and the database with combinations included a total of 31 city pairs with quality criteria.

Quality of rail service:

- snack car availability
- sleeper car availability
- lounge car availability
- baggage service

– package express
– on-time performance
– schedule match
– dining car availability,
– car type

Terminal quality:

– parking availability
– number of spaces
– parking fee
– parking lot lighting
– terminal snack bar
– local transportation
– distance to downtown
– modernness of terminal

Brand and co-authors conducted a demand analysis for high-speed rail forecasting in 1992 for the United States and examined the effects of building such a transportation system (Brand *et al.*, 1992). Three estimation methods were used: in the first two, total trip volumes were predicted and then split by mode (with two different estimator functions), in the third, mode was estimated and within each category a separate prediction of the impact of a possible high-speed rail network was made, so that the sum of these partial shares was the share of the new technology.

The basic model was as follows:

$$T_{OD}^{m} = f(P_{OD}, I_{OD}, LOS_{OD})$$

where:

T_{OD}^{m}	number of trips by mode *m* made between *O* and *D*
P_{OD}	population levels in *O* and *D*
I_{OD}	income of travellers between *O* and *D*
LOS_{OD}	level of service on existing modes between *O* and *D*

Thakuriah et al. based their demand estimates on data from the 1995 American Travel Survey (also used by (Ashiabor, Baik and Trani, 2007)), cited earlier (Thakuriah *et al.*, 2010). They used data sorted by routes, so that a single data series contained the available records in the database by a route defined by origin and destination.

The gravity model was described as follows:

$$E(N_{ij}) = T_{ij} = A_i B_j F(c_{ij}) \qquad i \in I, j \in J$$

where:

T_{ij}	expected flow between zone *i* and *j*
A_i, B_j	origin and destination functions, respectively
$F(c_{ij})$	set of costs (such as travel time, distance), which separates *i* from *j*

Martín and Nombela studied the demand effects of the results of investment in high-speed rail in Spain (Martín and Nombela, 2007). The data source for the analysis was the Movilia survey, a questionnaire survey of travel behaviour carried out in 2000. The analysis included trips over 100 km, 187 trips were included in the database. The variables used and how they were defined are summarised in Table 7.

Table 7. Variables of the Spanish high-speed rail model

Variable	Definition	Source
Total number of trips	Estimates obtained from total number of sample trips (subject to sampling errors)	Movilia survey
Travel time	Time spent on vehicles (planes, trains, bus or cars), without including boarding or waiting times.	Service companies/ viaMichelin
Cost	One-way tourist regular fare. For cars 0.20 €/ km, considering only variable costs.	Service companies
Time-interval	Average time between two departures (scheduled public transport). For cars: a value of zero is used (perfect availability).	Service companies
Distance	Measured from geographic coordinates of origins and destinations of trips done by sampled individuals (straight lines).	Movilia survey
Income	GDP per capita and per province (2000)	Statistical National Institute (INE)
Work trips	Percentage of trips done for working purposes, as declared by individuals.	Movilia survey
Population	Population per province	Statistical National Institute (INE)
Capital stock	Value of all transport infrastructure capital, measured by permanent inventory methodology.	Instituto Valenciano Investigaciones Economicas (IVIE)

Source: Author's compilation based on (Martín and Nombela, 2007)

They used a gravity model, with the following general logic:

$$flow_{ij} = f(activity\ in\ i,j; social\ and\ economic\ charact. i,j; interaction\ i,j)$$

Where *i* and *j* represent the two endpoints, the sources of activity data are GDP, population and income. The interaction coefficient is measured by distance. The estimation results are presented in Table 8 and the measured elasticities in Table 9.

Table 8. Results of the Spanish high-speed rail gravity model

Variable	Coef.	St. E.
Constant	2.14408	0.988133
log_distance	−1.47307	0.120261
log_pop_destination	0.393624	0.153476
log_pop_origin	0.807308	0.157858
log_Kstock_destination	0.241356	0.194177
log_Kstock_origin	0.0796245	0.191313

Source: Author's compilation based on (Martín and Nombela, 2007)

Table 9. Elasticities of the Spanish high-speed rail gravity model

Variable	Time-Elasticities				Cost-Elasticities			
	t_plane	t_train	t_bus	t_car	C_plane	C_train	C_bus	C_car
Plane	– 0.498	0.474	0.320	1.477	– 2.561	0.113	0.054	0.942
Train	0.090	– 2.527	0.320	1.477	0.463	– 0.604	0.054	0.942
Bus	0.090	0.474	– 2.936	1.477	0.463	0.113	-0.491	0.942
Car	0.090	0.474	0.320	– 1.024	0.463	0.113	0.054	-0.653

Source: Author's compilation based on (Martín and Nombela, 2007)

In their 2004 conference paper Lythgoe and co-authors (Lythgoe, Wardman and Toner, 2004) mainly investigated the potential for extension of rail demand models to station selection and access to the rail network. In doing so, they described the basic summary model as:

$$Q_{aijb} = Kp_a F_{ai} p_b F_{jb} F_{ij}$$

where:

Q_{aijb}	number of journeys from origin zone *a* to destination zone *b*
p_a, p_b	populations in zone *a* and zone *b*, respectively
F_{ai}	function of the utility of access zone *a* to the origin station *i*
F_{jb}	function of the utility of egress from the destination station *j* to zone *b*
F_{ij}	function of the utility of the rail journey from station *i* to station *j*

In 2008 Dai and Jin analysed Chinese rail traffic data using a gravity model (Dai and Jin, 2008). The basic gravity model was applied to the estimation in the following form:

$$\mathrm{Ln}T_{ij} = \mathrm{Ln}C + \alpha \mathrm{Ln}P_i + \gamma \mathrm{Ln}P_j + \beta \mathrm{Ln}d_{ij}$$

where:

T_{ij}	railway passenger traffic volume from city *i* to *j*
C	constant
P_i, P_j	non-agricultural population of cities *i* and *j*
d_{ij}	railway distance between city *i* and *j*

In 2009, Debrezion, Pels and Rietvel investigated the effects of access modes on station choice using a nested logit model on Dutch data (Debrezion, Pels and Rietveld, 2009). The basic spatial model was defined as follows:

$$T_{ij} = A_i O_j B_j D_j f(GJT_{ij}) g(GJT_{ij} | d_{ij}) exp(\xi_{ij})$$

where:

O_i	total number of trips originated in station *i*
D_i	total number of trips attracted by station *j*
GJT_{ij}	function of the generalized journey time between stations *i* and *j*
d_{ij}	distance between stations *i* and *j*

The GJT indicator is calculated according to the following formula, which defines a quality variable, Generalised Journey Time, for a route:

$$GJT_{ij} = \frac{1}{2}F_{ij} + IT_{ij} + TT_{ij} + 10NT_{ij}$$

where:

F_{ij}	frequency between stations *i* and *j*
IT_{ij}	in-vehicle time between stations *i* and *j*
TT_{ij}	transfer time between stations *i* and *j*
NT_{ij}	number of transfers between stations *i* and *j*

This indicator is calculated and used by the Dutch Railways (Nederlandse Spoorwegen) to measure the service quality of a route, and is structured in a way that is virtually identical to a similar tool used by British Rail (Wardman, 1994).

In their 2014 paper Clewlow, Sussman and Balakrishnan examined the impact of high-speed rail and low-cost airlines in European air transport (Clewlow, Sussman and Balakrishnan, 2014). In terms of measurement methodologies, the authors present the overall impact of the proliferation of non-aggregate modal choice models in transport demand forecasting, while pointing out that at the same time, methods based on regression-based aggregate data are still present, especially in the analysis of long-distance markets. In the case of air transport, the following standard model is typical:

$$\ln(D) = \beta_0 + \beta_1 \ln(GDP) + \beta_2 \ln(Yield) + \beta_3 X + \varepsilon$$

where:

D	demand, measured by passengers (or flights)
GDP	combination of the average gross domestic product at the origin city and destination city
$Yield$	average fares paid for air transportation service
X	number of additional parameters that are expected to influence demand
ε	random error

The basic data is often complemented by additional local information – such as population density, hub status of the airport, presence of low-cost airlines. For the measurement, a database of 35 airports and 90 origin-destination-pairs was compiled, including information for the years 1995-2009. The key variables of the model are presented in Table 10.

Table 10. Variables of the model applied to European air transport

Dimension	Variable	Source
Aviation demand	Air traffic (passengers/ year)	Eurostat
Price/ fares	Jet fuel (Euros per barrel/ year)	EIA
GDP	GDP (millions of Euros/ year)	Eurostat
Population	Population (population/ year)	Eurostat
Population density	Population density (inhabitants per km2/year)	Eurostat
Rail competition	Rail travel time (in-vehicle travel time, in minutes)	RENFE, SNCF, DB Bahn, National Rail, British Rail
Hub status	Hub of a major airline (binary)	Airport publications and airline data
Low-cost carrier presence	Low-cost carrier service (binary)	Airport publications, news articles, and press releases announcing low-cost carrier routes

Source: Author's compilation based on (Clewlow, Sussman and Balakrishnan, 2014)

Based on the standard model, the estimating function was used in this form:

$$\ln(OD\ Demand_{it}) = \beta_0 + \beta_1 \ln(Rail_{it}) + \beta_2 \ln(X_{it}) + \mu_i + \varepsilon_{it}$$

where:

X_{ij} parameters that are known to influence air traffic demand: GDP, population, density, and fuel price (as a proxy for airfares)

t time (year)

The impact of high-speed rail has been studied in two ways. In the first method, time spent on board (travel time) was used as an explanatory variable, the results of which are shown in Table 11.

Table 11. Estimating the competitive impact of rail based on time spent on board

Parameter	(1)	(2)	(3)	(4)
ln(GDP)	0.297 (0.714)	2.280 (0.872)	3.548 (1.273)	5.192 (1.696)
ln(fuel price)		– 1.863 (0.476)	– 2.360 (0.714)	– 2.304 (0.906)
ln(population)			1.961 (0.997)	1.818 (1.1269
ln(density)			– 0.427 (0.549)	– 0.376 (0.603)
ln (rail time)				4.734 (0.834)
Constant	8.066 (7.389)	– 5.471 (8.111)	– 41.774 (19.187)	– 82.329 (24,589)

Source: Author's compilation based on (Clewlow, Sussman and Balakrishnan, 2014)

In the second estimation method, the presence of rail is included as a dummy variable (so rail travel speed is no longer included in the model), as shown in Table 12.

Table 12. Estimating the competitive impact of rail based on the presence of rail

Parameter	OLS	RE
ln(GDP)	0.215 (0.166)	0.064 (0.284)
ln(fuel price)	– 0.203 (0.032)	– 0.201 (0.093)
ln(population)	0.639 (0.090)	0.724 (0.254)
ln(density)	0.207 (0.050)	0.253 (0.119)
Hub status	0.896 (0.151)	0.824 (0.429)
High-speed rail	– 0.488 (0.100)	– 0.123 (0.059)
Low-cost carrier	– 0.018 (0.122)	– 0.054 (0.058)
Constant	2.441 (2.188)	2.170 (4.747)

Source: Author's compilation based on (Clewlow, Sussman and Balakrishnan, 2014)

2.5 Models focusing on quality measurement

At first sight, the impact of service quality on the demand for transport services is a methodologically more difficult problem to grasp, since the assessment of quality opens up definitional issues, bringing in issues with many subjective elements into the well-established, well-measured demand dimensions (price and quantity). The variations of possible models can thus in principle be realised in a broader space, but there are also often recurrent elements due to the inherent limitations of data sources. In the following review, I have tried to present some examples of good practice ideas and to illustrate the diversity of the literature with some more specific and individual projects.

Rahaman and Rahaman published their research on quality in rail transport in Bangladesh in 2009 (Rahaman and Rahaman, 2009). Railway plays an important role in the life of the country, accounting for 20% of all passenger transport. The study collected data on one railway line (Bheramara-Khulna) and examined the relationship between quality and passenger satisfaction. Table 13 presents the evaluation criteria of the survey. At first glance, one might think that a developing country like Bangladesh might use a quality assessment table that is far removed from the specificities of Hungary, but the dimensions are in fact also applicable.[7]

Some elements show the difference between the two transport service cultures (e.g. in-vehicle catering is essential in the East and is declining in Hungary in line with Western trends), but most of them are adequate aspects in Hungary. However, the problems of timetable regularity, delays and timetable parameters are not mentioned at all.

Table 13. Evaluation criteria for the quality questionnaire

No.	Attributes	No.	Attributes
V1	waiting arrangement	V11	journey time
V2	seat condition	V12	train announcement
V3	spacing among seats	V13	waiting arrangement
V4	spacing for moving on train	V14	ticketing time
V5	luggage storage facilities	V15	information availability
V6	window condition	V16	toilet facilities in the station
V7	environment inside the train	V17	security in the station
V8	condition of toilet inside the train	V18	safety caution
V9	fooding inside the train	V19	announcement in the station
V10	security inside the train	V20	behaviour of the staffs in station

Source: Author's compilation based on (Rahaman and Rahaman, 2009)

In 1997, Germá Bel published a study analysing the effects of travel time on long-distance rail demand (Bel, 1997). The data are transport data for Spain in the period 1988-1991. The dependent variable is the change in rail ticket sales over the period by route, and the source is the sales database of the Spanish National Railways (Red National de Ferrocarriles Espanoles). The two explanatory variables (also by route) are the change in travel time by rail and road over the period. While in the first case the data are timetable data, in the second

[7] It should be noted, however, that if we look at the photographic illustrations included in the study, there are of course huge differences.

case the average speed available on the roads on a given route has been calculated by taking into account the length of the express and conventional road sections and the average journey times for the two types of road. A dummy variable was also added to the model, which took the value 1 if air transport supply in the destination increased over the period.

The author also examined a separation of the types of service, as previous analyses had shown that these segments differ significantly in terms of demand characteristics – thus, short-haul services (under 400 km) were included separately, and for services over 400 km, an integrated category and a second category containing only night trains were included.

The estimation produced the results shown in Table 14. In general, the observed elasticities are in line with intuition, with the road cross-elasticity for all trips above 400 km being inconsistent (the t-statistics for this value are also weaker). As expected, the impact of air traffic growth on short-distance rail is weak in the model, significant above 400 km, but less so on night rail services. This can easily be explained by the inelasticity of demand for these special services, interpreted another way: passengers who stick to the less fashionable form of sleeper travel in this period also react inelastically to the increase in air transport supply.

Table 14. Impact of journey times on rail demand, Spain, 1988-1991

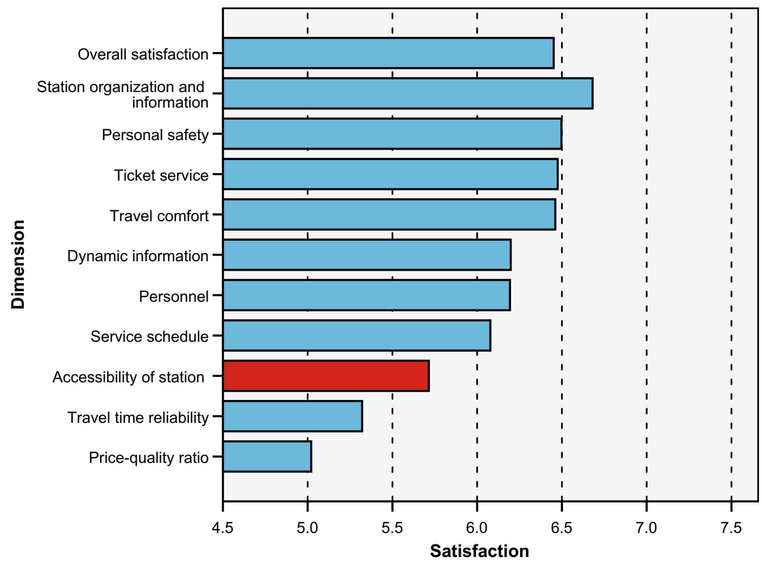
	100 < x < 400, day trains	400 km < x, day and night trains	400 km < x, night trains
Constant	0.158	– 0.050	– 0.077
Change in journey time by rail	– 2.660 (4.581)	– 2.372 (3.771)	– 1.750 (2.114)
Change in journey time by road (coach)	2.550 (3.916)	– 0.345 (0.491)	1.207 (1.639)
Increase of air service, dummy	– 0.067 (0.800)	– 0.225 (3.946)	– 0.124 (2.190)
Decrease of air service, dummy	–	0.173 (2.111)	–
R^2	0.642	0.760	0.377

Source: Author's compilation based on (Bel, 1997)

In a 2009 paper, Brons and co-authors examined the impact of the availability of railway stations on demand (Brons, Givoni and Rietveld, 2009). As seen earlier, the literature tended to take a more abstract approach to transport services in the early period, but as models became more detailed and the information technology tools became more sophisticated, more and more additional elements began to appear. At the same time, the role of congestion avoidance or environmental motivations in urban transport began to play an increasingly important role, rather than that of constraints (lack of own vehicle), and transport policy began to focus on positive incentives and user motivations. With this transformation, the role of a whole journey approach, which focuses on the whole transport value chain, has increased in transport policy, and in parallel in development and research.

In this new framework, accessibility and access to stations is no longer an urban transport problem, but a key element of the rail service value chain. The source of the data is the Dutch transport network, which is one of the most well organised and integrated transport systems in Europe. Brons and co-authors also refer to a survey conducted by the Dutch railway company, in which four of the 37 satisfaction factors examined addressed issues related to the accessibility of station services, such as connections to public transport, parking (car and bicycle), etc. For the estimation, the authors used the categories of this standard, regularly asked questionnaire in their own measurement. To assess importance, an indirect methodology was chosen, in which categories being close, moving together and having similar meanings were grouped in several steps, resulting in a total of eight main groups. The results of the evaluation are summarised in Figure 13 and the regression results are presented in Table 15.

Figure 13. Accessibility of stations in the evaluation criteria



Source: (Brons, Givoni and Rietveld, 2009)

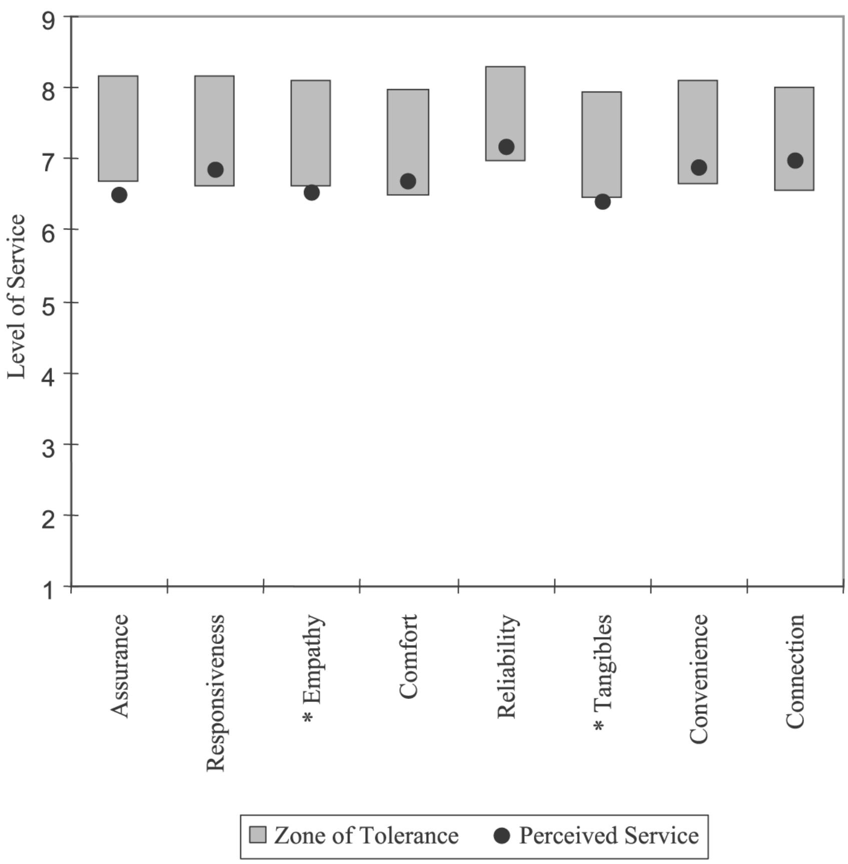
Table 15. Results of regression on the importance of evaluation criteria

Variable	Coefficient	*t*-value
(Constant)	1.20	18.66
Service dimensions		
Travel time reliability	0.20	35.74
Price/quality ratio	0.08	14.72
Travel comfort	0.21	22.02
Dynamic information	0.09	10.25
Ticket service	0.02	2.98
Station organization and information	0.10	10.70
Services schedule	0.90	10.29
Personnel	0.01	1.07
Personal safety	0.02	2.47
Access		
Guarded bicycle parking facilities	0.00	– 0.07
Unguarded bicycle parking facilities	0.01	2.17
Connection train with other public transport	0.03	6.66
Car park capacity	0.02	3.49

Source: Author's compilation based on (Brons, Givoni and Rietveld, 2009)

In 2007, Cavana and colleagues investigated the question in, what extremes rail passengers are willing to tolerate in different quality-dimensions in a given area (Cavana, Corbett and lo, 2007). The methodology was based on the SERVQUAL methodology developed by Parasuraman, and the data source was a survey conducted in Wellington, New Zealand, where 429 out of 800 questionnaires sent out were completed by passengers. The measurement data for the assessment tolerance zones are presented in Figure 14. It is striking that the observed values are typically at the bottom of the range, sometimes outside it.

Figure 14. Service quality tolerance zones and observed data

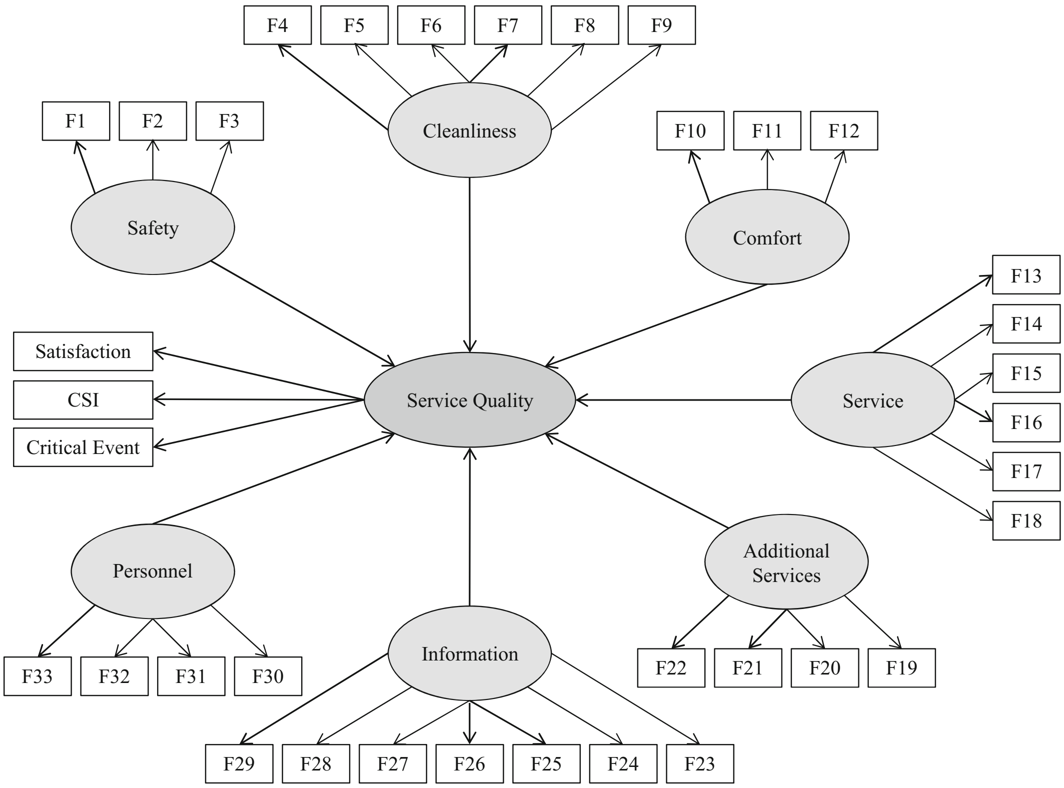


Source: (Cavana, Corbett and lo, 2007)

In 2015, Eboli and Mazzulla studied the relationship between rail satisfaction and quality in northern Italy (Eboli and Mazzulla, 2015). The sample included 32 regional rail lines, 9 suburban lines and two express lines. The data source was based on the data of more than

16,000 people, which were used to calibrate the model using the Structural Equation Modeling (SEM), a methodology to detect latent effects. The authors defined the service level factors in the structure shown in Figure 15.

Figure 15. Service quality factors



Source: (Eboli and Mazzulla, 2015)

The measurement results are presented in Table 16. In addition to the importance and satisfaction indicators for the factors assessed, it also shows the value of the Customer Satisfaction Indicator (CSI).

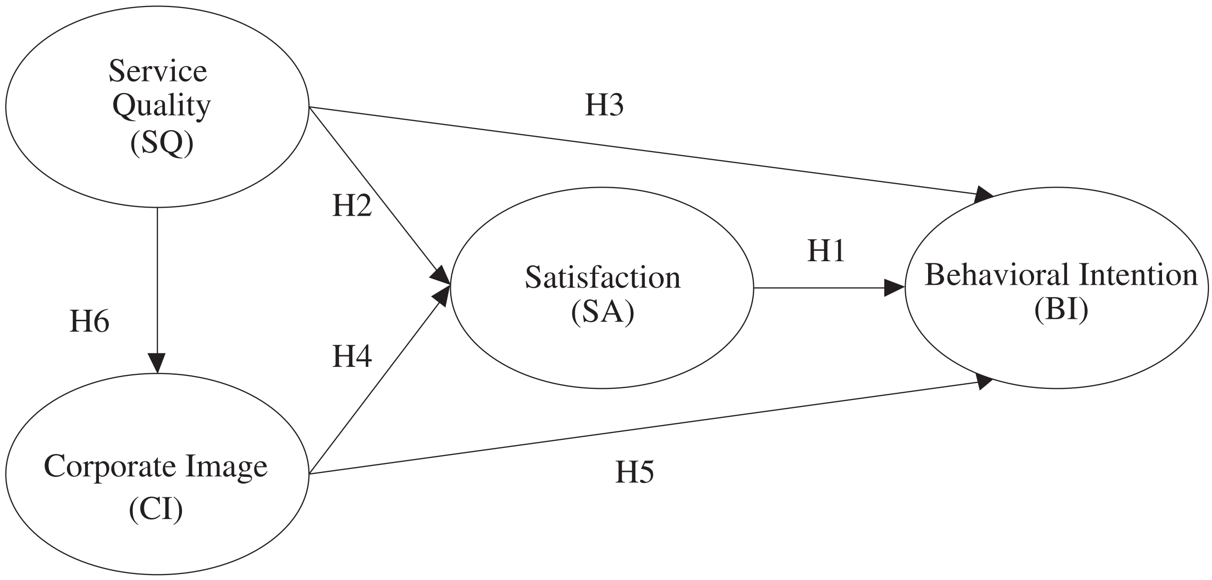
Table 16. Importance and satisfaction scores for quality factors

Service quality attribute	Importance rate	(var)	Importance rate	(var)	CSI index
F1 Travel safety	9.2	2.6	7.4	4.6	0.24
F2 Personal security on board	9.1	2.6	6.7	4.9	0.24
F3 Personal security at station	9.1	2.8	6.5	5.1	0.21
F4 Cleanliness of vehicles	8.9	3.4	5.0	5.5	0.16
F5 Cleanliness of seats	8.9	3.5	4.8	5.6	0.15
F6 Maintenance of seats	8.6	3.5	5.1	5.6	0.16
F7 Cleanliness of toilet facilities	8.8	3.8	4.4	5.7	0.14
F8 Cleanliness of stations	8.5	3.3	5.3	4.9	0.16
F9 Maintenance of stations	8.3	3.7	5.4	4.9	0.16
F10 Crowding on board	8.4	3.5	5.4	5.7	0.16
F11 Air-conditioning on board	8.7	3.1	5.1	6.0	0.16
F12 Comfort on board	8.4	3.1	5.6	5.2	0.16
F13 Fare/service ratio	8.8	3.4	5.1	5.5	0.16
F14 Frequency of runs	8.8	2.6	5.9	5.1	0.19
F15 Punctuality of runs	9.0	2.9	5.4	5.8	0.17
F16 Regularity of runs	9.0	2.6	5.7	5.2	0.18
F17 Integration with PT	8.7	3.0	6.0	4.9	0.18
F18 Localization of Stations	8.6	2.7	6.5	4.4	0.20
F19 Parking	8.0	4.9	5.7	5.5	0.16
F20 Bicycle transport on board	7.3	5.8	5.8	4.6	0.15
F21 Facilities for the disabled	8.8	3.8	5.2	5.7	0.16
F22 Substitute services	8.4	4.0	5.4	4.9	0.16
F23 Information at stations	8.7	2.9	5.9	4.7	0.18
F24 Information on board	8.5	3.3	5.5	5.0	0.17
F25 Info timeliness at stations	8.7	3.0	5.5	5.0	0.17
F26 Info timeliness on board	8.6	3.2	5.3	5.2	0.16
F27 Complaints	8.5	3.7	5.0	5.5	0.15
F28 Communication to office	8.3	3.7	5.1	5.3	0.15
F29 Info connections with PT	8.5	3.3	5.4	5.0	0.16
F30 Kindness on board	8.5	2.7	6.6	4.2	0.20
F31 Competence on board	8.7	2.5	6.6	4.1	0.20
F32 Ticket inspection	8.3	3.8	6.3	5.2	0.18
F33 Kindness at station	8.6	2.9	6.4	4.9	0.19
Overall service			5.8	4.6	5.7

Source: Author’s compilation based on (Eboli and Mazzulla, 2015)

Kuo has investigated issues of service quality, corporate image, satisfaction and attitudes in Taiwan, focusing specifically on elderly passengers (Kuo and Tang, 2013). The subject of this study is the high-speed railway line in Taiwan, the design model of which was presented in the article by Chang et al. The line started operating in 2007. The measurement concept is based on the SERVQUAL model developed by Parasuraman, and the logical framework developed for the railway in Taiwan is summarised in the schematic model shown in Figure 16.

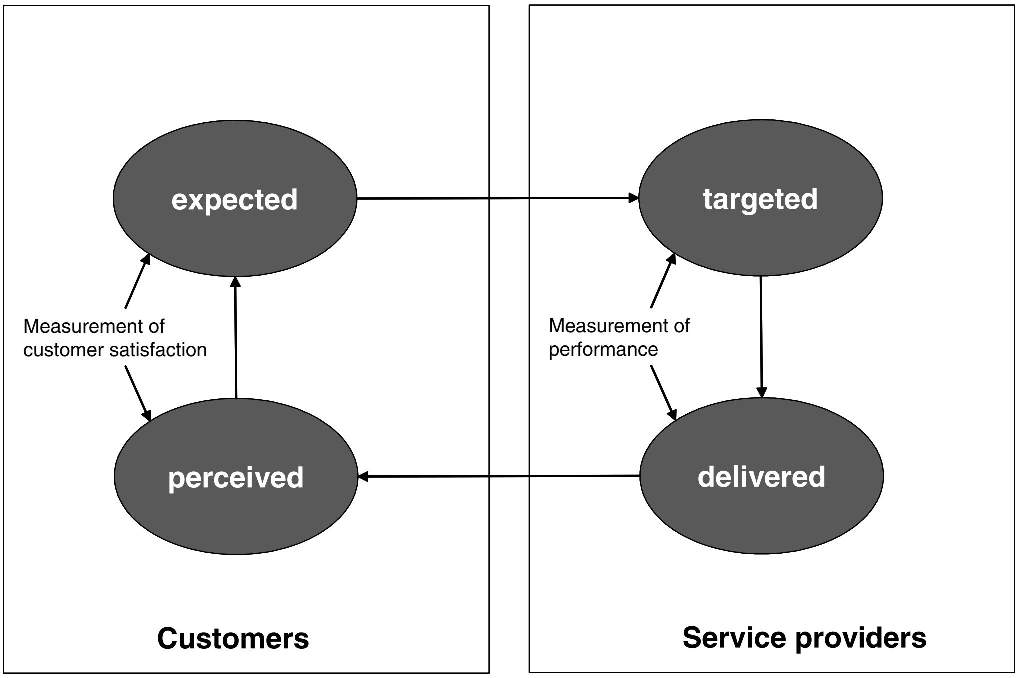
Figure 16. The decision-making process for elderly passengers



Source: (Kuo and Tang, 2013)

In 2008, Nathanail looked at rail data in Greece, using data from a 22-criteria rating system for railways (Nathanail, 2008). She plotted the evolution of customer experience and expectations as shown in Figure 17.

Figure 17. Quality loop of public transportation



Source: (Nathanail, 2008)

The 22-item system includes the quality factors assessed, grouped thematically. These factors are summarised in Table 17.

Table 17. Evaluation criteria and passenger service indicators of a railway operator

No.	Criteria	No.	Indicator	Sub-indicator
1	Itinerary accuracy			
2	System safety	2.1	Safety during trip	
		2.2	Safety at stations	
3	Cleanness	3.1	Train interior cleanness	
		3.2	Station cleanness	
		3.3	Train exterior cleanness	
4	Passenger comfort	4.1	Train temperature	
		4.2	Seat comfort	
		4.3	Rest comforts	4.3.1 Noise
				4.3.2 Vibrations
				4.3.3 Illumination
5	Servicing	5.1	Personnel behavior	5.1.1 Train personnel
				5.1.2 Station personnel
		5.2	Frequency of service	
		5.3	Quality and price of catering	5.3.1 Train catering
				5.3.2 Station catering
		5.4	Easiness of ticket purchasing at station	5.4.1 Waiting time
				5.4.2 Ticket availability
		5.5	Speed	
		5.6	Personnel appearance	5.6.1 Train personnel
				5.6.2 Station personnel
5	Servicing	5.7	Ticket purchasing facilities	5.7.1 At station
				5.7.2 Via telephone
				5.7.3 Via INTERNET
		5.8	Bed services	
		5.9	Escorted vehicles services	
6	Passenger information	6.1	Information during trip	
		6.2	Information at station	6.2.1 Announcements
				6.2.2 Bulletin board
		6.3	Pre trip information	

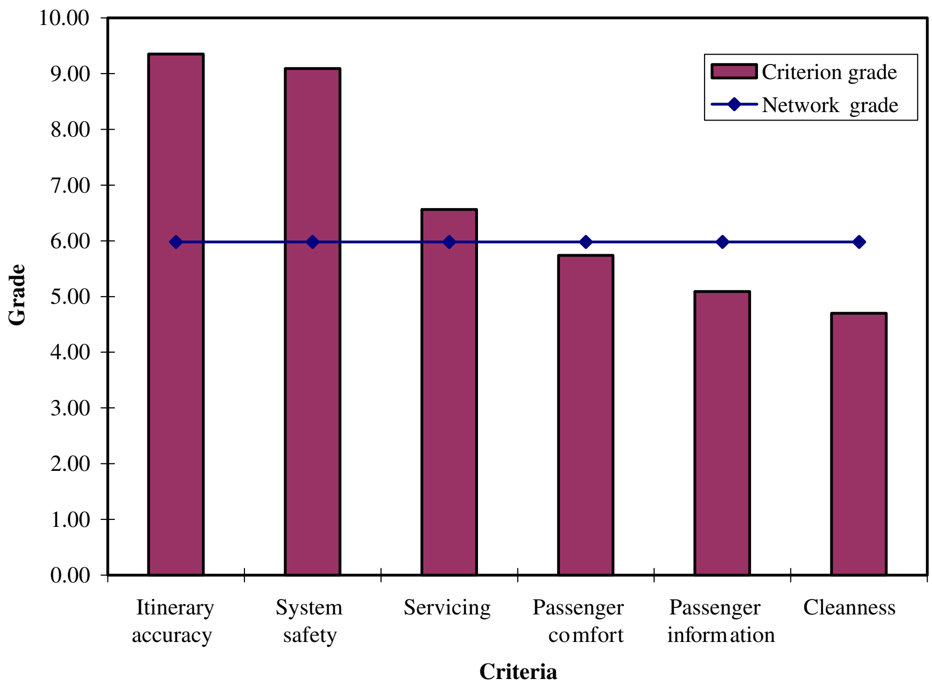
Source: Author's compilation based on (Nathanail, 2008)

The analysis relied on two data sources. A questionnaire survey of passengers was carried out and a mystery shopping survey was also conducted, supplemented by operator data. The first set of data was qualitative, the second and the third quantitative. After filtering out co-movements and further aggregation, the following seven data types remained from the broad set:

1. Itinerary accuracy
2. System safety,
3. Seat comfort,
4. Frequency of service
5. Speed,
6. Bed services,
7. Escorted vehicles services.

A total of 29 railway lines are included in the database. The measurement results are presented in Figure 18.

Figure 18. Results of the quality assessments



Source: (Nathanail, 2008)

In his 2008 paper, Litman presents data from a survey in Australia (Litman, 2008). The questionnaire survey was conducted by RailCorp, an Australian rail operator, and asked passengers to rate various improvements according to how much extra cost or extra journey time they would be willing to tolerate in return for a ten percentage point improvement in a given dimension. The results are summarised in Table 18.

Table 18. Subjective consumer evaluation of the benefits of improvements

Type of improvement	Additional fares aust. cent, (%)	Additional onboard time minutes (%)
Train layout and design improvements	5.6 (2.2%)	0.38 (1.0%)
Cleanliness	3.8 (1.5%)	0.26 (0.7%)
Ease of train boarding	3.2 (1.2%)	0.22 (0.6%)
Quietness	3.2 (1.2%)	0.22 (0.6%)
Train outside appearance	2.3 (0.9%)	0.15 (0.4%)
On-train announcements improved	2.3 (0.9%)	0.16 (0.4%)
Heating and air-conditioning	2.2 (0.8%)	0.15 (0.4%)
Improved lightning	1.9 (0.7%)	0.13 (0.4%)
Smoothness of ride	1.5 (0.6%)	0.10 (0.3%)
Graffiti removed	1.2 (0.6%)	0.08 (0.2%)
Seat comfort	1.1 (0.5%)	0.07 (0.2%)
Tickets at station	2.4 (0.9%)	16 (43.2%)
Station cleaning	1.9 (0.7%)	13 (35.1%)
Station building	1.4 (0.5%)	10 (27.0%)
Station staff	1.3 (0.5%)	9.0 (24.3%)
Ease of train on and off	1.1 (0.4%)	8.0 (21.6%)
Platform surface	1.0 (0.4%)	7.0 (18.9%)
Station announcements	0.8 (0.3%)	5.0 (13.5%)
Station safety	0.8 (0.3%)	6.0 (16.2%)
Station signing	0.7 (0.3%)	5.0 (13.5%)
Station graffiti removed	0.7 (0.3%)	5.0 (13.5%)

Retail	0.7 (0.3%)	5.0 (13.5%)
Platform seating	0.6 (0.2%)	4.0 (10.8%)
Lifts/escalators	0.4 (0.2%)	3.0 (8.1%)
Station information	0.4 (0.2%)	3.0 (8.1%)
Station lightning	0.4 (0.2%)	3.0 (8.1%)
Bus access	0.3 (0.1%)	2.0 (5.1%)
Bike access	0.3 (0.1%)	2.0 (5.1%)
Toilets	0.2 (0.1%)	1.0 (2.7%)
Car park	0.2 (0.1%)	1.0 (2.7%)
Car park drop-off	0.2 (0.1%)	1.0 (2.7%)
Platform weather protection	0.1 (0.0%)	0.4 (1.1%)
Subway/overbridge	0.1 (0.0%)	0.1 (0.3%)
Taxi	0.1 (0.0%)	0.1 (0.3%)
Telephone	0.1 (0.0%)	0.1 (0.3%)

Source: Author's compilation based on (Douglas Economics, 2006), cited by (Litman, 2008)

Wardman's article, published in 2011, reviews twenty years of literature on the topic of congestion. The analysis focuses on the multiplier approach, as we have seen earlier, e.g. in the summary work of Paulley et al. In the case of quality assessment, problems of translating results to monetary terms can be addressed by time-based accounting, or in some cases, as we have seen in the data of the Douglas study cited by Litman, both time and monetary assessment can be done simultaneously. The table in Wardman's summary that is most worth examining is the one in which he presents the results of a 2008 survey. Here you can browse in detail how a level of congestion in different situations is rated by travellers, the data is summarised in Table 19. The source data is from a published preference survey by MVA Consultancy for the UK Department for Transport.

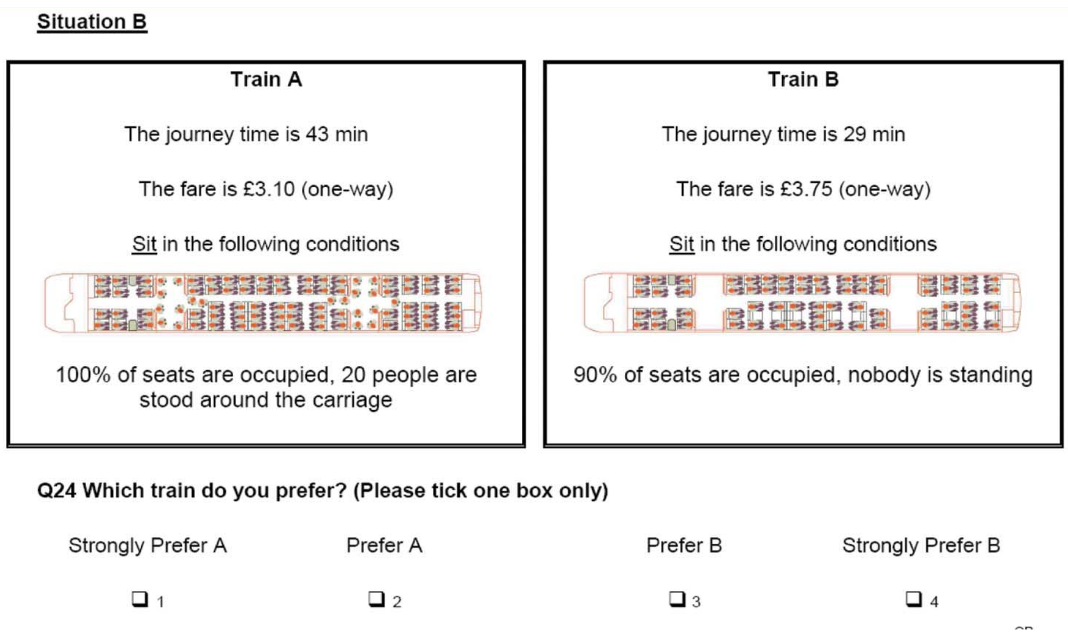
Table 19. Evaluation of congestion levels by trip type

Passenger /m^2	Non-business		Business		LSE		Regional		Interurban	
	Sit	Stand	Sit	Stand	Sit	Stand	Sit	Stand	Sit	Stand
0	1.00	1.48	1.00	1.91	1.00	1.43	1.00	1.34	1.00	1.77
1	1.10	1.58	1.13	1.95	1.09	1.56	1.24	1.61	1.11	1.81
2	1.21	1.68	1.27	1.99	1.18	1.69	1.48	1.88	1.23	1.85
3	1.31	1.77	1.40	2.03	1.27	1.82	1.72	2.16	1.34	1.89
4	1.41	1.87	1.54	2.08	1.36	1.95	1.96	2.43	1.46	1.92
5	1.52	1.97	1.67	2.12	1.45	2.08	2.20	2.70	1.57	1.96
6	1.62	2.06	1.81	2.16	1.54	2.21	2.44	2.97	1.69	2.00

Source: Author's compilation based on MVA Cons. (2008), cited by (Wardman and Whelan, 2011)

The logic of the questionnaire methodology used in the research is illustrated by the sample of questions shown in Figure 19.

Figure 19. Sample question from the stated preference research questionnaire



Source: MVA Consultancy (2008), cited by (Wardman and Whelan, 2011)

2.6 Conclusions

The review of the literature on this topic can be concluded with several lessons. Although the subject is largely related to corporate or public policy issues, and thus a significant part of the measurements is not typically carried out in the context of academic public life, there is a significant literature on modelling rail demand, more broadly transport demand, and on the analysis of quality factors.

While the role of method selection and simulation models and other bottom-up methods is significant, the available data and the purpose of the study primarily determine the choice of methodological tools. For the analysis of demand, the use of an origin-destination-pair database of traffic sales data within the framework of the gravity base model is therefore a realistic methodological solution. The use of the model is a direct consequence of the nature of the data available, but the experience in the literature suggests that it is indeed a suitable analytical tool for the study of the issues in question.

For the gravity model, there are more available basic data sources for the relevant data types and, in particular because the analysis is not aimed at forecasting or impact assessment, this format does not represent a limited methodological framework, as it is possible to include proxies and quality factors. Based on the literature, it cannot be clearly stated that more complex, multi-step models lead to more accurate results, but are more suitable for other purposes.

In the literature presented here, there are numerous examples of abstract approaches and quantification of various soft quality factors. The summarised results (thus mostly elasticity data) allow the evaluation of estimates of Hungarian rail demand and their comparison with the experience of transport systems in other countries.

3 METHODOLOGY

In the next section, I will describe the logic, structure and elements of the analytical model, the range of data used for analysis, and the main problems, their consequences and how to deal with them. I will briefly describe the process of building the model, highlighting the factors relevant to the outcome.

3.1 Research questions

The aim of my research, as stated in the introduction, is to build a demand model that allows to investigate the effects of substitution and quality indicators on the Hungarian passenger rail market.

I will attempt to build on the literature and available data sources to compile a database that can adequately describe the demand for passenger rail transport in Hungary. If the model has sufficient explanatory power, it will be suitable to investigate my research question, the different quality, service level factors and other influences on demand development.

A further research question is the relationship between quality effects: this is not a frequently studied issue based on the literature reviewed. My concept is based on the assumption that service quality on the Hungarian rail network is rather heterogeneous along the different dimension and that the overall service levels are unbalanced. For this reason, in addition to the formation of composite quality indicators, the relationship between highly variable quality elements will be also investigated. To test this, a hypothesis is proposed in which I assume that the customer experience, composed of many quality elements, is influenced by its minimum level rather than the average, i.e. that a very poor quality element pulls down the ratings of the others.

The research questions addressed are therefore:

- Can an appropriate explanatory model be developed for rail sales data based on the available data sources?
- What are the impacts of substitutes, economic background variables and quality factors on rail demand?
- Is there a relevant relationship or correlation between quality factors?

Given the right results, the following two important directions of inquiry emerge:

- What policy implications follow from the above results?
- What further directions of investigation do the results suggest?

3.2 Structure of the model

As the literature review shows, in demand analysis, unless a dedicated data collection is specifically designed for the research, the range of data available or obtainable is the most important determinant of the choice of methodology. Given that in the present case, initial data sources for travellers are typically available in aggregate format, data sources determining individual choice are scarce, and data showing the details of the modal split are not available, the literature reviewed suggests that a gravity equation-based model is the most appropriate methodological framework. A significant positive factor due to the time factor is that it is possible to build a panel database with a span of a decade.

3.2.1 The basic model

According to the gravity approach, the mobility demand between two geographical points is determined by the weight (population and/or economic activity) of those points and the distance between them, the mathematical form of which is inspired by Newton's law of gravity. The concept can be applied to various spatially definable economic activities, and is thus used to analyse trade and territorial integration issues (Anderson, 2011).

In the case of passenger rail transport, the population of the settlements (districts) concerned and the distance between points in the rail network are the basic elements of the

gravity model. The demand for rail passenger transport is represented by the number of passengers travelling on a given route (between two points) at a given time.
The gravity model can be described as follows:

$$D_{ij} = \frac{P_i P_j}{d_{ij}}$$

where:

- D_{ij} rail traffic demand between points *i* and *j*
- P_i, P_j population of settlements *i* and *j*
- d_{ij} distance of points *i* and *j*

The baseline model alone is not suitable for examining the effects of service quality: to rule out distortions due to additional effects not taken into account, it is necessary to look at the factors affecting demand more generally.
Demand is generally defined as the quantity of a good that market participants are able and willing to purchase at a given price. For the purposes of our model, this factor should be taken into account in the sense that only those transactions take place, where transport needs are matched by adequate welfare status. Therefore, in addition to the basic factors of the gravity model (population, distance), it is usually necessary to use some indicator of economic status representing the ability to pay. Such an indicator could be the value added of the local economy or data on activity and unemployment. Most research uses GDP for this purpose (as in Jones and Nichols, or Owen and Phillips), but income may also be appropriate (e.g. Brand et al.) This approach of the economic background can be used to filter out the distorting effects of the fact that the values predicted by the model for a region or destination do not take account of economic realities. In other words, if examining, for example, the transport demand of two regions of the same size and distance measured from the centre, where the economic situation of the two areas differs significantly, the model will also predict different transport demand in proportion to income and development, so that the estimates of the impact of other factors are not distorted by these differences.
A further problem that may arise for rail transport is its possible inferiority, as there are many arguments that it may be a non-preferred service for some customers, who may not use it above a certain income level. Thus, the income elasticity of demand for rail transport may

become negative in a higher range. This parameter is therefore an interesting research question in itself.
The basic model is therefore completed as follows:

$$D_{ij} = \left(\frac{P_i P_j}{d_{ij}}\right)^{\beta_1} \left(I_i I_j\right)^{\beta_2}$$

where:

$\beta_{1,2}$ coefficients of the factors

I_i, I_j economic development/income level of settlements/areas *i* and *j*

The population-based gravity model, adjusted for economic development, estimates the local demand for mobility, defined between two points, and incorporates the ability to pay, according to the definition of demand.

3.2.2 Substitutes

In the logical process of building a demand-model for rail passenger services, the next step after the previously created demand is the inclusion of mode choice. The decision involves a choice between a myriad of transport modes, where the most important are logically distinct types:

- Pedestrian transport
- Short distance private transport (scooter, bicycle, etc.)
- Local public transport (metro, tram, local bus, etc.)
- Individual transport by car
- Interurban public transport by road (coach)
- Interurban public transport by rail (rail)
- Air transport

Of course, further refinements may be feasible, but the vast majority of cases can be modelled along these main categories. In the context of mode choice and rail demand estimation, substitution appears as a relevant factor. In the case of rail transport, walking and short-distance individual and local transport are essentially not alternatives. The use of rail as a substitute for local transport is relevant, however, due to the underdeveloped nature of the Hungarian rail network and services, its presence is minimal in Budapest, where such a transport service would make the most sense, due to the inadequate supply (Budapest Fejlesztési Központ, 2020). Similarly, air transport can be ruled out as a realistic alternative for domestic rail transport due to lack of supply.

Therefore in the vast majority of cases, private and public road transport is a realistic substitute for rail transport. Both modes of transport are based on the road network and, although there are significant differences in journey times and other service dimensions, the road alternatives are essentially determined by the nature, route and condition of the road network. Although private transport is necessarily faster and more convenient than long-distance bus services, so there is fundamental difference between the two, but at the same time there is also a proportionality. A key question is how this proportionality is achieved for the journey times of private and public road transport in each direction. Table 20 gives some examples showing that proportionality can be observed. The variation is mainly due to the fact that, for a given pair of destinations, the route taken by the driver is not necessarily the same as the fastest route by car, and also due to the number of stops and the type of vehicle (e.g. on motorways only specially authorised buses can travel at higher speeds).

Table 20. Proportionality of bus and car journey times

Route	Travel time		Car/coach ratio
	Coach	Car	
Békéscsaba – Szeged	112 min	83 min	74.1%
Budapest – Baja	105 min	91 min	86.7%
Budapest – Salgótarján	97 min	80 min	82.5%
Budapest – Zalaegerszeg	174 min	144 min	82.8%
Siófok-Pécs	187 min	140 min	74.9%
Székesfehérvár – Tatabánya	80 min	55 min	68.8 %

Source: Author's compilation based on (Google, 2021b)

The choice of road mode, regardless of its individual or collective form, is therefore in competition with rail transport in terms of journey times, and the two alternatives can be approximated in terms of time by a single journey time. This is important because of the scarcity of available data, especially in the absence of an adequate database on bus services, which makes it difficult to estimate the latter. Data from journey planner systems do not usually provide pure bus journey times, as the overlap between rail and bus services has been further reduced in recent times, so that public transport journey times often include the use of rail services. Car trip planning is therefore more complexly suited to represent both road alternatives.

The impact of road transport, both individual and collective, can be well measured by the travel time by road between the two points. In Table 21, I show the conclusions on the expected demand for different combinations of road and rail travel times.

Table 21. Comparison of bus and car journey times in relation to rail

Travel time by coach	Travel time by car	Result
More favourable than rail	More favourable than rail	All passengers choose the road
Worse than rail	More favourable than rail	Those who can travel by car, other passengers typically travel by rail
Worse than rail	Worse than rail	All passengers choose the railways

Source: Author's compilation

Of course, preferences evolve in a much more complex way, so the effects summarised in the table do not hold in a function-like way, but there is no way to capture a more subtle mechanism than this in the assumptions of the model. In order to investigate road substitution as well as possible, ideally it is necessary to know the individual and the collective road travel times, but even with one of these data the model can be improved substantially, as a significant correlation between the two values can be assumed.

Based on the above, the model can therefore be further refined with road substitution according to the following logic:

$$D_{ij} = \left(\frac{P_i P_j}{d_{ij}}\right)^{\beta_1} \left(I_i I_j\right)^{\beta_2} {T_{ij}}^{\beta_3}$$

where:

T_{ij} Journey time on road

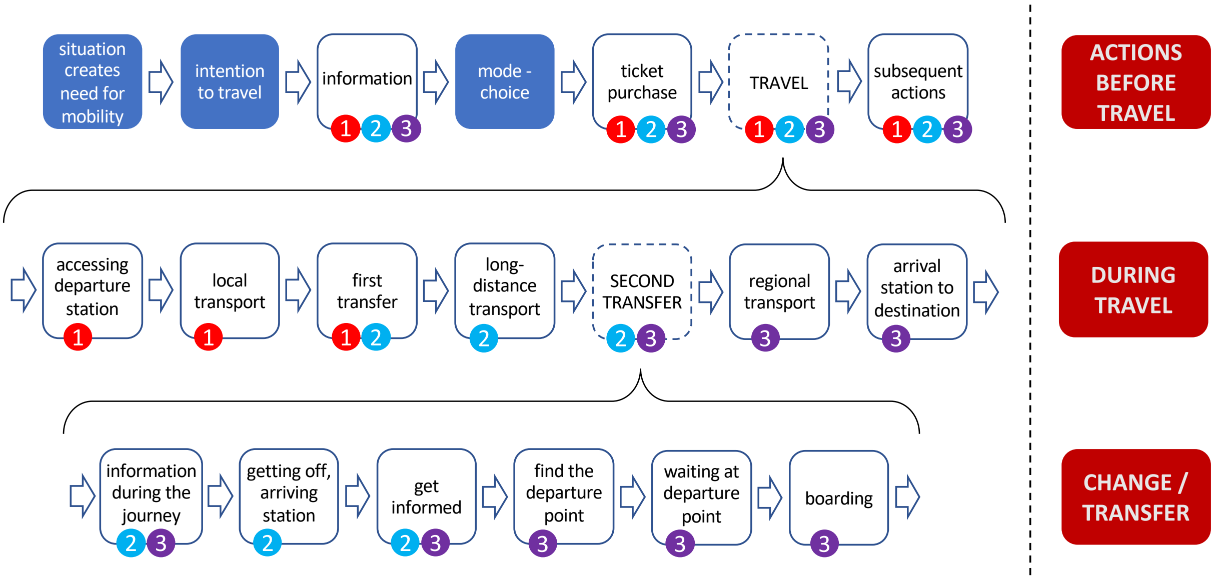
The model thus defined combines the primary determinants of demand with the effects of substitution. However, other explanatory variables may be used in the model in relation to substitution, such as the price of fuel or the supply of passenger cars, which will be discussed later.

3.2.3 Quality

With mobility decisions we have reached the logical point where the demand for a specific destination has emerged, incorporating the economic background variables, and then the remaining demand, taking into account the substitutes apply to rail service. It is at this point that the quality factors that are the primary object of the analysis are introduced into the logic model.

In the case of rail service, we need to assess the quality of a complex service package, which is made up of many elements. There are a large number of interactions throughout the whole process. The passenger meets many interfaces and goes through a lot of situations, from the decision to travel to the arrival and possible follow-up. During the process, the customer's experience depends on different factors, which are not necessarily interlinked. A schematic diagram of a typical travel chain is shown in Figure 20, which illustrates this complexity, ranging from the soft elements (e.g. information gathering) prior to the journey to the physical parameters of the service (e.g. journey time). Particularly relevant is the role of mode change, which typically receives less attention as a no-man's land between two service providers, while its role is clearly crucial for the perception of quality (Édes, 2019).

Figure 20. Customer experience-focused customer journey schema for a travel chain



Source: (Édes, 2019)

In the present research, transfers and related elements are only indirectly investigated in the data of one operator (mainly the quality of station services). However, through the quality data sources, significant soft elements such as information services and cleanliness are incorporated, allowing a wider range of quality effects to be investigated.

Due to the specificities of the assets in the Hungarian railway network, the operating conditions and the institutional system, it is particularly characteristic that the travelling experience does not show a uniform, standardised quality: large differences can be observed. This randomness means that the differences in service quality are not necessarily the result of a well-founded business decision, some kind of price discrimination. It is not the case that higher value-added customers paying higher prices receive a better quality of service, or even that stations with higher traffic volumes are in a privileged position (the unacceptable conditions of the main stations in Budapest are good examples). Completely different aspects shape the quality differences, which from the customer's point of view appear to be almost random and show a significant variation.

In the following section, I will review the main factors determining the complex quality of service in terms of their suitability for inclusion in the model:

1. *Enquiry:* Prior information before travel-decision provides the individual with information on the possible ways to meet the mobility need. In the case of individual travel, e.g. leisure, this is a very important moment, and often depends on the availability (a typical example is low-cost flights, where people often choose a destination from the available options and do not look for flights to a destination). In the case of business travel, for example, the traveller collects information with some basic experience, typically knowing the options, their quality and the conditions of use, but even for regular travel, prior information may be relevant – e.g. checking the current timetable or the traffic situation on the roads using real-time navigation. The availability, processability and quality of advance information is an essential service element, but in the model it would only be relevant towards substitutes, since for public rail services, a uniform level of information is available on a uniform interface, and thus no heterogeneity is present in the model. An exception to this is the station information, the use of which is not relevant for the preliminary orientation, but will be discussed later in the station phase.

2. *Travel decision:* the travel decision is based on prior information, other experiences and assumptions. In the case of rail demand, the latter can have a significant negative influence, which can affect decisions to varying degrees for a destination. However, this mechanism is so indirect that its representation in the model, even with adequate data, would be more distorting than accurate.

3. *Ticket purchase:* the ticket purchase process is a transaction cost, every second spent on it is a technically necessary, but inconvenient time for the traveller. Accordingly, it can be described by two important factors. Firstly, there is the question of the platform and interface on which the purchase is made, e.g. face-to-face only or online in a more flexible way. Secondly, the barriers to the consumer in the shopping process, from queuing to over-complicated online interfaces. Although the database contains information on the way tickets are purchased, this is not directly suitable as an explanatory variable: the share of digital channels is growing significantly year on year, this underlying trend would mask the additive effect of vending machines as an available service, and the online sales confuses further the picture. In a more

refined approach, a dummy variable with a list of ever changing list of stations with vending machines could be used to test this solution, but since the operator makes traffic a priority when deploying vending machines, causality would be very difficult to prove here due to the endogeneity problem.

4. *Visiting the departure station:* the customer experience of accessing a station is a particularly important element of the transport value chain. This would require good quality local access and transport data, which is not available from the data sources I use for this analysis. On the other hand, with adequate transport data, a more refined model could provide the distance between the city centre and the station, or the value of the time required to travel there. In this case, similar data from substitutes could be included.

5. *Station services:* the condition of the stations and the services they provide are key factors in customer experience. A wide range of services are available, including parking, bicycle storage, heated lounges, internet access, catering, etc. At the same time, stations of the Hungarian rail network often lacks some of the most basic elements, and it is even possible to imagine cases where the lack of an object would result in a more pleasant passenger experience due to its bad condition. Therefore, not only the existence of the various factors, but also their condition and cleanliness are important factors.

6. *Vehicle, travel conditions:* the role of travel conditions in the customer experience is similar to that of stations, but even more important due to the temporal specificity of the journey, which has unavoidable effects over a longer period. The availability of services and their condition are decisive, with the availability of seats and the functioning of cooling/heating being particularly important. The age of the rolling stock is an important issue: a large proportion of passengers prefer new rolling stock, but in terms of comfort, a 20-30 year old carriage in good condition and cleanliness might be more comfortable than, for example, a modern EMU trainset.

7. *Finding your way around while travelling:* a separate question, independent of the travel circumstances, is whether it is possible to obtain information while travelling. Beyond traditional solutions such as looking out of the window for information, or using enquiries and timetables, the possibilities are now extensive with electronic on-board information tools. More and more vehicles have displays that provide continuous and real-time data, and online navigation on your own devices can be even more reliable. Of course, the quality of these could be questionable also, and the damage caused by incorrect information is often greater than the cost of not providing it. These data are particularly difficult to measure, and comprehensive information is not available. Further complicating the issue is the fact that inadequate quality of information at stations and on-board is becoming less of an issue, as good quality, real-time information is available in the railway company's mobile application.

8. *Staff communication:* contact with staff is an absolutely critical part of the travel experience, a factor that is highly subjective and difficult to measure, and not even constant within a year, so it is not realistic to assign this type of explanatory variable to the annual turnover of an origin-destination pair.

9. *Timetable:* the most basic element of a transport service is speed, the planned (scheduled) journey time. This data is an indispensable part of the model, and is available retrospectively from the timetable data. However, its measurement may raise methodological issues, as some form of averaging is required in the calculation.

10. *Accuracy:* journey times as a theoretical timetable target may not be achieved in reality. As has been shown in the literature on relative time evaluation, psychologically this factor is particularly relevant, with delays and reliability featuring prominently in all rankings. However, different events may have different time ratings, so a delay of ten minutes may affect travellers worse than a ten-minute longer, but previously announced journey time. Measuring the data is possible, but again there are many uncertainties about usability and connectivity.

11. *Destination services:* the destination is of somewhat less importance in the travel value chain, as it is typically a shorter stay than the departure. However, the evaluation criteria may be the same as for the departure. An interesting issue in distinguishing between the two is the presumed symmetry of outward and return journeys: I will return to this issue in the results.

12. *Onward journey:* the onward journey is a factor similar to, and measured in the same way as, the visit to the departure station.

13. *Transfer:* as I briefly discussed in the first chapter, one of the major problems of transport systems is the presence of isolated elements – the user's journey consists of a combination of transport modes, but such interconnections are not taken into account sufficiently by service systems. However, transfers are often made within a single operator's system, and given the right data such an analysis is possible, in particular to compare the impact on demand of serving a destination pair with or without transfers.

14. *Post-trip services:* an essential element of the complex service is the way in which post-travel services – typically complaint handling, handling of lost property etc. – are provided. These data are not really measurable by any means other than targeted questionnaires and data collection. While there is a long tradition of such surveys in the service sectors – e.g. the prominent use of NPS (Net Promoter Score) indicators – no similar information is available for the Hungarian State Railways.

My hypothesis is that the quality of rail services is a determinant of the demand for the service. In other words, there is a strong relationship between the quality of services and the demand for them. Of course, the quality of substitutes may also be a relevant factor, if adequate data are available on the subject.

In a more detailed analysis, the effects of the quality determinants should be examined individually, and the nature of the relationship between the effects, which is expected to be more difficult to detect due to the complexity of the model, is an additional research question. The totality of services together determines demand, and a single service element of blatantly low quality can drastically worsen the overall effect, so the role of these factors will be examined as a minimum function in the demand model also.

The final model can therefore be formulated in the following logical framework:

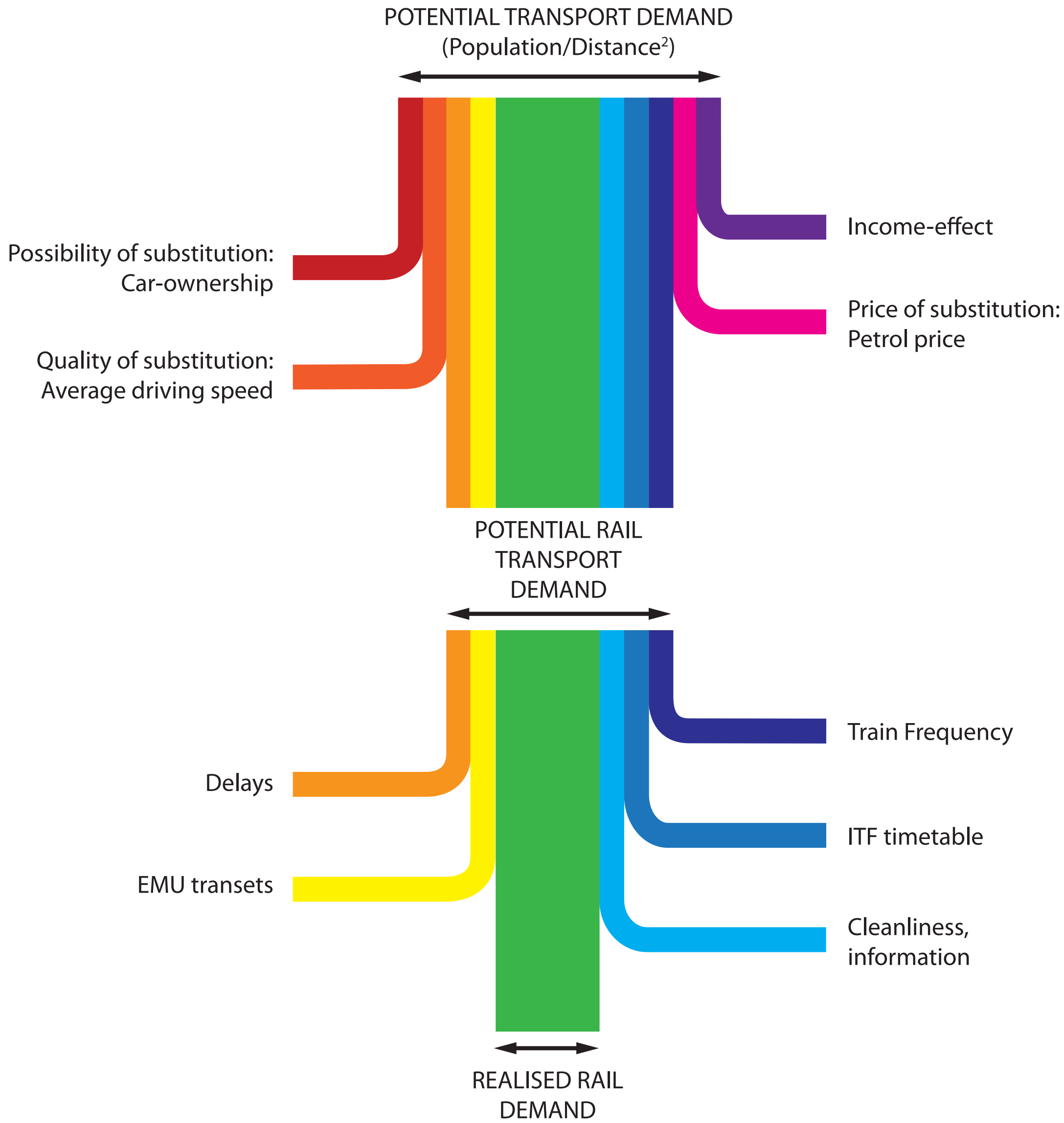
$$D_{ij} = \left(\frac{P_i P_j}{d_{ij}}\right)^{\beta_1} \left(GDP_i GDP_j\right)^{\beta_2} T_{ij}{}^{\beta_3} f\left(Q_{1_{ij}}{}^{\gamma_1} Q_{2_{ij}}{}^{\gamma_2} \ldots Q_{x_{ij}}{}^{\gamma_x} \ldots Q_{s_{ij}}{}^{\gamma_s}\right)$$

where

$Q_{x_{ij}}$ value of quality indicator *x* for route *ij*

The logical framework of the full model is shown in Figure 21.

Figure 21. Logical framework of the model



Source: Author's compilation

The estimation process involves testing several versions of each factor, with the statistical results allowing the most appropriate variables to be identified. The final model produced is then suitable for estimating the effects sought and accepting or rejecting the hypotheses.

3.3 Scope of the data used

3.3.1 Basic data for the gravity model

Demand data

The source of data for estimating demand for rail services is the database of the ticketing system of MÁV-Start. The dataset analysed contains sales data for a total of 120 months between January 2010 and December 2019.

MÁV introduced origin-destination ticketing in 2007. The previous system operated with a very inconvenient solution for data analysis, where only the distance of the journey was displayed on the ticket media (typically the amounts rounded to 10 km according to the fare structure), two classic examples of which are shown in Figure 22.

Figure 22. Types of rail tickets before 2007: bark tickets and computerised tickets

Source: (delpestibusz.hu, 2021) and (retronom.hu, 2021)

A major change was introduced in 2007 (followed by further improvements and increasingly refined data recording methods), from which time onwards tickets will record both the departure and arrival station, and will include route information if the specificity of the transfer or route requires it (i.e. in cases where several equally reasonable routes are possible). A new type of ticket is shown in Figure 23. Significant innovations since then have further improved the refinement of the sales data, but these details are less relevant to my research.

Figure 23. Printed computer ticket after 2007

Source: (Wikimedia.org, 2009)

Sales data since the 2007 reform allow the concept to be analysed by origin-destination pairs. Of course, in order to be able to properly estimate the turnover between two settlements, further questions arise in relation to the database:

1. Is the volume sold representative of the total passenger traffic (type I and type II error)?
2. Is the number of journeys by ticket type clear?
3. How can the problem of weekday and holiday traffic be addressed?
4. At what level of aggregation should traffic be analysed?
5. Is the route (and therefore length) of the journey clearly inferred from the ticket data?

The answer to the first question is, in many ways, a clear no. Presumably less relevant, but not negligible, is the number of cases where ticket changes are made but travel is not – it is reasonable to assume that the incidence of this problem is relatively low and not highly correlated with the explanatory variables, no such bias can be assumed. It could also be argued that the volume sold is more closely related to the concept of demand than the number of actual travel events.

The problem of trips that are not included is much more significant. This contains both unauthorised use and the number of unregistered free trips. For the first category, there is no reliable source of data, while for the second category there is an unmeasured but very significant number of passengers.

The magnitude of the latter can best be estimated on the basis of data for 2011. In 2009, a decision was taken to introduce a registration ticket for public long-distance transport in order to ensure that the state compensation for free travel is accurately accounted for between the state and the railway company (wikipedia.org, 2021b). This type of ticket, while bringing significant advantages in terms of the accounting system, also brought inconveniences on a day-to-day level, and in line with the public policy concept of the time, its use was discontinued on 1 January 2012. However, sales data for 2011 provide an estimate of the proportion of people travelling without paying and without any registration: at that time, 13.5% of the total number of tickets were purchased by over-65s with a registration fee of HUF 0, while it can still be assumed that those travelling free of charge did not fully comply with their obligation to 'buy' tickets.

A relevant question for this research is the impact of the lack of data on registration and unpaid travellers on the results. Two factors suggest that this alone does not cause a disproportionate bias:

– The sensitivity of paying users to quality is understandable, while free users are presumably less quality-oriented. Accordingly, the absence of free passengers just removes a potential risk of bias from the data.
– The primary transport policy objective at Community level, as discussed in detail in Chapter 1, is to increase the number of people choosing to travel by rail, and in particular the alternative of private motorisation. Thus, the target group for quality improvement is primarily the paying passenger, and the greatest potential benefits can be realised via this group. The behaviour of this target group is well represented by existing fare-paying passengers.

In relation to the problem of free riders, it is worth modifying the research question to include only the non-railway population under 65, so that this bias is excluded – for the above reasons, this is not a problem in terms of public policy relevance.

The second question is the relationship between the number of units and the number of trips. This question can be divided into two main themes. On the one hand, a distinction has to be made between return tickets (round trip) and one-way tickets, since in the former case two travel events are behind one transaction. On the other hand, the situation is much more complicated for season tickets, where, in the absence of check-in/check-out systems, there is no data available on the average number of trips that can be assumed to be behind a season ticket purchase. However, due to the specificities of rail pricing the latter problem is not nearly as difficult as for example in the case of metropolitan systems. For example in Budapest, the price of a monthly BKK season ticket is equal to the price of 27.1 tickets (BKK, 2021), so if we assume 2x1 trips per day, the price of the season ticket is already recovered for 14 trips. Accordingly, there can be a very large variation in the number of journeys behind the season ticket, since with many transfers a day, one can expect up to 100-200 boardings, but it is worth buying a season ticket for two weeks with a minimum of 2x1 boardings a day. On rail, the ratio is narrower: according to the tariffs, 38 tickets can be bought for the price of a monthly rail pass (MÁV-START, 2018a), so the ticket change is worth at least 19 working days, which means practically every working day.

The database includes the passenger flows according to MÁV's own internal accounting logic, in both cases calculating a passenger headcount figure from the volume sold based on a standard ratio. For return tickets, a multiplier of 2x is used as appropriate, while for season tickets it is 52x. This means that the database contains 26-day values, which are higher than the average number of working days per month (19-23 days) and include a certain number of weekend trips. As it is assumed that this is somewhat of an overestimate, I have also made a revised headcount estimate of 20 days as a more conservative lower bound. Although this calculation might underestimate the actual passenger numbers, it is definitely suitable to ensure that the number of season ticket passengers does not overwhelm the other passenger values. The model focuses on quality, and the reactions of more flexible travellers who do not commit to every day may be very relevant, so this is possible bias appears in a direction that is more favourable to this analysis.
The varying proportion of weekday traffic and public holidays, as well as the summer period – and of course a myriad of similar issues – touch on the problem of seasonality in rail traffic. The finer the level of data processing, the greater the significance of the influence of the calendar effect. There are a number of sophisticated ways of dealing with these situations, in the present case, we simply include annual aggregated data in the model, so most of these effects are not confounding. The calendar bias is still present in the case of annual data: in an extreme situation, comparing two years may result in an additional working day surplus, representing a difference of 0.27%. But the complex model is not fine enough to allow such an effect to significantly bias any results.
For the summer surplus, we are dealing with similar trends on a year/year basis, e.g. the summer surplus in the case of lake Balaton settlements occurs every year, so the model consistently underestimates the traffic based on population. In a similar way, such effects can be estimated by different calculations, but there is not enough input data of sufficient quality. For an accurate calculation, it would be necessary to know the additional traffic and its distribution over time for every specific settlement at the lake Balaton region. Thus, the question can be treated with reverse logic, by including a dummy variable to investigate the impact of additional traffic in these areas. Of course, there are other similar point factors that significantly influence demand, such as festivals, holidays, etc. An aggregate macro-level model over a decade is not the task for this kind of analysis, as such additive effects are mainly due to unobserved heterogeneity.

Based on the sales data recorded in the system, the exact departure and arrival point of a given trip can be determined. Due to the characteristics of the network, there are only a limited number of end points between which it is meaningful (i.e. without significant detours) to imagine the relevant existence of several alternative routes. In the case of inter-city traffic, such routes are typically Budapest-Szombathely or Budapest-Nyíregyháza. However, these situations can be clarified by specifying an additional relevant point, and in such cases the system records the relevant station for the route. The role of the route is not only relevant for the length of the journey, as the data on quality may vary from one specific section to another.

After reviewing the general issues raised by the use of the sales data, the structure of the relational database is presented. The level of detail of the data sources has changed over time: while monthly aggregates are available for previous years, the post-2016 system provides daily aggregates. The descriptive metadata of the database are presented in Table 22. The rows of the database contain the volume sold for a given (1) date (depending on the time period, a given month or day) for a given (2) destination pair (i.e. between two stations) for a given (3) discount category for a given (4) ticket type, where ticket types are return and season ticket forms, which are therefore representing the travel frequency.

Table 22. Characteristics of the demand database

Year	Data frequency	Observations
2010	monthly	5 276 659
2011	monthly	5 884 717
2012	monthly	4 939 246
2013	monthly	5 821 435
2014	monthly	6 347 629
2015	monthly	6 497 404
2016	daily	23 261 443
2017	daily	26 683 332
2018	daily	29 092 539
2019	daily	34 124 597

Source: Author's compilation

It is striking that the multiplier between the monthly and daily data is significantly lower than the first expected value of around 30. One reason for this is that the variable for the type of ticket can take 383 different types, which is combined with other dimensions of the ticket (e.g. car class). Thus most of the categories do not take a value every calendar day. Another important reason is that the possible combination for origin-destination pairs is extremely high (865 270 possible pairs). Of course, among these there are a significant number of pairs which only generate traffic on a few days a year or not at all (the value of the number of real pairs is relatively stable at around 65 000 origin-destination pairs per year through the decade). Accordingly, as the unit of observation is increased (e.g. from day to month), the number of origin-destination pairs involved increases. The variables from the database used for this research are summarised in Table 23.

Table 23. Variables used in the demand database

Data type	Unit/content
Date	Date of ticket sales
Departure station	Name of station/stop
Arrival station	Name of station/stop
Distance	Km
Ticket type	Discounts, ticket types
Volume of ticket sales	Unit
Calculated passenger headcount	Nr of passengers

Source: Author's compilation

For each record, the localised data can be used to determine the specific path of the travel event associated with the ticket, which is an important input to the basic gravity model variables. Station (stop) data do not coincide with settlement data, as many stops are located between two municipalities, and in the case of larger cities, there are often several stops or stations. Therefore, it is not feasible to establish a clear correspondence between the two categories of data, but each departure or arrival location can be assigned to one or more municipalities, and thus linking with the municipal level databases is feasible. This task is already part of the interconnection, so further details will be presented in the section on interconnection.

Municipal background variables

The basic logic of the gravity model is based on the characteristics of a specific geographical area or settlement, so in this case the settlements are the basic unit of observation. The source of the background variables in the research is the spatial statistical database of the KSH (Central Statistical Office of Hungary), the T-STAR system (MTA KRTK, 2021). This database adds a large number of annual background variables to high-quality census data. The set of variables relevant to the research includes both spatial economic development and data suitable for testing substitution. The data fully cover the time series under study (2010-2019). The range of data used is presented in Table 24.

Table 24. The municipal background variables

T-STAR Variable	Definition	Measurement unit
de01	Population in the middle of the year (extrapolated from census data)	population
egylado	Taxable income per resident	HUF/capita
szgksur	Car density	car/1000 ppl
mnado	Unemployed/taxpayers ratio	%
mnkr	Regional unemployment rate	%

Source: Author's compilation

Data for the population are from the T-STAR database. The primary source of data is the decennial complete census dataset, in this case the 2011 census, and for subsequent years, calculation based natural population statistics (live births, deaths) and migration data. Population data refer to the resident population according to the methodology of the KSH in line with international recommendations.

With regard to economic background, several data types may be suitable for the role expected in the model, since in general economic development, the ability to pay, may be related to many factors and may correlate them. In the case of transport demand, some factors may also be strongly related to the variables in the model, and a comparative analysis of these may also lead to relevant results.

For effective demand, income data are the most appropriate first approximation, and therefore the average of the taxable income tax base per population per municipality is used in the model.

Two additional factors that are also suitable for representing welfare position are more directly related to transport demand. In the case of economic activity, the unemployment rate has an impact on the ability to pay, but also carries a more direct effect, the transport demand from commuting to work.

Another such direct dataset could be the number of vehicles in operation in the area. In many ways, this indicator is a good proxy for economic development – and in Hungary, because of the geographical and settlement structure, there are no specific differences between areas that would distort the model by some other external effect (such as high mountain areas, wetlands, islands, difficult access), nor is there a lower proportion of car use in large cities. It can therefore be assumed that car ownership is generally evenly distributed spatially and that differences are mainly due to economic development.

However, another aspect that complicates the use of car ownership in the estimation is that this variable is strongly related to travel choices, namely that those who own a car are much more likely not to choose rail, and those who do not own a car are more likely to choose it. Thus, depending on the size of the car population in a given municipality, the level of demand may vary. The combined result of the two effects is also depending on whether rail travel is an inferior or a normal good for consumers.

3.3.2 Access data

Data on substitution

In terms of substitution, the model should be supplemented with two basic, real substitution paths, since private motorised transport and public road transport are primarily alternatives to rail transport.

This alternative is represented by several data sources. For the private motorised mode, in line with the literature presented, the fuel price is a suitable tool for estimating demand effects. The source of the data series is the KSH Stadat system, where annual average consumer prices are available (KSH, 2021). The average price of petrol and diesel fuels recorded here is used in the model to represent substitution through vehicle use in this sector.

The other, more general data representing these two substitution modes is road speed. In the absence of a timetable database for the public road transport alternative, as already discussed in the subsection on the concept, both the individual and public road alternative is represented by road travel times in the model. In both cases, the ratio of average individual motorised travel (typically car) time to rail travel time is suitable to represent the value of the alternative. Of course, there are a number of other factors that influence substitution, but these are similar for a given destination, e.g. the convenience of a car or the cost of a bus trip does not differ. On the other hand, the location, condition and utilisation of roads and rail networks vary considerably from one destination to another. Thus, relative differences can be used in the model to show the relative position of substitution options.

In reviewing the possible data sources, there is no available database that provides access data for the total of 300 638 origin-destination pairs studied for the ten-year period. Ideally, such a database would contain annual access data using the same methodology. Therefore, the model includes current data for 2021. The source is the background database behind the highly advanced, high-reliability travel search engine of the Google Maps Platform. This can be accessed via the Distance Matrix API (DM API) using a python query (Google, 2021a). A limitation of the Google DM API is that for public transport it is not possible to guarantee with absolute certainty which services are included in a given journey time. Preferences (bus/rail) can be set in the search, but the search engine does not necessarily follow these setting during the planning if it does not find suitable connections, and neither does it provide any information about this in the result files. The database includes rail access, in many cases involving municipalities that do not have or are not suitable for bus service in the given direction (a common example is the left bank of the Danube in the Danube bend, where bus service parallel to rail has practically disappeared since the introduction of the clock-face scheduling). Therefore the road car and rail public access modes are the ones that are certainly suitable for use in the model in the case of Google DM API data source.

The explanation for this is trivial for car travel. In the case of rail, we can build on the data source because serving rail origin-destination pairs appears in the database. Of course, there may be cases where the system also calculates the use of bus. The problems encountered are summarised in Table 25.

Table 25. Mixed community routes problems in Google DM API data

Travel time difference (coach/rail)	Which mode has a better journey time?	Result
Minimal	–	Incorrect design does not cause significant distortion
Significant	Coach	Not a rational option for paying rail passengers – presumably a minority
Significant	Rail	The Google DM API chooses the railway, so this case cannot be realistic

Source: Author's compilation

The Google DM API database can only be queried in real-time, therefore the data shows the status in July 2021, the query includes Wednesday daily hits due to the risk of distortion caused by the individual timetable offer at the beginning and end of the week.

An analysis of particular relevance for travel and access data is provided by the GEO database of the MTA KRTK, which contains public transport data for 2014, produced in 2013-2015, for a total of 45555 inter-district travel trips between census districts, recording distances, times and costs for both car and public transport. The database does not contain separate bus and rail data, but does record the mode of transport used for the journey included in the database. The database uses the population-weighted centroid of the enumeration area to determine the geographical location of the enumeration areas. The journey data:

- road journey time: by car, in minutes
- road travel distance: by car, in metres
- public transport journey time: time taken by public transport between the departure and arrival points (in minutes), including walking distance, transit, transfer and walking to destination
- distance to public transport, in metres

- means of public transport used: the variable shows the number of times the passenger used each means of transport to and from the two points

The road transport database was compiled as follows:

- in all cases, the route planning was carried out between the nearest road sections accessible by car
- where the distance between the location and the nearest road accessible by car is less than 100 m, the walking distance is included in the car journey time
- if the distance walked is greater than 100 m, the time was estimated for the length of the walk using an empirical multiplier of 1.4 and a speed of 4 km/h
- the same logic was used for the arrival locations
- the road travel time matrix was estimated between all locations and the one with the shortest travel time was entered into the database
- the achievable speeds on the road sections were generated depending on the nature and technical characteristics of the road and the traffic conditions at the time of the query

The public transport database was compiled as follows:

- timetable: exact timetable data for local and interurban services
- coordinates: geographical coordinates of all stops and stations where the given mean of transport stops
- timetable data: timetable for a Wednesday, January 2014
- services between 5 a.m. and 7.50 a.m.
- the most optimal mode was included in the model after generating all existing combinations
- maximum total travel time of 2.5 hours or 70 km, with no distance between neighbouring counties

The MTA GEO database is a data source of outstanding quality and detail, however, as can be seen from the description above, it focuses specifically on shorter distance, commuting trips, and is limited to inter-county data for public transport times. For this reason, there are a significant number of origin-destination pairs for which there is no recorded travel data, the details of which are summarised in Table 26. For public transport journeys, only 53.2% of the records are available, of which only half of the records have information on the journeys made, so for the whole database this proportion is only 26.2%. In this peculiar case only the road journeys part of the MTA GEO database are suitable for use in the model, the remaining information being primarily used for individual tests and data quality checks.

Table 26. Access data of the MTA GEO database compared to the total sample

Variable	Description	Obs.	Share in total database
kozutmeter	Distance by road	886 307	100.0%
kozutido	Travel time by road	886 307	100.0%
tkuthosz	Distance by public transport	471 903	53.2%
teljesido	Travel time by public transport	471 903	53.2%
x_a_b_v_h_f	Descripiton of public transport modes	232 204	26.2%

Source: Author's compilation

The use of the two access databases raises several questions about their relationship. Are the data consistent with each other, do they use similar optimisation to calculate the route plans? If so, is it possible to use the data from the two time points 2014 and 2021 to represent changes over time in order to show the impact of infrastructure development on road access? In terms of data consistency, the proportion of variables with the same data content is presented in Table 27.

Table 27. Comparison of GDM API and MTA GEO database data

Parameter	GDM API / MTA GEO	
	road distance ratio	road journey time ratio
Average	101.5 %	90.1 %
Standard deviation	0.49	0.28
Minimum	0.0 %	0.0 %
10% value	93.1 %	79.9 %
25% value	98.0 %	84.8 %
50% value	99.9 %	89.4 %
75% value	102.1 %	93.3 %
90% value	107.8 %	97.6 %
Maximum	10 692.7 %	4 545.2 %

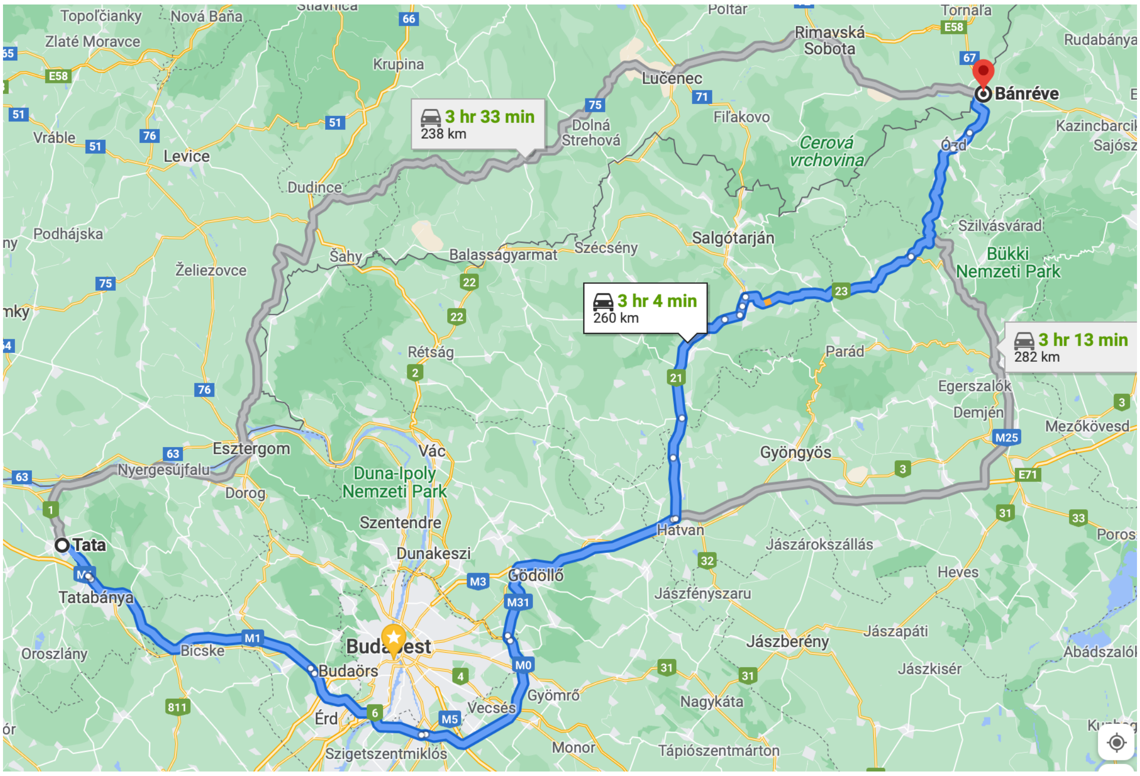
Source: Author's compilation

The first finding of the comparison is that there is a basic consistency, with 94.2% of the cases having a ratio between 0.8 and 1.2. This confirms the data quality of both data sources. However, it is questionable whether this phenomenon is caused by data errors, optimisation effects or genuine discrepancies. In the case of the time dimension, the GDM API database calculates significantly lower values. The methodological reason for this is that, as described above, the MTA GEO database also takes into account walking times for individual driving.

Further examination of the differences between variables shows that in many cases the difference is not due to road construction or other improvements, but to the effect of route choice. Figure 24 shows the route options for a specific destination pair. Looking at the example, the GDM API gives a distance of 260 km in the database, as shown in the figure, Google search returns this for an arbitrary search. In contrast, the MTA GEO database gives a distance of 237 km. It is clear that the optimisation here has taken into account the route through Slovakia in another of our cases.

Therefore, the discrepancy is not an error, nor does it include the fact that the road access has changed from 2014 to 2021, especially since the route in kilometres is longer here seven years later.

Figure 24. Route planning discrepancies in destination data



Source: (Google, 2021b)

When examining the relationship between the two databases, it is worth looking specifically at cases where the values are more divergent than realistic. In these cases, the samples examined clearly confirm the preliminary assumption: the verified data of the MTA GEO are realistic, while in the case of the GDM API data, there are cases where the geolocation identification based on the settlement name is incorrect, and incorrect location results in incorrect travel distances and time requirements. All these considerations make the MTA GEO database a better quality estimator. Accordingly, for road access, it is worth including data from the latter data source in the model.

3.3.3 Data on quality

Rail journey times

None of the databases received from the railway company contained usable data for the inclusion of rail journey times in the database. The only source available from MÁV-Start is the digital version of the official timetable books, which are no longer published in print since the 2018 timetable year (MÁV-csoport, 2017). An example of this format is shown in Figure 25. The classic rail timetable is clearly not suitable for recording journey times for a large number of origin-destination pairs (1.3 million observations in total over the ten-year period), especially as transfers are particularly complicated to find from this form of rail timetable.

Figure 25. Example of a traditional rail timetable format

A Bakonyszentlászló - Veszprém szakasz ipartörténeti műemlék. **11 Győr — Veszprém** 11

km	MÁV-START Zrt.	❄ 39510 / 39540		❄ 39512	39542	❄ 39514		❄ 39524 / 39544		39546	❄ 39556	❄ 39516		❄ 39526	39548		❄ 39558
					91					91					30		31
	Kiindulási állomás																
0	**Győr *1, 8*** 1 10	5 15	...	7 25	9 20	11 45	...	Ⓐ13 45	...	...	14 45	15 45	...	17 55	Ⓐ18 51	...	22 38
2	Győr-Gyárváros	5 18	...	7 28	9 23	11 48	...	13 48	...	...	14 48	15 48	...	17 58	18 54	...	22 41
6	Győrszabadhegy o	5 21	...	7 31	9 26	11 51	...	13 51	...	...	14 51	15 51	...	18 01	18 57	...	22 44
	Győrszabadhegy	5 22	...	7 32	9 27	11 52	...	13 52	...	...	14 52	15 52	...	18 02	18 58	...	22 45
16	Nyúl	5 33	...	7 43	9 38	12 03	...	14 03	...	...	15 03	16 03	...	18 13	19 09	...	22 56
21	Pannonhalma	5 40	...	7 50	9 47	12 10	...	14 10	...	...	15 10	16 10	...	18 20	19 16	...	23 03
27	Tarjánpuszta	5 48	...	7 58	9 55	12 22	...	14 18	...	...	15 18	16 18	...	18 28	19 24	...	23 11
28	Győrasszonyfa	5 50	...	8 00	9 57	12 24	...	14 20	...	...	15 20	16 20	...	18 30	19 26	...	23 13
34	Veszprémvarsány	5 57	...	8 07	10 04	12 31	...	14 27	...	...	15 27	16 27	...	18 37	19 33	...	23 20
38	Bakonygyirót	6 02	...	8 12	10 09	12 36	...	14 32	...	...	15 32	16 32	...	18 42	19 38	...	23 25
41	Bakonyszentlászló o	6 06	...	8 16	10 13	12 40	...	Ⓐ14 36	...	...	15 36	16 36	...	18 46	Ⓐ19 42	...	23 29
	Bakonyszentlászló	⚒ 6 07	...	8 17	10 14	12 41	...	5 14 37	...	14 37	...	16 37	...	18 47	...	...	...
45	Vinye	6 15	...	8 25	10 22	12 49	...	14 45	...	14 45	...	16 45	...	18 55	...	...	...
50	Porva-Csesznek	6 23	...	8 32	10 29	12 56	...	14 52	...	14 52	...	16 52	...	19 02	...	...	...
58	Zirc o	6 36	...	8 45	10 42	13 09	...	15 05	...	15 05	...	17 05	...	19 15	...	...	...
	Zirc	6 38	...	8 46	...	13 10	...	15 06	...	...	...	17 07	...	19 16	...	...	...
67	Eplény	6 51	...	8 59	...	13 23	...	15 22	...	...	...	17 20	...	19 32	...	...	...
79	**Veszprém *20*** o	⚒ 7 11	...	9 19	...	13 43	...	5 15 42	...	...	...	17 40	...	19 52	...	...	...
	Végállomás																

58

91 Közlekedik: III.15 - X.23-ig ©.

Felhívjuk az utasok figyelmét, hogy a Győr - Veszprém vasútvonalon közlekedő szerelvények befogadóképessége miatt a csoportos utazásokat, valamint a csoportos kerékpárszállítást legkésőbb az utazást megelőzően 7 nappal be kell jelenteni az ertekesites@mav-start.hu címen vagy a (+36) 30 497-3412 / (+36) 30 497-3070 telefonszámon. Nagy létszámú, előre nem bejelentett csoport utazását technológiai okokra hivatkozva a MÁV-START megtagadhatja.

Source: (MÁV-START, 2018b)

Because of the problems described above, I used data retrieved from the Google DM API database presented earlier as the source of the rail journey times in the model. Following the logic of Table 25, the risk of bias due to bus travel time data is minimal. An additional risk arises from the logic of the API query: the Google system determines the locations based on their names, a method that carries the potential for incorrect calculation due to misidentified locations. In all of the cases, the railway stations are used in the search instead of the names of the settlements, which allows a significantly more accurate process, as the system knows and identifies the railway stations. Experience gained from test queries has shown that in ambiguous situations the system returns missing results rather than incorrect data. In such cases, the solution was to add a search term (e.g. 'Csitténytegy' railway station instead of 'Csitténytegy railway station'). However, the possibility of a miscalculation cannot be completely excluded.

In terms of journey times, the most accurate approach is to use a complete timetable database, taking into account the average journey times of all the relevant trains between the two settlements. A key methodological question is then whether the use of a full average or a selective average is a better solution for several service types. Obviously, the relevant data for passengers is the journey time of the Intercity services between two large cities, which are the most regular, accessible and fastest connections, as these are the services typically used by passengers due to their significantly shorter journey times. The surcharges on these services do not represent a significant additional cost, so it is not expected that this approach would cause significant bias. In the case of the Google DM API data source, this calculation is done automatically, as the route planner optimises for the shortest route, and thus, in the absence of an explicit exclusion setting, it will favour the fastest option with the fewest transfers.

Timetable data

A narrower, but very important part of the information on quality and service level is the set of data that can be retrieved from timetables. The most important determinants of the quality of rail passenger services, whatever the logic of the approach, certainly include the number and frequency of available services, the structure of the timetable, in particular the existence of a clock-face scheduling. All these are accurately reflected in the published rail timetable which is used as a source of additional quality data. For these qualitative timetable characteristics, no suitable alternative source was available, so I compiled them manually from the timetables. I recorded the number of daily train pairs and the existence and the frequency of the clock-face scheduling on a line by processing the 10-year timetables of 86 railway lines. The latter data field was recorded according to a strict definition, i.e. lines with a clear structure based on a clock-face scheduling system, but with trains randomly omitted from the regular scheduling, were not included.

By default, the frequency of service is given by the number of services per day. In a strict approach, only the services considered as described above are counted (i.e. always the fastest access mode), but this logic is not feasible for manual evaluation, as it cannot be interpreted as one piece of information for the whole line. If, for example, on a long-distance connection, the faster trains have a running time of 2.5 hours and run every 2 hours, but a slower passenger train with a running time of say 3 hours is also available between two high speed services, then this is still a good proxy and can be taken into account as an hourly offer. However, if the passenger trains are only available with a transfer, or if their running time is e.g. 4 hours, or if their departure times are not in the middle of the 2-hour range but close to the faster services, then in these cases they do not represent a real quality surplus in the offer. It is therefore partly a subjective matter to determine the quality of supply in terms of frequency in a given timetable field. In order to reduce the subjective element in the data collection, only the number of trains running along the whole line or the relevant section of the line has been included in the database.

For the supply density at Lake Balaton, the summer timetable was taken into account, as it involves a higher number of passengers, and a dummy variable for Lake Balaton was included in the estimation factors, identifying all coastal municipalities.

Similarly, the GYSEV network was an issue to be investigated in the database: the two railway companies operate joint systems, and based on mutually cleared ticket sales and joint services, it can be assumed that GYSEV data would be integrated, but it is important to test for possible specific effects. For this reason, a dummy variable representing the company's stations has been added to the model.

Punctuality

An important component of the quality of rail services is punctuality. The delay database of the railway infrastructure division of MÁV records 16 million deviation events from the timetable over a ten-year period. Despite the accuracy of the data recording, the usefulness of the resource is limited by the lack of precise information on the exact context of the recorded delay events, and inconsistencies in the data. On the one hand, in the absence of appropriate timetable data, it is not possible to accurately record the number of train delay events per station in relation to the number of exact movements (day, month). On the other hand, it is not known what delays may occur for trains or lines not included in the database. If certain lines, periods or trains are not included in the database, this can cause significant distortion. A further difficulty is that a very significant proportion of delays fall within the range of a few minutes, so the low variance and noise of a few minutes' records can mask the impact of delays, which are likely to have a greater negative impact on customer experience. In terms of reliability, it is not necessarily only the magnitude of delays, but also their unpredictability that can have a significant impact on customer experience. To test these two factors, I have defined two variables that specifically include delay events over 10 minutes and the standard deviation of delays.

Infrastructure and services

The quality factors reviewed earlier can be grouped into three logical categories according to the factors on which quality and customer experience depend:

- Infrastructure-defined factors, built-in elements, services
- Availability, status, contents of available services
- Condition, cleanliness of surfaces

The first category includes information that shows the level of infrastructure a traveller will encounter on a trip. These are elements that are the result of long-term decisions, and thus include in particular the properties of real estate and rolling stock. At one end of the theoretical scale, there are large terminals with lifts, shops, catering facilities, good transport connections and heated waiting rooms, while at the other end of the scale there are stations without a proper high platform, a rain shelter, toilets or even a timetable information board. The services a station has to offer affect the travelling experience in many ways and can be measured on a relative basis and described with factual data.

The second logical category is the availability and condition of the services, as the infrastructure that is supposed to be available is often far from being functional. This is a decades-old problem on the Hungarian rail network, with most passengers having experiences ranging from unheated waiting rooms to non-functioning lifts or closed toilets, and the services that are installed but not functioning are actually more unpleasant from a customer experience perspective than those that are not available at all. Likewise, information services can be of many different types and often the information that is supposed to be available is missing, a phenomenon that can be observed in all formats from timetable boards to voice announcements or electronic displays. This data can often vary over time and is more difficult to measure, and regular on-site monitoring is essential to produce a good quality data source.

The third category is the cleanliness and condition of surfaces. These are closely related to the previous factor, but logically worth distinguishing, as in the case of the rail network this is typically a sensitive issue that often has a very negative impact on the customer experience. An established service that is available and working can also be an extremely unpleasant experience if not properly cleaned and operated. Typical of station services, this is the case from waiting rooms to toilets. The lack of regular cleaning means that even the windows of vehicles are often in such a state that looking out of them is not a positive experience. This type of data also requires a database based on regular monitoring.
This is an elusive area in terms of accessible data. There is no database that can provide a comprehensive picture of the services that are supposed to be available in the stations, and the data sources available in the group's own systems are incomplete and inaccessible. However, there are public, accessible basic sources of information that could be used to create a database of stations and stops on the Hungarian rail network, which could be classified according to a list of services on a simpler scale. The most important such source is the railway station database vasutallomasok.hu, which is edited by a groups of civil volunteers and nowadays contains a large amount of information, and in most cases could form the basis of such a database with photo documentation and basic information. This portal also provides a summary, typically over a longer period of time, of changes in places where substantial modifications and improvements are taking place. An excellent example of this is the somewhat peculiar case of the Balatonfőkajár felső stop, which was often mentioned in the press, and the growing scandal led to the unexpected renovation (or rather demolition and rebuilding) of the otherwise still non-functional building at the stop (Index.hu, 2015). If you visit the profile of this stop on vasutallomasok.hu, the full story is displayed in the comments, so you can follow and reconstruct the process (vasutallomasok.hu, 2011). However, the processing of this source would be significantly beyond the scope of the present analysis due to the large number of locations (varying depending on the period under study, more than 1200 stations and stops).

In the model, data on this segment of quality has been examined through the MÁV-Start Quality Status Assessment (MÁF) system. The Public Service Contract of MÁV-Start for the period 2010-12 (which was prepared in the period 2008-09) included a detailed quality assurance regulation. Its intended role was to monitor quality and service level performance on the basis of regular measurements, and to ensure that the consequences of under-performance or better-than-expected performance would have a financial impact on MÁV-Start's Community funding through a bonus/malus system based on this measurement system.

The evaluation methodology was included in the contract with a sophisticated content, and although the practical application of the bonus/malus was not implemented at the end, a monitoring system was nevertheless set up within the railway company. The Quality Status Assessment System (MÁF), which was the result of this process, was launched in February 2010, but its data were not used by the company in the first years and this element was not integrated into its operation. The data source is nevertheless available.

The source of the data is an evaluation dataset resulting from a regular and well-structured audit activity carried out incognito by several persons. Inspections are carried out on the entire rail network according to a schedule (which of course is not known to the operators of the services under investigation), which is based on various considerations. The inspectors assess the service quality and cleanliness of a station, a stop or a train (vehicle) according to a very complex, predefined structure and a set of criteria.

The MÁF system recorded 52 945 measurements in the first period, 2010-2013, with a further 13 959 observations in the two years after 2018. In total, the system has carried out almost 70 000 full service quality assessment measurements on the network of MÁV, each time assessing and evaluating the condition and cleanliness of 50 to 150 services, premises and vehicles at a given location on a 3-degree scale, according to the quality characteristics required by the Public Service Contract. The available measurement data are summarised in Table 28.

Table 28. Summary of the measurement data of the MÁF system

Year	Type of measurement	Obs.	Average number of evaluation criteria
2010	station	2430	90.6
2011	station	3445	91.4
2012	station	3717	85.1
2013	station	2887	86.7
2018	station	3640	100.4
2019	station	3715	111.7
2020	station	2815	113.2
2010	train	8239	146.7
2011	train	11142	141.0
2012	train	10735	149.2
2013	train	10350	145.2
2018	train	1601*	55.1
2020	train	2188**	57.1

* data for one month

** data for two months

Source: Author's compilation

In the system, each item is given a separate measurement value (0-1-2 points), from which a fixed weighting is applied to give a value for the quality of service at the higher aggregation level. The evaluation system has been taken forward in the Public Service Contract 2013-2023, concluded on 15 November 2013 and still in force, which sets out the evaluation methodology in Annex 8 (MÁV-csoport, 2021a).

The criteria are assessed against the following main categories:

– Station - Static passenger information
– Station - Cleanliness
– Station - Passenger service
– Station - Passenger information

The following areas of measurement are distinguished at the stations:

– Rain shelter
– Walkways, stairs, under- and overpasses
– Toilettes
– Rail in contact with the platform
– Platforms
– Ticket offices
– Cash hall
– Passenger hall
– Track end
– Waiting room (1st and 2nd class)

For each location and criterion, the system evaluates the cleanliness criteria that are valid for the given location. These criteria are summarised in Table 29.

Table 29. Quality evaluation criteria of the Public Service Contract

Measurement question	Measurement question
Cleanliness of windows inside and outside	Existence of benches/seats
Cleanliness of doors/handles	Cleanliness of benches/seats
Access	Cleanliness of pavement/sidewalk
Risk of tripping	Cleanliness of ceiling
Drift limit (safety barrier)	Technical condition of cash desk (reviewer)
Condition of heating/air conditioning	Cash desk capacity (queuing)
Exterior/interior cleanliness of walls/building	Cashier service behaviour
Speaker announcements – connectivity	START notices (passenger rights)
Speaker announcements — destination	Passenger information – availability
Speaker announcements – departure time	Passenger information – operational
Speaker announcements – train type	Passenger information – computer aided
Speaker announcements – track number	Technical condition of lighting
Speaker announcements – arrival time	Visual – connection options
Operation and existence of audio passenger information	Visual – destination
Opening hours of premises	Visual – departure time
Announcement – Wall timetable	Visual – train type
Announcement – train timetable	Visual – track number
Announcement – arriving/departing trains	Visual – arrival time
Possibility of drinking water	Visual passenger information functionality
Signs, inscriptions, cleanliness of clocks	WC access
Availability of trash bins	Toilet bowl cleanliness
Sufficiency, cleanliness of the pit	Lack of toilet facilities
Lamp cleanliness	Toilet facilities (hand towels)
Can you find the instruction manual (A3)?	Toilet facilities (paper)
Cleanliness of washbasin	Toilet functionality
Condition of shutters	Presence of station pictograms

Source: Author’s compilation based on MÁV MÁF system data

The evaluation tools are set out in Appendix 3 of Annex 8 to the Public Service Contract, an example of which (the vehicle evaluation template) is shown in Figure 26.

Figure 26. Train and vehicle quality assessment form in the public service contract

	VONAT ÉS JÁRMŰ MÉRÉSI SABLON **A MÁV-START/GYSEV FEDÉLZETI SZOLGÁLTATÁSOK SZOLGÁLTATÁSFELÜGYELETI ELLENŐRZÉSÉRE**									Gy...../............../2.......
Mérés ideje:	20.....	hó	nap	:	Induló állomás:			Érkező/végállomás		
Vonatszám:		Szegmens:			Ellenőrzést végezte:		név (aláírás)		beosztás	
Vonatösszeállítás, menetrendszerűség										
SZVÖR szerinti		Eltérés SzVÖR-től:							Előfűtés/hűtés:	Igen / Nem
Állomás:		Menetrend szerinti indulás	:		Tényleges indulás:	:	Menetrend szerinti érkezés	:	Tényleges érkezés:	:
Állomás:		Menetrend szerinti indulás	:		Tényleges indulás:	:	Menetrend szerinti érkezés	:	Tényleges érkezés:	:
Állomás:		Menetrend szerinti indulás	:		Tényleges indulás:	:	Menetrend szerinti érkezés	:	Tényleges érkezés:	:
Állomás:		Menetrend szerinti indulás	:		Tényleges indulás:	:	Menetrend szerinti érkezés	:	Tényleges érkezés:	:

A szolgáltatás műszaki tartalmának minősítése

	Külső			Előtér, Utastér																							
Kocsi sorozat- és pályaszáma	Külső tisztaság	Viszonylat-megjelölés	Kocsiszám-megjelölés	Feltáró ajtó működése	Átjáró ajtó működése	Előtér tisztasága	Előtér világítás	Ajtók, kilincsek tisztasága	Utastér-, fülke ajtó működése	Ablakok tiszt. kívül-belül	Oldalfalak tisztasága	Plafon tisztasága	Utastér világítás	Lámpák tisztasága	Csomagtartó tisztasága	Kapaszkodó tisztasága	Ülések, üléshuzatok megléte	Ülés tisztasága	Asztalok megléte	Szemét tartók megléte	Szeméttárolók tisztasága	Jelek, feliratok, órák tiszt.	Padló tisztasága	Hűtés-fűtés	Hangos utastájékoztató	Fedélzeti kijelző	Piktogramok megléte

	WC															
Kocsi sorozat- és pályaszáma	Bejutás, ajtó zárhatósága	WC világítás	Padló tiszt.	Oldalfalak tiszt.	WC ülőke megléte	WC ülőke tisztasága	WC csésze megléte	WC csésze tisztasága	Mosdó kagyló tisztasága	Tükör megléte	Tükör tisztasága	Szappanfeltöltés	WC-papír feltöltés	Mosdó vízellátás	WC vízöblítés	Megjegyzés

Operatív intézkedések:

Source: Author's compilation based on (MÁV-csoport, 2021a)

Vehicles

Vehicles are an essential determinant of service quality. In this context, the vehicle allocation and train composition data source received from the MÁV Group contains information in an inconsistent and varied form, often in text file format, by line group and by year. The processing of this database would go far beyond the scope of this research. Therefore, as a realistic solution to represent this aspect, I have used the data on the availability of Stadler EMU trains that have been continuously used on the MÁV network since 2006. These trains arrived in several waves and their presence can be well linked to particular railway lines. The relevant data are structured in the Hungarian wikipedia page of the Stadler Flirt type EMU train (wikipedia.org, 2021c). On this basis, I have created a dummy of EMU trains by year, which indicates the lines served by Stadler vehicles.

This type is by far the best representative of modern rolling stock, which is appreciated by passengers, and their maintenance and operation done in a different way, so they represent a different level of cleanliness. According to the data of the MÁF system, the data on quality rating of trains, despite the low variance, show this difference even though there these vehicles are not exclusive on any line. The data referred to are shown in Table 30.

Table 30. Availability and average cleanliness of Stadler EMU trainsets

Line group	Average cleanliness	Std. Dev.	Obs.
All lines	1.79	0.081	855 462
Lines without EMUs	1.77	0.089	475 789
Lines with EMUs	1.82	0.058	379 673

Source: Author's compilation

The remaining vehicle fleet is extremely diverse, with many possible criteria and evaluation factors, but without adequate evaluation data for each vehicle, it is not realistic to use them. Moreover, for older, much used vehicles, the assessment is very subjective and age does not necessarily correlate with the valuation of the vehicles. Thus, I can examine this aspect using the most representative new vehicles.

3.4 Building the model

A key problem when building the model was the extensive nature of the data sources, the uncleaned state and the missing connectivity options. The lack of a complex timetable database not only caused difficulties because of the missing data content in the model, but also difficulties in linking data sources. As the construction of the database alone would have been well beyond the realistic scope of the present work, it was necessary to find ways of simplifying the process in order to produce a high quality, analysable database that could be produced in a realistic timeframe. In the following section, I summarise the steps of the data building process that are relevant for the evaluation of the results and the conclusions. The data sources and the structure of the database are presented in Table 31.

Table 31. Compilation of the demand data – structure, sources and links

Source	Data content	Identical identifiers	Primary connection	Secondary connection
MÁV sales data	Sales quantities	Orig./dest. pair	Orig./dest. pair + Year	–
Google Distance Matrix API	Distance and time, road and rail	Orig./dest. pair	Orig./dest. pair	–
MTA KRTK GEO	Travel distance and time by road	Orig./dest. pair	Orig./dest. pair	
KSH T-STAR	Spatial data (demography, economy)	Settlement	Settlement + Year	–
KSH petrol price series	Petrol prices	Year	Year	–
MÁV delay database	Delay data	Date+ station + train nr.	Station + Year	Railway line + Year
MÁV MÁF system	Quality data (cleanliness, availability of information)	Date+ station + train nr.	Station + Year	Railway line + Year
Supply data based on timetable	Daily train numbers, regular timetable	Railway line	Railway line + Year	–
EMU trainset dummy	Electric Motor Units available	Railway line	Railway line + Year	–
GYSEV dummy	GYSEV Stations	Station	Station	–
Balaton dummy	Satitons at Lake Balaton	Station	Station	–

Source: Author's compilation

The primary principle in compiling sales data was, of course, to keep the related data at the lowest possible level of aggregation and to achieve the most accurate linkage possible.
The initial database contains inter-station origin-destination pairs over a ten-year period. However, some of the data are only available at the level of settlements. The final versions is a gravity model based on population data, so it was necessary to transform the station-station logic to a settlement-settlement aggregated model structure. Accordingly, the data that could be fitted at station level was assigned in the station state of the model, while the data organised at the settlement level was assigned in the final structure. Data sources that

could be organised in other ways were included in the database at the most accurate point of connection. In some cases, the linkage did not involve unique pairing, these problems could even cause data corruption; in other situations, more significant risks were encountered.

Unambiguous mapping was achieved for the Google DM API, the MTA KRTK GEO location databases, as well as at the municipal level for the KSH T-STAR. Due to the municipality logic, the whole settlement is included in the model, in contrast to the usual practice in Budapest, which relies on district level data, and this may cause inconsistencies in the travel time values due to the spatial extent. For the rail data, logical starting stations are generally included according to the direction of travel, and it is important to note that in the case of Budapest, I have assigned Google DM API data to origin-destination pairs at the station level. For simpler data sources the fit is 100%, such as fuel prices, GYSEV and Balaton dummy variables.

The most problematic linkages were with data from the MÁV systems. The data are often not identified at station level or do not even contain such variables, the solution was to link them to railway lines. However, the categorisation of railway lines is not straightforward for a pair of destinations, especially for hubs. For stations with more than one line, a multi-stage approach was used. When an intersection was obtained by comparing the trip with another endpoint, the common rail line was included in both endpoints. If no such line could be identified, the line with the highest traffic volume was used, or the line with the lowest line ID in the case of lacking data (the latter calculation also approximates the traffic volume, but with lower efficiency).

For the majority of the delay database, the station level data recording allows for a good linkage to the stations. Where no station data were available in a given year, the average annual delay value associated with the railway line was entered into the model.

The situation is similar for the station data of the MÁF system, where the data are also included in the model by station linkage, with line-level averages in the case of data gaps. The measurement data for the trains were averaged by line and year based on the line designation in the model and linked this way to the model. In the case of the MÁF system, the measurement values for the missing years were filled with the station averages for the other years.

The EMU trainset dummy contains information for the year and line, so this data source was incorporated into the model by linking it to these two variables.

4 RESULTS

In the following I have used a stepwise approach to build the model. I first set up the gravity model, then added the economic background variables, the substitution data and finally the quality variables in several steps. The results are presented in this structure below.

4.1 The basic model

As a first test of the compiled database, it is worth examining the explanatory power of the basic gravity model. Given that the basic assumptions of the gravity model are best met by the log-log regression, since we are examining the impact of different factors on demand that are not comparable on an absolute scale, the natural logistic regression, is a reasonable choice, which approximates elasticity well (Békés and Kézdi, 2021). Accordingly, I use the natural logarithm of the variables (except for the dummy variables) in the calculations. To test and compare the variables, the database allows basically two approaches. In a purely cross-sectional analysis, the time factor is ignored (in this case, years are only used to link the data) and each row of the database is considered as an independent, stand-alone observation. The results of the cross-sectional version of the basic model are presented in Table 32.

Table 32. The basic model with cross-sectional estimation

Variable	Coefficient		S.E.
$ln\_Population_{dest.; arr.}$	0.914	***	(0.00151)
$ln\_Distance_{rail}$	– 1.362	***	(0.00202)
Constant	0.330	***	(0.0141)
Obs.	888 185		–
R^2	0.4072		–

*** p<0.01

Source: Author's compilation

The baseline model explains a significant part of the heterogeneity as expected, the direction of coefficients are corresponding to the intuition in the case of population and distance between settlements (in this case the distance according to the rail ticket purchased). The population variable is itself the product of the population of the two settlements in the origion-destination pair, so that the two values appear in one variable. The ratio of the effects of population and distance also partially confirms the assumption of the gravity model, where the square of the distance appears in the physics formula. The basic model thus confirms the concept that factors affecting demand can be examined within this framework.

Before moving on to the next variables, we need to return to the question of the estimation method. A definite advantage of the cross-sectional approach is the very large number of cases: around 890,000 observations (this value decreases slightly as the number of explanatory variables increases, as the value of some variables is not known for all observations). With this method we can basically compare the effects of individual variables and their relationship to each other. However, there is a fundamental problem of endogeneity in this case: effects that are persistent but not observed for a single observation (origin-destinations pairs) are omitted from the analysis, distorting the results by appearing to be the effect of other, observed factors. Panel analysis provides a way to reduce this error. In the panel, we identify the observations (in our case, origin-destination pairs) and the time factor (in our case, 10 years of data are included), and the model interprets the data along these two dimensions. The calculation thus does not consider the data of the same destination pair observed over several years as independent observations, but bases the estimation on the differences between the values of these observations in different years. By repeatedly observing more and more values of exactly the same cases or origin-destination pairs over a longer period, it is possible to look for differences within these groups by examining the effects of explanatory variables.

The primary problem with the basic model is therefore the presence of unobserved effects and their possible bias. These difficulties could be particularly significant for a complex issue such as the evolution of demand for transport services. Moreover, in many respects, the range and quality of the data is incomplete, so even compared to the assumptions of the basic model, the problems of data scarcity and data quality can cause uncertainties. It is therefore particularly important to exploit the potential of the panel database.

Panel estimation requires a choice between random effect and fixed effect estimation. The random effect assumes that the omitted variables are independent of the explanatory variables. This assumption is difficult to defend in a case where a large number of explanatory variables covering many domains are planned to be used in the model. If we consider factors that are in principle independent of everything but are clearly determinants, such as consumers' subjective taste preferences or their habits in previous periods, it is easy to see that they are not independent, since they can affect, for example, the choices of the previous period just as much as they can be reflected in the effects of quality factors. To choose between a fixed or random effect, I performed the Hausman test (Balázsi *et al.*, 2014) on several variables, and in all cases the results strongly supported the rejection of the hypothesis of using a random effect estimator. Random-effects estimation, like cross-sectional testing, may be suitable to further investigate a factor that for some reason does not yield a significant result in the panel, but the fixed-effects estimation should be used in the baseline model regardless.

The risk of internal correlations is particularly high in the case of a large number of background variables representing interrelated factors. The software used will flag and automatically exclude variables that are subject to collinearity, but even so, the estimation should aim to include only one explanatory variable with the same or overlapping content. The results of the fixed effect estimation for the baseline model panel are presented in Table 33.

Table 33. The basic model with fixed-effects panel estimation

Variable	Coefficient		S.E.
$ln\_Population_{dest.;\,arr.}$	1.930	***	(0.0597)
$ln\_Distance_{rail}$	– 0.368	***	(0.0137)
Constant	– 13.92	***	(0.562)
Obs.	888 185	–	–
Origin-destination pairs	188 226	–	–
Within R^2	0.0055	–	–
Overall R^2	0.1568	–	–

*** p<0.01

Source: Author's compilation

For the dependent variable, I tested the other two versions presented in section 3.3.1, with no significant difference either with the corrected season ticket count (lower number of passengers for season tickets) or with the paying passengers (narrower definition of the scope of discounts).

Before extending the model, a very important step is to check whether some lag should be included in the estimation. The inclusion of past period consumption is absolutely justified from the point of view that it is a good proxy for several factors that otherwise can not be or are difficult to investigate, for example, the role of subjective, taste-based factors in mode choice. In addition, the choices made in previous years may in themselves partly explain the evolution of demand in a given period, due to the possible presence of habits, trip dependence and lower willingness to switch. Table 34 presents the model with the one-year lagged variable.

Table 34. Adding an annual lag variable to the base model

Variable	Coefficient		S.E.
$ln\_Population_{destination;\ arrival}$	2.240	***	(0.0629)
$ln\_Distance_{rail}$	– 0.271	***	(0.0152)
ln_Lag_var	0.254	***	(0.00206)
Constant	– 17.84	***	(0.599)
Obs.	597 240		–
Origin-destination pairs	115 460		–
Within R^2	0.0722		–
Overall R^2	0.2305		–

*** p<0.01

Source: Author's compilation

The effect of the lagged variable is significant, of a similar magnitude and significant to distance. These all support the above hypothesis, so the retention of the lag variable in the extended model is justified even if it is used at the expense of a reduction in the number of usable observations due to the lag. Next, we incorporate the economic effects on demand into the model (Table 35).

Table 35. Adding economic variables to the basic model

Variable	Coefficient		S.E.
ln_Population$_{destination; arrival}$	1.844	***	(0.0690)
ln_Distance$_{rail}$	– 0.276	***	(0.0153)
ln_Lag_var	0.255	***	(0.00207)
ln_Income	– 0.122	***	(0.0104)
ln_Unemployment	– 0.0263	***	(0.00600)
Constant	– 13.16	***	(0.679)
Obs.	597 204		–
Origin-destination pairs	115 457		–
Within R^2	0.0728		–
Overall R^2	0.2677		–

*** p<0.01

Source: Author's compilation

With this step, two new variables appear at the same time. The tax base per capita confirms the previous assumption that rail transport also has a certain inferior nature. If it appears in the model alone, it is clearly a negative effect: the higher the average wage level in a given area, the less people choose rail. The coefficient of the unemployment indicator is also negative, which is consistent with the basic hypothesis that lower incomes and lower unemployment also lead to lower levels of demand. The simultaneity of the two effects is not in itself contradictory, but their correlation means that their parallel presence should be avoided. Next, it is worth controlling for the primary means of substitution, the car.

4.2 Substitutes and other structural effects

In total, I have examined three concepts of substitutes in more detail. Here, the primary role is played by car ownership as a long-term, asset-side proxy, fuel price as a variable cost element, and travel time, which can also represent road public transport substitution. At this stage of the model, the fuel estimate does not work well – it will be re-tested in the full model, along with the variables controlling for quality. The results are summarised in Table 36.

Table 36. Results of the substitution model, fixed-effects panel

Variable	Coefficient		S.E.
ln_Population$_{destination; arrival}$	1.867	***	(0.0699)
ln_Distance$_{rail}$	– 0.277	***	(0.0153)
ln_Lag_var	0.254	***	(0.00207)
ln_Unemployment	– 0.00192		(0.00555)
ln_Car_ownership	– 0.186	***	(0.0237)
ln_Speed$_{car}$	– 0.105		(0.0958)
Constant	– 12.87	***	(0.774)
Obs.	597 081		–
Origin-destination pairs	115 441		–
Within R^2	0.0726		–
Overall R^2	0.2683		–

*** p<0.01

Source: Author's compilation

The result of the estimation is consistent with the intuition: an increase in car density will significantly reduce rail demand. Here, of course, income and asset-side effects occur simultaneously, but as we have seen earlier, they do not weaken each other in the higher ranges because of the inferiority. From this point of view, the question of the precise mechanism of action through which income and asset ownership affect demand is less relevant.

With regard to the speed, I have examined several indicators. The results for the estimation of the effects of the different variables representing the competitiveness of rail and road services were as shown in Table 37. The ratio between rail and road speed did not become a suitable indicator as an explanatory variable due to the different correlations. The values calculated from the otherwise very good quality MTA GEO data showed very poor significance for the comparison of speeds, and the absolute road speed was excluded from the model in advance. The primary reason for this may be the constant nature of the value over time, although a similar effect is also present in the GDM API data. Finally, the relatively strongest indicator was the absolute average speed calculated from the latter source.

Table 37. Comparison of estimates of substitution variables

Variable	Coefficient	S.E.	Obs.	Orig.-dest. pairs	Within R^2
Speed of car/train (based on GDM API)	– 0.0111	(0.0181)	586 930	113 503	0.073
Speed of car/train (based on MTA GEO)	– 0.00310	(0.0952)	586 930	113 503	0.073
Car travel speed (based on GDM API)	– 0.0525	(0.0181)	595 518	115 249	0.072
Car travel speed (based on MTA GEO)	–	(omitted)	595 560	115 257	0.072

Source: Author's compilation

Thus, based on the tests, the absolute speed of road travel is the best proxy for the quality of road access, but even for this variable the relationship is weaker, not significant in the version not yet controlled for further data.

Its interpretation is best understood as representing the effect of a higher average speed of a car trip, e.g. if there is a motorway road link between a given origin-destination pair. In the final model, it is worth revisiting and refining the set of proxy variables used.

Two more parameters should be examined before the quality is tested, the Balaton and GYSEV dummies. Their values are unchanged for one observation in each of the ten years, and accordingly do not work well in the fixed effect estimation, with zero variance within origin-destination pairs. Their presumed effect can therefore be assessed in the otherwise inadequate random effects estimation, the results of which are presented in Table 38. It is important to stress again that, due to the estimation procedure, these values are not relevant for the full model, as it can be seen that the parameters of several factors do not match the results of the fixed-effects estimation.

Table 38. Impacts of Balaton and GYSEV dummies, random effects estimation

Variable	Coefficient		S.E.
ln_Population$_{destination; arrival}$	0.190	***	(0.00141)
ln_Distance$_{rail}$	– 0.306	***	(0.00246)
ln_Lag_var	0.805	***	(0.000767)
ln_Unemployment	– 0.0215		(0.00306)
ln_Car_ownership	– 0.128	***	(0.0104)
ln_Speed$_{car}$	– 0.0806		(0.00938)
GYSEV	0.0470	***	(0.00546)
Balaton	0.164	***	(0.00449)
Constant	0.68	***	(0.0661)
Obs.	597 081		–
Origin-destination pairs	115 441		–
Within R^2	0.0684		–
Overall R^2	0.8409		–

*** p<0.01

Source: Author's compilation

The results show that the additional traffic at Lake Balaton has a significant positive effect in the model – which is perfectly intuitive. The dummy effect of GYSEV is low, which mostly supports the assumption that GYSEV sales data have not been omitted, but this estimate is probably not suitable for a more complex conclusion.

4.3 Quality effects

When examining quality impacts, the effects of several possible indicators for a given dimension have been compared. First, the effect of the minimum or average of each indicator has been examined: at the basis of differences between the departure and arrival stations, the question is whether the average or just the smaller value of the two indicators has a greater effect on demand. Both logic can be supported by intuitive arguments, so the variable with the stronger effect is included in the model.

To give a simple example, if a trip is split into two rail line segments, the timetable for the outbound and inbound stations is different. This bidirectional effect can be captured by taking the average of the two values, i.e. lower supply on one line is adjusted upwards by the other, and vice versa. It can also be understood that quality depends primarily on the level of lower supply, i.e. if one line has a lower frequency service, the better supply of the other line cannot correct it.

Next, it has been examined whether only the data for the outgoing station or only the incoming station showed a larger effect than the mean or minimum value? In principle, an intuitive explanation for this can easily be found, e.g. station services are assumed to have a greater effect at the departure than at the arrival point, since we stay longer and wait longer at the former, while we typically leave the station area as soon as possible at the arrival. If such an unaveraged effect is stronger than the effect at both endpoints, it will be a significant result.

A further important consideration on this last point is: does it make sense to distinguish between departure and arrival stations? As a starting point, we can assume that the distribution of rail traffic between destinations is balanced, since the traffic of a single settlement cannot be asymmetric in the long run. From this point of view, for a given origin-destination pair, the traffic flows towards each other should compensate for the asymmetric quality effects described above. This can be seen in the second column of Table 39, where the level of correlation between the annual aggregate values of the inbound and outbound traffic of a station is above 98% in all cases.

Table 39. Balance of departure and arrival stations in the database

Year	Correlation of annual aggregate values	Correlation of origin-destination pairs
2010	98.57 %	77.23 %
2011	98.57 %	77.60 %
2012	98.07 %	74.78 %
2013	98.51 %	81.25 %
2014	98.60 %	80.96 %
2015	98.30 %	79.40 %
2016	98.47 %	81.91 %
2017	98.67 %	82.98 %
2018	98.70 %	82.75 %
2019	98.70 %	82.84 %

Source: Author's compilation

However, in reality, the smoothing is not nearly so functional. On the one hand, for regular travellers (season tickets, frequent return journeys), the choice of where to buy tickets is in fact optional. In the case of return journeys, especially for longer distances, there is an interval of several days for the return trip, and for season tickets this is even less limited. So, for some part of the journeys the choice of where to buy tickets is not randomly distributed between the two end points: it is presumably not independent of the quality of the station services. In addition, there are many possible situations where rail travel is only in one direction (round trip, return by other means, etc.). The latter is illustrated in the second column of Table 39, where I have compared data for origin-destination pairs rather than aggregated values. Only 64.3% of the pairs had two-way traffic, and in the remaining cases the correlation value varied around 80% (72.1% weighted by traffic). So, in principle, the annual traffic of a settlement is balanced in both directions, but asymmetry is observed for origin-destination pairs.

Returning to the analysis of quality indicators, I will summarise the approach using the example of delays. The data always includes station and line delays (where no station delay data are available, the average of the associated rail line is used). Here I have tested three versions of three indicators in the model. The first indicator is the average of departure and arrival delays, this value represents the extent to which the service of a given origin-destination pair is prone to have delays. The standard deviation represents the uncertainty that should be expected for the same service, rather than the average delay. To approximate the quality thresholds, I form a third indicator, which represents the proportion of delays of more than ten minutes in the delay data for a given origin-destination pair. There is a sensible and logical intuitive reasoning behind all three indicators, so the comparison carries real substance. For all three indicators, I examined the departure and arrival station and the average value by location also. A summary of the results is shown in Table 40.

Table 40. Statistics of different delay indicators in the model

Variable	Indicator	Location	Coefficient	S.E.	t
ln_ind_keses	Average delay length	Departure	– 0.0035779	0.0033482	– 1.069
ln_erk_keses		Arrival	– 0.0035574	0.0033328	– 1.067
ln_keses		Average	– 0.0035149	0.0033502	– 1.049
ln_ind_keses_sd	Standard deviation of delay	Departure	– 0.0094218	0.0025003	– 3.768
ln_erk_keses_sd		Arrival	– 0.0093134	0.0025037	– 3.720
ln_keses_sd		Average	– 0.0093992	0.0025038	– 3.754
ln_ind_tkeses	Ratio of delays longer than 10 min.	Departure	– 0.0070447	0.0020126	– 3.500
ln_erk_tkeses		Arrival	– 0.0041081	0.0020837	– 1.972
ln_tkeses		Average	– 0.0063744	0.0021859	– 2.916

Source: Author's compilation

In general, the results show relatively weaker effects, we will return to this in the conclusions. Sequentially, we see that the strongest effect is for the standard deviation of delays, about three times the average delay and one and a half times the effect for delays beyond ten minutes. For the statistics by location, it is striking that departure delays tend to show the strongest effect (the differences are only outstanding for delays over ten minutes). Overall, the standard deviation of departure delays is included in the construction of the baseline model.

The strongest indicator for the number of daily services representing supply is the minimum value of the supply level at the departure and arrival station, i.e. the narrowest supply is the most significant. This parameter shows a stronger significant effect in the model, than in the case of delays, as shown in Table 41.

Table 41. Statistics on the number of daily services in the model

Variable	Indicator	Location	Coefficient	S.E.	t
ln_ind_vdb	Pairs of trains/day	Arrival	0.0614348	0.008672	7.08
ln_erk_vdb	Pairs of trains/day	Departure	0.0413869	0.009131	4.53
ln_vdb	Pairs of trains/day	Minimal	0.0869765	0.009139	9.52
ln_vdba	Pairs of trains/day	Average	0.0780829	0.011907	6.56

Source: Author's compilation

The supply side is also characterised by the data on the clock-face scheduling. Variations of this include the frequency at departure and arrival stations, as well as the intensity of the frequency (which captures the frequency of train departures). The results are presented in Table 42. The average frequency at the departure and arrival stations has a significant effect, of which the intensity is the strongest indicator.

Table 42. Statistics of the clock-face schedule in the model

Variable	Indicator	Location	Coefficient	S.E.	t
ind_itf	existence of ITF	Departure	0.024543	0.008443	2.91
erk_itf	existence of ITF	Arrival	0.023699	0.008657	2.74
itf	existence of ITF	Minimal	0.033394	0.007340	4.55
itfa	existence of ITF	Average	0.042015	0.011190	3.75
ind_itf_ints	intensity of ITF	Departure	0.087102	0.087102	5.42
erk_itf_ints	intensity of ITF	Arrival	0.136098	0.016894	8.06
itf_ints	intensity of ITF	Minimal	0.154068	0.017307	8.90
itf_intsa	intensity of ITF	Average	0.212367	0.021903	9.70

Source: Author's compilation

The statistics for the four variables that can be similarly derived from the dummy for the EMU trainsets are presented in Table 43. The most powerful is the indicator for the average supply of EMU trains, so this is included in the model.

Table 43. Statistics of EMU trainsets in the model

Variable	Indicator	Location	Coefficient	S.E.	t
ind_mot	Availability of Stadler EMUs	Departure	0.15927	0.01212	13.14
erk_mot		Arrival	0.13775	0.01218	11.31
mot		Minimal	0.16988	0.16988	15.19
mota		Average	0.25598	0.25598	16.19

Source: Author's compilation

For the statistics on the quality data of the MÁF system, the impact of a total of three values (minimum, average and weighted average) is examined, by departure station, arrival station and average of these two, as shown in Table 44. The coefficients confirm the intuitive hypothesis made earlier that the significance of the values by departure is stronger than the average quality, this is observed for all three indicators.

Table 44. Quality statistics of the MÁF system in the model

Variable	Indicator	Location	Coefficient	S.E.	t
ind_mafmin	Quality minimum	Departure	0.029592	0.022347	1.32
erk_mafmin	Quality minimum	Arrival	– 0.033981	0.022365	– 1.52
mafmin	Quality minimum	Average	0.005279	0.029892	0.18
ind_mafmean	Quality average	Departure	0.229137	0.049011	4.68
erk_mafmean	Quality average	Arrival	0.011291	0.050175	0.23
mafmean	Quality average	Average	0.193782	0.060993	3.18
ind_mafsmean	Quality weighted average	Departure	0.270162	0.055486	4.87
erk_mafsmean	Quality weighted average	Arrival	– 0.060290	0.056823	1.06
mafsmean	Quality weighted average	Average	0.178695	0.069147	2.58

Source: Author's compilation

It can also be seen that the weighted indicator has the greatest explanatory power. The significance of the weighting is due to the fact that very different subjective factors are included in the formula (this is particularly the case for the relationship between cleanliness and information). Annex 8 to the public service contract document also sets out calculation weights, so the significance of these is not considered equal (MÁV-csoport, 2021a). For the calculation, a simple logic has been applied: individual coefficients were calculated by fitting each quality factor to the random effect estimate, and these weights were then assigned to the factors in the averaging process, as presented in Table 45.

Table 45. Weights of quality statistics of the MÁF system

Variable	Location	Weight
Cleanliness	Station	0.891
Information	Station	0.492
Cleanliness	Train	0.967
Information	Train	0.199

Source: Author's compilation

4.4 Estimation results of the full model

When constructing the full model, the most significant variable has been included for every given subject to minimise the risk of internal correlation. The results of the fixed-effects estimation form the basis of the main conclusions of my analysis and are presented in Table 46. In the final equation, a total of ten explanatory variables are included in addition to the two basic variables of the gravity model. The coefficients of the log-log panel regression, which can be linearly estimated using natural logarithms, are most commonly interpreted as elasticities. This means, in the example of the lag variable, that a 10% increase in consumption in the previous year leads to a 2% increase in demand. Expressed as a percentage, we can talk about an elasticity of +21.3%.

The income situation of the two settlements concerned in the origin-destination pair has a positive impact on demand with an elasticity of 13.9%. The negative impact of the availability of cars is particularly large, at -58.4%. The increase in the price of fuel contributes to a lesser extent to the increase in rail demand at a level of 5.8%, while the increase in the average speed of car travel (for a given origin-destination pair) reduces it by 20.3%.

Table 46. Estimation results of the complex demand model

Variable	Coefficient		S.E.
$ln\_Population_{destination;\ arrival}$	1.1800	***	(0.0874)
$ln\_Distance_{rail}$	– 0.2662	***	(0.0165)
ln_Lag_var	0.2129	***	(0.0022)
ln_Income	0.1391	***	(0.0031)
ln_Car_ownership	– 0.5837	***	(0.0731)
ln_Petrol_price	0.0575	***	(0.0156)
ln_Speed_{car}	– 0.2028	**	(0.1026)
$ln\_Trains\_p\_day_{min}$	0.0354	***	(0.0089)
ITF_{mean}	0.2045	***	(0.0216)
EMU_{mean}	0.2351	***	(0.0146)
$ln\_Departure\_delay_{st.\ dev.}$	– 0.0101	***	(0.0024)
$Departure\_quailty_{weitghted\ mean}$	0.3980	***	(0.0551)
Constant	– 5.6468	***	(0.9719)
Obs.	499 590		–
Origin-destination pairs	104 442		–
Within R^2	0.0530		–
Overall R^2	0.3486		–

*** p<0.01, ** p<0.05

Source: Author’s compilation

The analysis of quality data, which was the main focus of the research, yielded substantial results. The various factors of supply quality generally have a significant and substantial impact on demand, as estimated by the model.

First, we look at the 'hard' elements of supply quality. The minimum number of train services at two stations has a positive, significant, but small effect on demand (3.5%) when measured together with the effect of the integrated clock-face scheduling (20.4%). The correlation between the two factors results in a higher supply effect when run without ITF (6.8%), and the reverse is also true (pure ITF effect: 23.1%). Similarly, we find a very strong effect of 23.5 % for the use of modern EMU trainsets on a given line.

Among the 'soft' quality factors, the model estimates a significant and intuitive negative impact on the delay results, while the -1% level is very low compared to the other quality effects. However, it is also worth noting that when the effect of an additional 10% delay is considered, this indicator predicts a -2.3% reduction using the natural logarithm calculation, which is now of considerable magnitude.

The other quality factors, which are derived from a single source, the MÁF system data, are treated as a composite weighted variable, a very strong explanatory variable with a coefficient of 39.8%. This means that a 10% increase in the average score in the MÁF system is expected to result in an increase in sales of around 4%.

It is important to highlight which indicators have the greatest impact on demand from data on quality. For the number of services, the minimum of the departure and arrival is the determining factor. For the clock-face scheduling and the effect of EMU trainsets, the average of the departure and arrival is the determining factor. For delays, the delay before departure is more of a problem, and the effect of the standard deviation of delays is more important than the length of the delay itself. For the complex quality indicator of cleanliness and information, again the role of the departure value is the dominant factor.

5 CONCLUSIONS

The aim of my research was to develop a demand model to investigate the effects of substitution and quality indicators on the Hungarian passenger rail market. The research concerns only paying passengers, so the analysis does not cover passengers over 65 years of age, railway and their family members and other free travellers. This condition is in line with public policy objectives to increase the role of rail among transportation modes.

Building on the literature and available data sources, I have succeeded in compiling a database that can adequately describe the demand for passenger rail transport in Hungary and has produced acceptable results within the framework of the literature and intuitive concept presented. This allowed a deeper analysis of certain quality and service level factors and other influences on the evolution of demand.

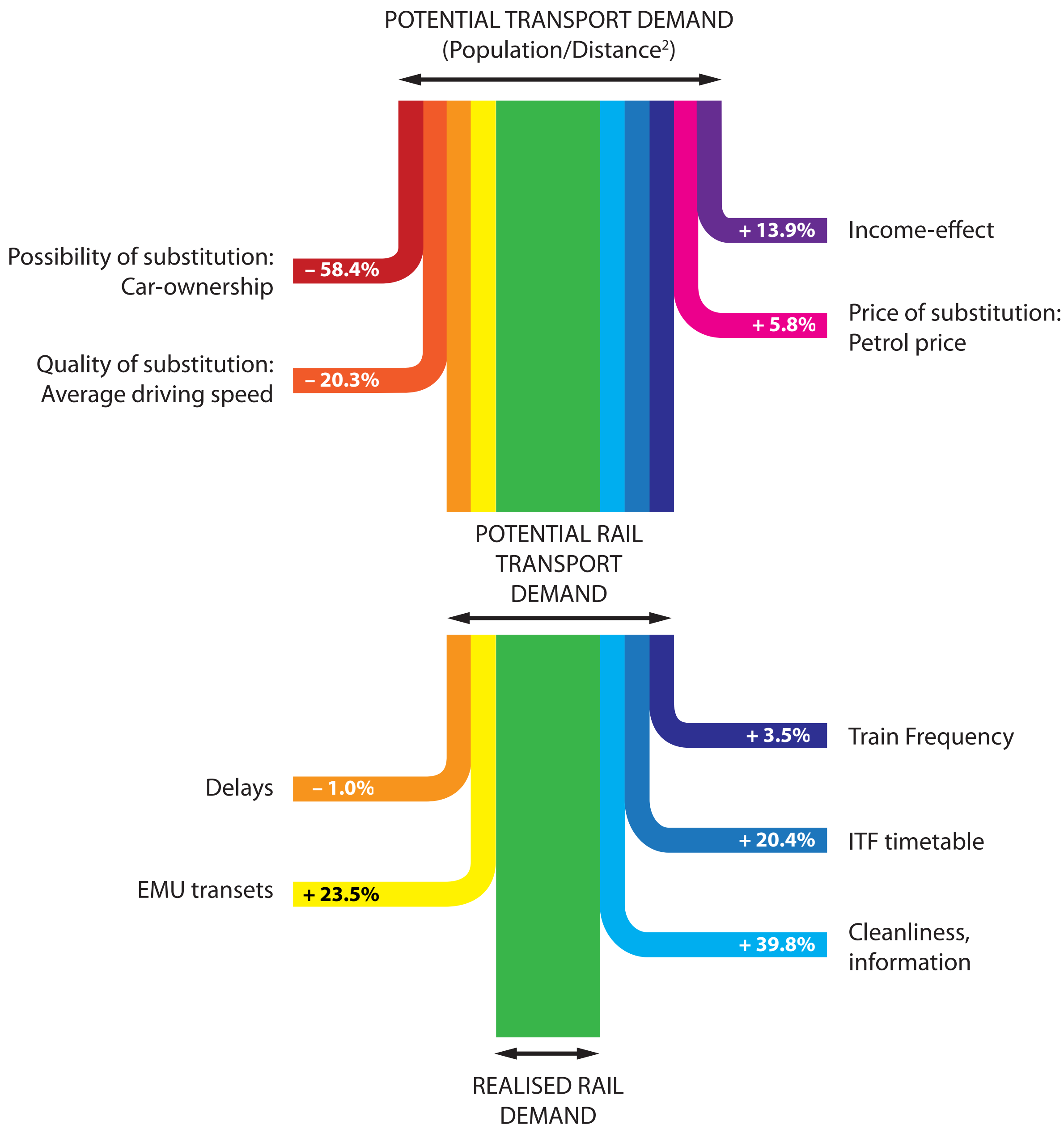
5.1 Interpretation of the results

The results of the estimation carried out in this research allow us to draw conclusions on the impact of the different demand factors and their relationship to each other. According to the estimating function, the coefficients can be interpreted as elasticities. These values are presented in Figure 26, following the structure of the previously summarised logical framework of the model (see Figure 21).

Among the demand factors, the effect of the per capita municipal income tax base is relatively smaller, 13.9%, within the framework of the basic gravity model, but this value should be seen in conjunction with the fact that it is already controlled by car ownership, so the indirect income effect through car ownership is not included in this value. It is also important to note that, without the other variables of car substitution, the income effect is negative at -10%, which clearly confirms the hypothesis that rail transport is an inferior service in Hungary today.

The model shows an indirect positive effect of the increase in fuel prices, but its low value also indicates a more inelastic response, i.e. it is more difficult to 'stop' using cars, while the other elasticities of substitution show significantly higher values.

Figure 27. Results of the estimation in integrated logical framework



Source: Author's compilation

The strong effect of a lagged demand variable suggests that mode-switching is inelastic: ceteris paribus more people are more likely to travel between two cities, if the proportion of rail passengers was higher in the earlier period.

Together, these effects can be interpreted so, that it is harder to increase the share of rail, because it is 'harder to enter' this mode: car ownership reduces demand very much, the

effect of fuel price increase is smaller, and the historical demand – the share of people already travelling by rail – is an important factor.
In the case of travel times, only the speed of car travel showed a significant effect, which can be interpreted as a stronger role of car substitution in the range of higher average speeds. This effect can best be interpreted as the role of motorways and expressways.
Other approximations of the time dimension did not show a strong effect in the model – this supports the hypothesis that a large proportion of today's passengers are less sensitive to travel time, which may be due to individual preferences (preferring rail for subjective reasons) or to constraint (no realistic substitution option). A more accurate analysis of this factor would be greatly assisted by the generation of travel data that is representative of changes over a ten-year period, and the variability that this would introduce would allow the impact of this factor to be measured much more accurately in the panel estimate.
Consistent with the above hypothesis about time, the weight for delays was also relatively low, with coefficients never exceeding 1%. Here, although the quality of the database is presumably good, the assumptions made in the construction of the station averages, as presented earlier, may lead to different results with a more accurate and better connected database. Nevertheless, it is clear that the relative weight of this effect is smaller for other quality indicators. The different calculation methods examined show that the standard deviation of delays is more important than their absolute average. Estimating the proportion of delays of more than ten minutes resulted in coefficients that approximate the standard deviation and are much more important than the average. The effect of departure delays is clearly the dominant factor.
For the timetable supply, the effect of the minimum departure and arrival frequencies (number of train pairs per day) is stronger than the average. Thus, in this dimension, the minimum level determines the value of the service for a passenger. To give a simple example: a passenger has little benefit from a 30-minute frequency if the second section of the route has trains departing every 3 hours. This latter scarcity of supply also reduces the value of the first section.
The variable for the cycle frequency of the timetable represents an additional value, as it appears in the model alongside the number of services. Clock-face scheduling has a noticeable additional benefit, and here it is less important whether it covers the whole travel chain, with the average value of the origin and destination having a stronger effect than the

minimum. Similarly, for the effect of EMU trainsets, the average is the dominant value, and the effect itself is the second strongest among the quality variables after cleanliness.

The estimation of the soft quality factors – cleanliness and information – is based on a database with a very high frequency, a fixed methodology, but a relatively low variance scoring system. Nevertheless, the effects found are significant and substantial. The hypothesis concerning the effect of the minimum quality is not confirmed by the results. Among the quality effects, it is striking that the effect of the service quality of the outgoing station is much higher than that of the arrival station. The most significant version of this indicator is a the weighted average of several factors. The sources of the weights were the coefficients of the different quality indicators estimated in a random effects model.

To summarise, the minimum of the departure and arrival data is therefore the determinant for the number of services, and the average of the two for the frequency and EMU trainset effects. For delays, the delay before departure is more of a problem, and the effect of the standard deviation of delays is more important. For the complex quality indicator of cleanliness and information, again the role of the departure value is the dominant factor.

5.2 Evaluation of the research questions

Based on the results, I got the following answers to the research questions I had previously asked:

1. Can an appropriate explanatory model be developed for rail sales data based on the available data sources?

The complex model resulting from connecting of the databases processed has sufficient explanatory power. By including additional data sources, improving the data quality or the level of detail of a particular area, this analytical framework can be used to explore certain sub-questions in more depth, but the model even at this level is already capable of drawing conclusions.

2. What are the impacts of substitutes, economic background variables and quality factors on rail demand?

Quality factors have a significant impact on demand when the data sources are fine scaled and of good quality, and their effect is confirmed by the model. For a given quality factor, it makes sense to use specific sub-indicators or calculated values (average of different values in some cases, minimum in others, etc.).

3. Is there a relevant relationship between the quality factors?

For the quality dimensions of rail supply, the minimum service level wasn't observed stronger than the average, so this general hypothesis was not confirmed. However, such effects can be observed within some of the quality factors, so it is worth using specific sub-indicators or derived values in the estimation, which are summarised in Table 47.

Table 47. Comparison of the characteristics of quality impacts

Variable	Location	Statistics
Income	Departure and arrival settlement	Average
Car-ownership	Departure and arrival settlement	Average
Number of services	Departure and arrival station	Minimum
ITF schedule	Departure and arrival station	Average
EMU trainset	Departure and arrival station	Average
Delay	Departure station	Standard deviation
Cleanliness and information	Departure station	Weighted average

Source: Author's compilation

Based on the results, the examination of two additional questions has been reasonable.

4. What policy implications follow from the above results?

In terms of transport theories, the results are in line with expectations based on the literature and intuition. However, when looking at the details, several questions emerge that may have direct applications. First of all, at some points the demand for rail services is inelastic: it follows that a very conscious and complex transport development is needed to achieve a modal shift from private motorised transport to rail, which is clearly in the public interest from an ecological and quality of life point of view. It must be recognised that demand depends on a delicate balance of many factors and that a single intervention alone will not, or may not, be effective.

A key question is to what extent rail can be considered an inferior service today. The results lead to a twofold conclusion: on the one hand, rail is no longer locked into the inferior, undesirable service category it was in after the regime change. Quality sensitive demand and its components show that timetable improvements, modern rolling stock, cleaner trains, evolving information have a real impact on travellers, there is a quality-oriented part of demand.

On the other hand, the evidence of inflexible demand points to the fact that rail is still much less preferred by those for whom other modes are readily available. For example, with highly developed rail services, transport culture and, of course, a more ecologically aware population, car ownership should explain much less of the evolution of demand for rail services.

5. What further lines of investigation do the results suggest?

Probably the most important conclusion of the research is that the database compiled is particularly well suited to assessing the demand for rail services and the quality of service. With such a large amount of data, it is possible to analyse even finer details and thus answer not only general questions but also more detailed ones.

In particular, the incompleteness of the data sources and the necessary simplifications in the case of interconnections point to further possibilities: building on substantial research resources, by including additional available data and eliminating most of the simplifications, a much more accurate and detailed database.

In addition, the use of empirical sources rather than the weights in the Public Service Contract for quality factors would be an excellent basis for research that measures consumer preferences with a focus on soft factors. Complementing regular, relatively narrowly focused customer surveys with a specific, larger-scale survey could provide answers to many questions which, when combined with demand data, would significantly increase our knowledge of demand for rail services, and on which effective and efficient public policy interventions can be based.

6 LIST OF REFERENCES

Adler, N., Pels, E. and Nash, C. (2010) 'High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis', *Transportation Research Part B: Methodological*. Elsevier Ltd, 44(7), pp. 812–833. doi: 10.1016/j.trb.2010.01.001.

Ahern, A. A. and Tapley, N. (2008) 'The use of stated preference techniques to model modal choices on interurban trips in Ireland', *Transportation Research Part A: Policy and Practice*, 42(1), pp. 15–27. doi: 10.1016/j.tra.2007.06.005.

Allianz pro Schiene (2007) *Város, vidék, vasút. Kötöttpályás megoldások a regionális közlekedésben: 16 sikeres példa*. Berlin. Available at: https://www.allianz-pro-schiene.de/wp-content/uploads/2015/09/allianz_pro_schiene_stadtlandschiene_ungarische_version_3Auflage.pdf (Accessed: 10 August 2017).

Anderson, J. E. (2011) 'The Gravity Model', *Annual Review of Economics*, 3(1), pp. 133–160. doi: 10.1146/annurev-economics-111809-125114.

Andrés López-Pita, F. R. (2005) 'Impact of High-Speed Lines in Relation to Very High Frequency Air Services.pdf', *Journal of Public Transportation*, 8(2), pp. 17–35. Available at: http://www.nctr.usf.edu/jpt/pdf/JPT 8-2 Lopez-Pita_Robuste.pdf.

Ashiabor, S., Baik, H. and Trani, A. (2007) 'Logit Models for Forecasting Nationwide Intercity Travel Demand in the United States', *Transportation Research Record: Journal of the Transportation Research Board*, 2007, pp. 1–12. doi: 10.3141/2007-01.

Balázsi, L. *et al.* (2014) 'A közgazdasági adatforradalom és a panelökonometria', *Közgazdasági Szemle*, 61(11), p. 1319—1340. Available at: http://www.kszemle.hu/tartalom/cikk.php?id=1515.

Barrow, K. (2018) *Ten years, 27,000km: China celebrates a decade of high-speed*, *International Railway Journal*. Available at: https://www.railjournal.com/passenger/high-speed/ten-years-27000km-china-celebrates-a-decade-of-high-speed (Accessed: 12 September 2018).

Békés, G. and Kézdi, G. (2021) *Data Analysis for Business, Economics, and Policy*. Cambridge University Press. doi: 10.1017/9781108591102.

Bel, G. (1997) 'Changes in travel time across modes and its impact on the demand for inter-urban rail travel', *Transportation Research Part E: Logistics and Transportation Review*, 33(1), pp. 43–52. doi: 10.1016/S1366-5545(96)00004-X.

BKK (2021) *Jegyek és bérletek*. Available at: https://bkk.hu/jegyek-es-berletek/arak/ (Accessed: 21 August 2021).

Bode, M. *et al.* (2018) *Special Report*. Available at: https://www.eca.europa.eu/Lists/ECADocuments/SR18_19/SR_HIGH_SPEED_RAIL_EN.pdf (Accessed: 21 December 2019).

Brand, D. *et al.* (1992) 'Forecasting High-Speed Rail Ridership', *Transportation Research Record*, (1341), pp. 12–18.

Brons, M., Givoni, M. and Rietveld, P. (2009) 'Access to railway stations and its potential in increasing rail use', *Transportation Research Part A: Policy and Practice*. Elsevier Ltd, 43(2), pp. 136–149. doi: 10.1016/j.tra.2008.08.002.

Budapest Fejlesztési Központ (2020) *Budapesti Agglomerációs Vasúti Stratégia*. Available at: https://budapestvasut2040.hu/wp-content/uploads/2021/01/Strategia.pdf?utm_source=pestbuda&utm_medium=link&utm_campaign=pestbuda_202106.

Button, K. (2006) 'Transportation Economics: Some Developments Over the Past 30 Years', *Transportation Research*, 45(2), pp. 7–30.

Cadarso, L. *et al.* (2017) 'Integrated Airline Scheduling: Considering Competition Effects and the Entry of the High Speed Rail', *Transportation Science*. INFORMS , 51(1), pp. 132–154. doi: 10.1287/trsc.2015.0617.

Cavana, R. Y., Corbett, L. M. and lo, Y. L. G. (2007) 'Developing zones of tolerance for managing passenger rail service quality', *International Journal of Quality & Reliability Management*, 24(1), pp. 7–31. doi: 10.1108/02656710710720303.

Centre for Aviation (2011) *A decade of change for the global airline industry: A new and altered reality in the rankings*. Available at: https://centreforaviation.com/analysis/reports/a-decade-of-change-for-the-global-airline-industry-a-new-and-altered-reality-in-the-rankings-55209 (Accessed: 19 January 2019).

Chang, Y. H., Yeh, C. H. and Shen, C. C. (2000) 'A multiobjective model for passenger train services planning: Application to Taiwan's high-speed rail line', *Transportation Research Part B: Methodological*, 34(2), pp. 91–106. doi: 10.1016/S0191-2615(99)00013-2.

Clewlow, R. R., Sussman, J. M. and Balakrishnan, H. (2014) 'The impact of high-speed rail and low-cost carriers on European air passenger traffic', *Transport Policy*. Elsevier, 33, pp. 136–143. doi: 10.1016/j.tranpol.2014.01.015.

Cohen, G. S., Erlbaum, N. S. and Hartgen, D. T. (1978) 'Intercity Rail Travel Models', *Transportation Research Record: Journal of the Transportation Research Board*, (673), p. pp 21-25.

LA CoMotion (2017) *User-Experience & the Future of Public Transit*. Available at: https://www.lacomotion.com/archives/2840 (Accessed: 9 January 2019).

Connolly, K. (2018) *'We are becoming a joke': Germans turn on Deutsche Bahn*, *The Guardian*. Available at: https://www.theguardian.com/world/2018/dec/20/trains-on-time-germans-deutsche-bahn-railway (Accessed: 21 December 2018).

coonan, clifford (2011) *Outrage at Wenzhou disaster pushes China to suspend bullet train project | The Independent*. Available at: https://www.independent.co.uk/news/world/asia/outrage-at-wenzhou-disaster-pushes-china-to-suspend-bullet-train-project-2336254.html (Accessed: 12 January 2019).

Dai, T. and Jin, F. (2008) 'Spatial interaction and network structure evolvement of cities in terms of China's rail passenger flows', *Chinese Geographical Science*, 18(3), pp. 206–213. doi: 10.1007/s11769-008-0206-2.

Debrezion, G., Pels, E. and Rietveld, P. (2009) 'Modelling the joint access mode and railway station choice', *Transportation Research Part E: Logistics and Transportation Review*, 45(1), pp. 270–283. doi: 10.1016/j.tre.2008.07.001.

delpestibusz.hu (2021) *Jegyarchívum*. Available at: http://www.delpestibusz.hu/jegyb.html#gyvas (Accessed: 10 August 2021).

Deutsche Bahn (2018) *Flinkster Carsharing | Deutsche Bahn Connect*. Available at: https://www.deutschebahnconnect.com/en/mobilitaetskonzepte/carsharing/flinkster_carsharing (Accessed: 19 December 2018).

Diaconu, L. (2012) 'The Evolution of the European Low-cost Airlines'Business Models. Ryanair Case Study', *Procedia - Social and Behavioral Sciences*, 62, pp. 342–346. doi: 10.1016/j.sbspro.2012.09.054.

Douglas Economics (2006) *Value and Demand Effect of Rail Service Attributes*.

Duranton, S. *et al.* (2017) *The 2017 European Railway Performance Index*, *BCG*. Available at: https://www.bcg.com/publications/2017/transportation-travel-tourism-2017-european-railway-performance-index.aspx (Accessed: 19 January 2018).

Eboli, L. and Mazzulla, G. (2015) 'Relationships between rail passengers' satisfaction and service quality: a framework for identifying key service factors', *Public Transport*, 7(2), pp.

185–201. doi: 10.1007/s12469-014-0096-x.

Édes, B. (2019) 'Egy hálózatos iparág élő hálózatok nélkül. Miért múlt századi az ügyfélélmény a magyar közösségi közlekedési rendszerben, és mit lehetne tanulni más iparágaktól?', *Vezetéstudomány / Budapest Management Review*, 50(9), pp. 81–94. doi: 10.14267/VEZTUD.2019.09.08.

Édes, B., Gerhardt, E. and Micski, J. (2011) 'A liberalizáció első időszakának versenyszempontú értékelése a magyar asúti teherszállítási piacon', in *Verseny és Szabályozás*. Budapest. Available at: http://old.mtakti.hu/file/download/vesz2011/vasuti_teher.pdf (Accessed: 19 January 2019).

European Commission (1992) *The Future Development Of The Common Transport Policy*. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM%3A1992%3A0494%3AFIN (Accessed: 1 August 2021).

European Commission (1996) *White Paper: A STRATEGY FOR REVITALISING THE COMMUNITY'S RAILWAYS CONTENTS*. Available at: http://europa.eu/documents/comm/white_papers/pdf/com96_421_en.pdf (Accessed: 19 January 2019).

European Commission (2018) *Multimodal Sustainable Transport: which role for the internalisation of external costs?* Available at: https://ec.europa.eu/transport/sites/transport/files/2018-year-multimodality-external-costs-note.pdf (Accessed: 10 January 2019).

European Union, A. for R. (2016) *Big data in railways*. Available at: https://www.era.europa.eu/sites/default/files/activities/docs/cor_big_data_en.pdf (Accessed: 20 January 2019).

Fleischer, T. (2006) 'A vasúti pályaudvarok új szerepe a kibôvített', *Európai Tükör*, 11(5), pp. 53–63. Available at: http://www.vki.hu/~tfleisch/PDF/pdf06/fleischer_vasutallomas_eutukor06-5.pdf.

Global Railway Review (2018) *Barcelona-Madrid high-speed line has more than 85 million passengers*. Available at: https://www.globalrailwayreview.com/news/66620/barcelona-madrid-85-million-passengers/ (Accessed: 13 December 2018).

Goodwin, P. B. (1992) 'A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes', *Journal of Transport Economics and Policy*, 26(2), pp. 155–169. doi: 10.2307/20052977.

Google (2021a) *Distance Matrix Service*. Available at: https://developers.google.com/maps/documentation/javascript/distancematrix (Accessed: 15 July 2021).

Google (2021b) *Googlemaps utazástervező*. Available at: https://www.google.com/maps/dir/budapest/Zalaegerszeg/@46.9878181,17.4111407,9z/am=t/data=!4m17!4m16!1m5!1m1!1s0x4741c334d1d4cfc9:0x400c4290c1e1160!2m2!1d19.040235!2d47.497912!1m5!1m1!1s0x47692817e6d34a01:0x400c4290c1e1250!2m2!1d16.8416322!2d46.8416936!2m (Accessed: 10 August 2021).

GYSEV (2018) *Éves jelentés 2017*. Available at: https://www2.gysev.hu/sites/gysev/files/atoms/files/gysev_eves_jelentes_web_oldalparban_175x251.pdf (Accessed: 11 January 2019).

Hansson, T., Ringbeck, J. and Franke, M. (2003) 'Flight for Survival: A New Business Model for the Airline Industry', *Strategy+Business*. Available at: https://www.strategy-business.com/article/21966?gko=6e1b2 (Accessed: 19 January 2019).

Holmid, S. (2005) 'INTERACTION DESIGN AND SERVICE DESIGN: EXPANDING A COMPARISON OF DESIGN DISCIPLINES.', in *Nordic design research conference*. Nordic Design Research. Available at: http://www.nordes.org/opj/index.php/n13/article/view/157/140 (Accessed: 19 January 2019).

Homolya, R. (2018) *A magyar vasút jelene és jövője*. Available at: http://binx.ktenet.hu/rendezvenyek/tagozati/2018-11-13-15_KeT_iiMKK/eloadasok/Plenaris-este/HomolyaRobert.pdf (Accessed: 11 December 2018).

Hörcher, D. (2021) *Száguldó magyar szupervonat: közlekedéspolitikai reneszánsz vagy totális aránytévesztés? – Válasz Online*, *Száguldó magyar szupervonat: közlekedéspolitikai reneszánsz vagy totális aránytévesztés?* Available at: https://www.valaszonline.hu/2021/07/15/horcher-daniel-nsv-vasut-fejlesztes-terv-velemeny/ (Accessed: 26 August 2021).

iho.hu (2009) *MAEF – A gépi jegykiadás története*. Available at: http://iho.hu/hir/maef (Accessed: 10 December 2018).

Independent.ie (2018) 'Ryanair doubles the price of standard seat selection - "It's your choice"'. Available at: https://www.independent.ie/life/travel/travel-news/ryanair-doubles-the-price-of-standard-seat-selection-its-your-choice-36738777.html (Accessed: 19 November 2018).

Index.hu (2015) 'Index - Belföld - Látta már az ország legocsmányabb vasútállomását?

Megmutatjuk!' Available at: https://index.hu/belfold/2015/07/21/latta_mar_az_orszag_legocsmanyabb_vasutallomasat_megmutatjuk/ (Accessed: 19 December 2018).

itf.hu (2008) *itf.hu - Integrált ütemes menetrend (ITF)*. Available at: http://www.itf.hu/index.php/alapfogalmak/integralt-uetemes-menetrend-itf (Accessed: 9 January 2019).

itf.hu (2016) *itf.hu - 12 éve indult el az első, integrált ütemes menetrend Magyarországon*. Available at: http://www.itf.hu/index.php/cikk/109-12-eve-indult-el-az-elso-integralt-uetemes-menetrend-magyarorszagon (Accessed: 9 January 2019).

Jaffe, E. (2012) *Does Light Rail Really Alleviate Highway Congestion? - CityLab*. Available at: https://www.citylab.com/transportation/2012/02/does-light-rail-really-alleviate-highway-congestion/1358/ (Accessed: 19 January 2019).

Janic, M. (2003) 'High-speed rail and air passenger transport: A comparison of the operational environmental performance', *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 217(4), pp. 259–269. doi: 10.1243/095440903322712865.

Jones, I. S. and Nichols, A. J. (1983) 'The Demand for Inter-City Rail Travel in the United Kingdom: Some Evidence', *Journal of Transport Economics and Policy*, Vol. 17(No. 2 (May, 1983)), pp. 133–153.

de Jong, E. (2017) 'Lessons of liberalisation: How do Britain's railways really compare with others in Europe? - Global Railway Review'. Available at: https://www.globalrailwayreview.com/article/32888/lessons-liberalisation-britains-railways-re-ally-compare-others-europe/ (Accessed: 19 January 2019).

Közlekedő Tömeg Egyesület (2018) *Turista legyen a talpán, aki nálunk vonatozni akar - Közlekedő Tömeg*. Available at: https://kozlekedotomeg.blog.hu/2018/01/06/turista_legyen_a_talpan_aki_nalunk_vonatozni_akar (Accessed: 2 March 2018).

KSH (2021) *KSH stadat rendszer*. Available at: https://www.ksh.hu/stadat_files/ara/hu/ara0004.html (Accessed: 1 August 2021).

KTI (2021) *Trendek*, *Trendek Közlekedéstudományi Intézet*. Available at: http://www.kti.hu/trend-kti/ (Accessed: 19 January 2019).

Kuo, C. and Tang, M. (2013) 'Relationships among service quality, corporate image, customer satisfaction, and behavioral intention for the elderly in high speed rail services', *JOURNAL OF ADVANCED TRANSPORTATION*, 47, pp. 515–525.

Kwitny, J. (1981) 'The Great Transportation Conspiracy', *Harper's Magazine*, 262(1569), pp. 14–21.

Lalive, R., Luechinger, S. and Schmutzler, A. (2013) 'Does Supporting Passenger Railways Reduce Road Traffic Externalities?' Available at: http://www.econ.uzh.ch/static/wp/econwp110.pdf (Accessed: 19 January 2019).

Litman, T. (2008) 'Valuing Transit Service Quality Improvements', *Journal of Public Transportation*, 11(2), pp. 43–64. doi: 10.5038/2375-0901.11.2.3.

Litman, T. (2013) 'Transport Elasticities : Impacts on Travel Behaviour', *Sustainable Urban Transport Technical Document #11*, pp. 1–40. Available at: http://www.sutp.org/files/contents/documents/resources/B_Technical-Documents/GIZ_SUTP_TD11_Transport-Elasticities_EN.pdf.

Liu, Y. *et al.* (2014) 'Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data', *PLoS ONE*, 9(1). doi: 10.1371/journal.pone.0086026.

Luk, G. (2017) 'Why Doesn't The United States Have High-Speed Bullet Trains Like Europe And Asia?', *Forbes*. Available at: https://www.forbes.com/sites/quora/2017/03/11/why-doesnt-the-united-states-have-high-speed-bullet-trains-like-europe-and-asia/#7ff59bacc080 (Accessed: 19 May 2018).

Lythgoe, W., Wardman, M. and Toner, J. (2004) 'Enhancing rail passenger demand models to examine station choice and access to the rail network', in *European Transport Conference*, pp. 4–6. doi: 10.1038/nclimate2153.

Martín, J. C. and Nombela, G. (2007) 'Microeconomic impacts of investments in high speed trains in Spain', *Annals of Regional Science*, 41(3), pp. 715–733. doi: 10.1007/s00168-007-0116-8.

MÁV-csoport (2017) *Új kedvezmények is érkeznek az új vasúti menetrenddel*. Available at: https://www.mavcsoport.hu/mav-start/belfoldi-utazas-nemzetkozi-utazas/uj-kedvezmenyek-erkeznek-uj-vasuti-menetrenddel (Accessed: 10 July 2021).

MÁV-csoport (2021a) *2014-2023 évi vasúti személyszállítási közszolgáltatási szerződés*. Available at: https://www.mavcsoport.hu/mav-start/dokumentumok/2014-2023-evi-vasuti-szemelyszallitasi-kozszolgaltatasi-szerzodes-es (Accessed: 10 July 2021).

MÁV-csoport (2021b) *MÁV-csoport összevont (konszolidált) éves beszámoló 2020. december*

31. Available at: https://www.mavcsoport.hu/sites/default/files/upload/page/mav_konszolidalt_eves_beszamolo_2020.pdf.

MÁV-START (2014) *Rövidebb menetidő + több vonat = vonzó szolgáltatás, növekvő utasszám*. Available at: https://www.mavcsoport.hu/mav/rovidebb-menetido-tobb-vonat-vonzo-szolgaltatas-novekvo-utasszam (Accessed: 2 August 2021).

MÁV-START (2018a) *A MÁV-START VASÚTI SZEMÉLYSZÁLLÍTÓ ZÁRTKÖRŰEN MŰKÖDŐ RÉSZVÉNYTÁRSASÁG SZEMÉLYSZÁLLÍTÁSI ÜZLETSZABÁLYZATA*. Available at: https://www.mavcsoport.hu/sites/default/files/upload/page/a_mav-start_szemelyszallitasi_uzletszabalyzata_6_mod._181206.pdf (Accessed: 23 May 2019).

MÁV-START (2018b) *Belföldi közforgalmú menetrend*.

MÁV (2018) *Üzemeltetés támogató szoftver adattárház kiépítése MS Power BI alapokon | MÁV-csoport*. Available at: https://www.mavcsoport.hu/mav-csoport/beszerzesi-hirdetmenyek/uzemeltetes-tamogato-szoftver-adattarhaz-kiepitese-ms-power-bi (Accessed: 19 December 2018).

McFadden, D. (1974) 'The Measurement of Urban Travel Demand', *Journal of Public Economics 3*, pp. 303-328.

Mihályi, P. (2004) 'Szempontok a magyar vasút távlati stratégiájának kidolgozásához, I. és II. rész', *Kożlekedéstudományi Szemle*, 54.(9-10.), pp. 322-336., 362-367.

Mobility (2021) *The Mobility Cooperative*. Available at: https://www.mobility.ch/en/mobility-cooperative (Accessed: 20 August 2021).

MTA KRTK (2021) *KSH területi-statisztikai rendszer*. Available at: https://adatbank.krtk.mta.hu/ksh-teruleti-statisztika-t-star/ (Accessed: 13 July 2021).

NASA (2001) *SATS: A bold vision NASA-led technology development aimed at increasing mobility, access for smaller communities*. Available at: www.nasa.gov/centers/langley/news/factsheets/SATS.html.

Nathanail, E. (2008) 'Measuring the quality of service for passengers on the hellenic railways', *Transportation Research Part A: Policy and Practice*, 42(1), pp. 48–66. doi: 10.1016/j.tra.2007.06.006.

Nohe, S. (2017) *TriRail… try a little harder. UX research on south Florida's train*. Available at: https://medium.com/@sarahnohe/trirail-try-a-little-harder-rethinking-the-south-florida-train-experience-b47b153cfa05 (Accessed: 13 December 2018).

Norman, D., Miller, J. and Henderson, A. (1995) ‘What you see, some of what’s in the future, and how we go about doing it’, in *Conference companion on Human factors in computing systems - CHI ’95*. New York, New York, USA: ACM Press, p. 155. doi: 10.1145/223355.223477.
ORR (2018) *Passenger Rail Usage: Quality and Methodology Report*. Available at: http://www.statisticsauthority.gov.uk/assessment/code-of-practice/index.html (Accessed: 12 December 2018).
Oum, T. H., Waters II, W. G. and Yong, J. S. (1990) ‘A survey of recent estimates of price elasticities of demand for transport’, *World Bank Working Papers*, p. 1. Available at: http://documents.worldbank.org/curated/en/573201468766481035/A-survey-of-recent-estimates-of-price-elasticities-of-demand-for-transport (Accessed: 12 January 2019).
Outwater, M. *et al.* (2010) ‘www.econstor.eu’, *Journal of Choice Modelling*, 3(1), pp. 58–83.
Owen, A. D. and Phillips, G. D. . (1987) ‘The Characteristics of Railway Passenger Demand. An Econometric Investigation’, *Journal of Transport Economics and Policy*, 21(3), pp. 231–253. doi: 10.2307/20052820.
Paulley, N. *et al.* (2006) ‘The demand for public transport: The effects of fares, quality of service, income and car ownership’, *Transport Policy*, 13(4), pp. 295–306. doi: 10.1016/j.tranpol.2005.12.004.
Petőfi, S. (1847) *Petőfi Sándor: VASUTON | Verstár - ötven költő összes verse | Kézikönyvtár*. Available at: https://www.arcanum.hu/hu/online-kiadvanyok/Verstar-verstar-otven-kolto-osszes-verse-2/petofi-sandor-DFB2/1847-EFBB/vasuton-F5F3/ (Accessed: 11 October 2018).
Rahaman, K. R. and Rahaman, A. (2009) ‘SERVICE QUALITY ATTRIBUTES AFFECTING THE SATISFACTION OF RAILWAY’, 4(3), pp. 115–125.
retronom.hu (2021) *retronom.hu*. Available at: http://retronom.hu/node/6510 (Accessed: 10 August 2021).
Rozycki, C. von, Koeser, H. and Schwarz, H. (2003) ‘Ecology profile of the german high-speed rail passenger transport system, ICE’, *The International Journal of Life Cycle Assessment*, 8(2), pp. 83–91. doi: 10.1007/BF02978431.
de Rus, G. (1990) ‘Public Transport Demand Elasticities in Spain’, *Journal of Transport Economics and Policy*, 24(2), pp. 189–201. Available at: http://www.jstor.org/stable/20052917.
Smart Rail World (2017) *Taiwan’s high-speed rail ten years on.* Available at: https://www.smartrailworld.com/taiwans-high-speed-rail-ten-years-on (Accessed: 12

January 2019).
Swisspass (2021) *SwissPass Plus Partners*. Available at: https://www.swisspass.ch/plus/partner?category=cat_auto (Accessed: 20 August 2021).
Thakuriah, P. *et al.* (2010) 'Estimation of the Demand for Inter-City Travel - Issues with Using the American Travel Survey', *Transportation Research E-Circular*, E-C026(Personal Travel: The Long and Short of It), pp. 255–269.
The Economist (2018) 'Why Europe's train network needs more, not less, competition - Free the rails'. Available at: https://www.economist.com/leaders/2018/06/30/why-europes-train-network-needs-more-not-less-competition (Accessed: 19 November 2018).
The High Line (2018) *History | The High Line*. Available at: https://www.thehighline.org/history/ (Accessed: 19 December 2018).
The Wall Street Journal (2017) *How Budget Carriers Transformed the Airline Industry—in 14 Charts - WSJ*. Available at: https://www.wsj.com/articles/how-budget-carriers-transformed-the-airline-industryin-14-charts-1503501624 (Accessed: 3 January 2019).
Topham, G. (2017) *Rail services lost under 1960s Beeching cuts may reopen*. Available at: http://www.theguardian.com/business/2017/nov/28/uk-rail-services-lost-beeching-cuts-could-reopen-chris-grayling (Accessed: 24 November 2018).
vasutallomasok.hu (2011) *Vasútállomások.hu - Balatonfőkajár felső*. Available at: http://www.vasutallomasok.hu/allomas.php?az=blff (Accessed: 18 October 2018).
Vitézy, D. (2014) 'A fővárosi közlekedésfejlesztés eredményei', in *A városi közlekedés aktuális kérdései*. Balatonföldvár. Available at: https://bkk.hu/wp-content/uploads/2014/09/KTE_Balatonfenyves_VD_20140811.pdf (Accessed: 19 December 2019).
VOR (2021) *Verkehrsverbund Ostregion - Über Uns*. Available at: https://www.vor.at/ueber-uns/ (Accessed: 10 August 2021).
Walker, J. (2017) *The Dangers of Elite Projection — Human Transit*. Available at: https://humantransit.org/2017/07/the-dangers-of-elite-projection.html?fbclid=IwAR01hwnrpm8whEDrxq3kTc_k-9pWD3EiSJIRcjj4z0HyLQDVY6q-Ny-Vfcc (Accessed: 12 January 2019).
Wardman, M. (1994) 'Forecasting the Impact of Service Quality Changes on the Demand for Interurban Rail Travel', *Journal of Transport Economics and Policy*, 28(3), pp. 287–306. doi: 10.1007/s00330-006-0309-z.

Wardman, M. (2001) 'A review of British evidence on time and service quality valuations', *Transportation Research Part E: Logistics and Transportation Review*, 37(2–3), pp. 107–128. doi: 10.1016/S1366-5545(00)00012-0.
Wardman, M. (2006) 'Demand for rail travel and the effects of external factors', *Transportation Research Part E: Logistics and Transportation Review*, 42(3), pp. 129–148. doi: 10.1016/j.tre.2004.07.003.
Wardman, M., Lythgoe, W. and Whelan, G. (2007) 'Rail Passenger Demand Forecasting: Cross-Sectional Models Revisited', *Research in Transportation Economics*, 20(07), pp. 119–152. doi: 10.1016/S0739-8859(07)20005-8.
Wardman, M. and Whelan, G. (2011) 'Twenty years of rail crowding valuation studies: Evidence and lessons from British experience', *Transport Reviews*, 31(3), pp. 379–398. doi: 10.1080/01441647.2010.519127.
Webster, F.V., Bly, P. H. (Eds. . (1980) *The Demand for Public Transport. Report of an International Collaborative Study*. Crowthorne, Berkshire.: Transport and Road Research Laboratory.
Whimapp (2021) *How it works*. Available at: https://whimapp.com/helsinki/en/how-it-works/ (Accessed: 30 August 2021).
Wikimedia.org (2009) *wikimedia médiatár*. Available at: https://commons.wikimedia.org/wiki/File:Vonatjegy,_MÁV,_Kaposvár-Szeged,_2009.08.31.jpg (Accessed: 30 August 2021).
wikipedia.org (2021a) *MÁV Magyar Államvasutak Zrt.* Available at: https://hu.wikipedia.org/wiki/MÁV_Magyar_Államvasutak_Zrt. (Accessed: 2 August 2021).
wikipedia.org (2021b) *Regisztrációs jegy*. Available at: https://hu.wikipedia.org/wiki/Regisztrációs_jegy (Accessed: 10 August 2021).
wikipedia.org (2021c) *Wikipédia: Stadler Flirt*. Available at: https://hu.wikipedia.org/wiki/Stadler_FLIRT (Accessed: 5 July 2021).
Wolmar, C. (2018) *How Britain's Colonial Railways Transformed India*. Available at: https://reconnectingasia.csis.org/analysis/entries/how-britians-colonial-railways-transformed-india/%0Ahttp://files/252/how-britians-colonial-railways-transformed-india.html (Accessed: 12 June 2018).
Wright, R. (2016) 'New York tram stages an unlikely comeback | Financial Times'. Available at: https://www.ft.com/content/133c1c58-cb5c-11e5-a8ef-ea66e967dd44 (Accessed: 19

January 2019).

ZüriMobil (2021) *Entdecken Sie mit ZüriMobil das Mobilitätsangebot von Zürich*. Available at: https://www.stadt-zuerich.ch/site/zuerimobil/de/index.html (Accessed: 20 August 2021).