University of Southampton

Faculty of Engineering and Physics Science
Department of Civil, Maritime and Environmental Engineering

The theoretical and practical aspects of railway reform: implications for Saudi Arabia

DOI: https://doi.org/10.5258/SOTON/D2268

by

Sultan Alsaedi
MSC in Transportation Engineering and Planning
ORCiD: 0000-0002-2486-8192

A thesis for the degree of
Doctor of Philosophy

April 2023

University of Southampton

Abstract

Faculty of Engineering and Physics Science
Department of Civil, Maritime and Environmental Engineering

Doctor of Philosophy

The theoretical and practical aspects of railway reform: implications for Saudi Arabia

by Sultan Alsaedi

The Saudi government has started to reform its railway industry by following the international trends in railway reform. These reforms can be described as organisational reforms, but the railway's industrial organisation can be seen as not fully developed compared to other industries. Practically, railway reform can be clustered into three blocks based on country experiences, which are regulation, structural and ownership reforms. Based on the literature, the impacts of railway reform on performance vary between each form. Thus, this research aims to determine the optimal organisation of rail services in Saudi Arabia as an example of a country that has not fully developed its rail network.

To deliver the research aim, this research conducted technical, cost and welfare assessments to select the optimum railway organisation. To carry out these three assessments, this research performed analytical techniques that are built on quantitative and qualitative analyses. This research also conducted interviews with Saudi rail experts to validate the research recommendations. The results of this research conclude that the vertical separation between infrastructure managers and train operations could improve the efficiency of the Saudi railway system, particularly the institutional separation. Meanwhile, the regional division of rail services is not recommended for the Saudi rail system, and the horizontal separation between passenger and freight rail services is not supported by the findings of this research. With regards to the competition forms, this research has a mixture of results between the three assessments to determine the impact of on-track and off-track competitions. Finally, the results show that combining rail policies (competition introduction, vertical and horizontal separation) reduces railway efficiency. Based on these results, this research does not recommend introducing these reforms as one package, although the Saudi rail experts recommend this option.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed:... Date:..................

Acknowledgements

I would like to express my deepest appreciation to Professor John Preston and Dr. Simon Blainey for accepting me as a PhD student. Words cannot express my gratitude to Professor John Preston, and I could not have undertaken this journey without his support and guidance throughout delivering this research.

I would like to extend my sincere thanks to my family and friends, particularly to my lovely wife, Rawya for the support during the journey of completing this research.

Abbreviations

COLS	Corrected Ordinary Least Squares.
CRS	Constant Returns to Scale.
DEA	Data Envelopment Analysis.
DMU	Decision-Making Units.
IMs	Infrastructure managers.
MoT	Ministry of Transport.
OLS	Ordinary Least Squares.
PFP	Partial Factor Productivity.
PIF	Public Investment Fund.
PTA	Public Transport Authority.
ROSCO	Rolling Stock Company.
RTD	Returns to Density.
RTS	Returns to Scale.
RUs	Railway Undertakings.
SAR	Saudi Railway Company.
SFA	Stochastic Frontier Analysis.
SRO	Saudi Railway Organisation.
SSTPC	Saudi Spanish Train Project Company.
SUR	Seemingly Unrelated Regression.
TACs	Track Access Charges.
TFP	Total Factor Productivity.
TOCs	Train Operating Companies.
VRS	Variable Returns to Scale.

Chapter 1

Introduction

1.1 Research Background

1.1.1 Motivations for Railway Reform

An efficient transport system is a substantial contributor to economic development, which could be summarised as follows (World Bank, 2017):

- **Network effects**: a well-developed transportation network allows for better connectivity to major destinations for the transportation of large numbers of passengers and freight.
- **Performance improvements**: an efficient transport system could contribute to cost and time reductions for transport users, which in turn has a positive impact on economic growth.
- **Reliability**: a well-developed transport system has significant impacts on time consumption and a reduction in loss and damage for transport users.
- **Market size**: an efficient transport system provides better access to wider markets and adds to economies of scale in production, distribution, and consumption.
- **Productivity**: an efficient transport system improves productivity gained from the efficient allocation of input sources.

For the railway industry, rail transport can be described as an efficient transport mode in terms of energy consumption, service capacity, and environmental damage. Rail services, particularly long-distance services, have the capacity to convey large volumes of passengers and freight compared to other modes of transportation. In terms of environmental damage, rail services also emit fewer emissions per traffic unit (measured in passenger/km or tonne/km) (World Bank, 2017).

However, in the middle of the 20th century, the railway industry in several countries was confronted with external and internal challenges, resulting in performance deterioration. The substantial development of the other transport modes had a negative impact on rail services, which experienced losses in the modal split. For example, in 1949, the Japanese National Railway (JNR) was the dominant transport mode for passenger and freight services, but the JNR struggled to compete and began to lose market share due to the development of other motorised transport modes. For instance, between 1960 and 1987, passenger transport measured by passenger-km decreased by 29%, while freight transport measured by tonne-km decreased by 34%, as shown in Figure 1.1 (Ishida, 2011).

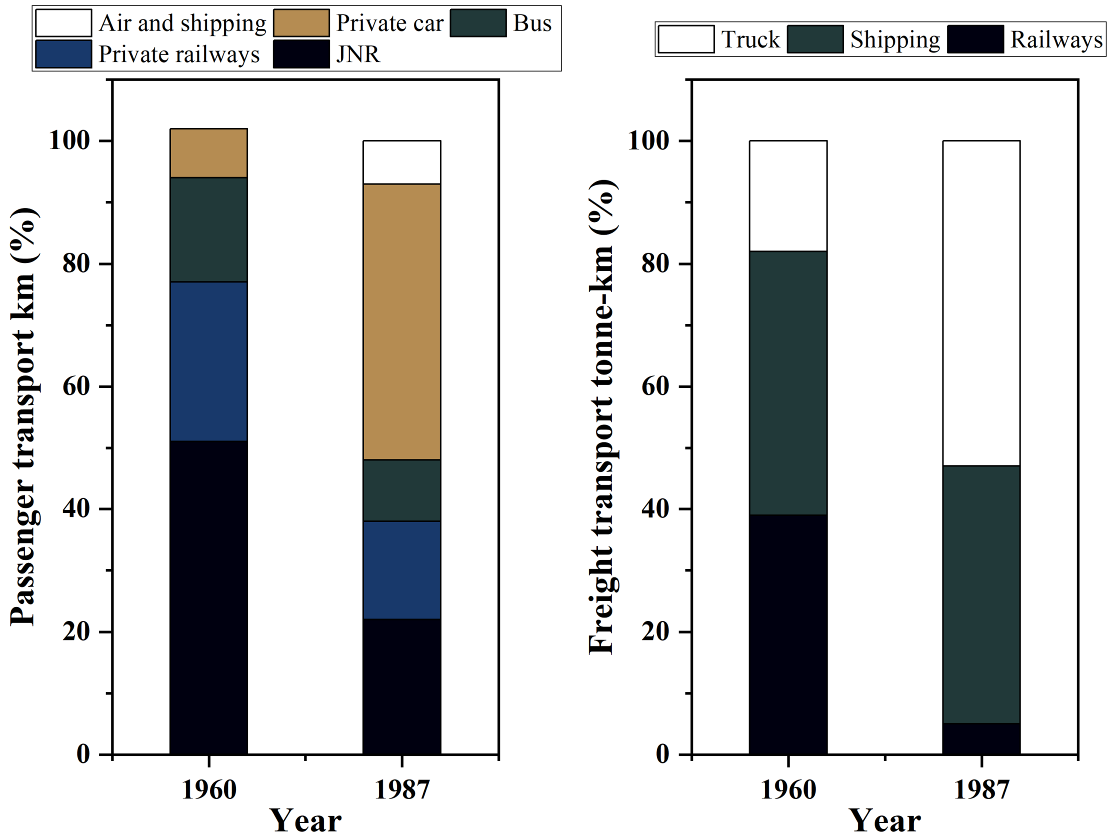


FIGURE 1.1: Market share of domestic transport modes in Japan. (Source: Ishida (2011)).

Besides the external impacts, the railway industry has internal challenges such as financial performance, high costs of rail infrastructure, and poor performance quality. For example, at the beginning of the 1990s, the cumulative deficit in the financial performance of Deutsche Bundesbahn (DB) (West Germany's state-owned railway business) was estimated to be approximately €25.5 billion (Brenck and Peter, 2007). In response to this challenge, the German Federal Government contributed roughly €7 billion per year to keep Deutsche Bundesbahn running, but the financial situation deteriorated to the point that income was insufficient to cover operational expenses (Brenck and Peter, 2007). Another example is the JNR, which had an inadequate financial performance from 1980 to 1986, with an annual deficit of almost one trillion yen ($4.53 billion) (Ishida, 2011). Despite the government's subsidies of roughly ¥700 billion ($3.17 billion) each year to alleviate financial difficulties, the JNR seemed to be on the verge of collapsing, with long-term debt of ¥25 trillion dollars ($113.37 billion) (Ishida, 2011). As a result, an immediate reform was necessary to save the JNR and keep it in operation.

1.1.2 Railway Reform and its Reasons

Railway systems across countries have faced rapid changes to tackle the challenges that are discussed above. These changes were intended to end the monopolistic behaviour of the railway systems. Furthermore, these changes have been addressed by governments and authorities as railway reforms. The implementation of these reforms is concentrated in three blocks, and Figure 1.2 illustrates the three blocks in which these changes were implemented.

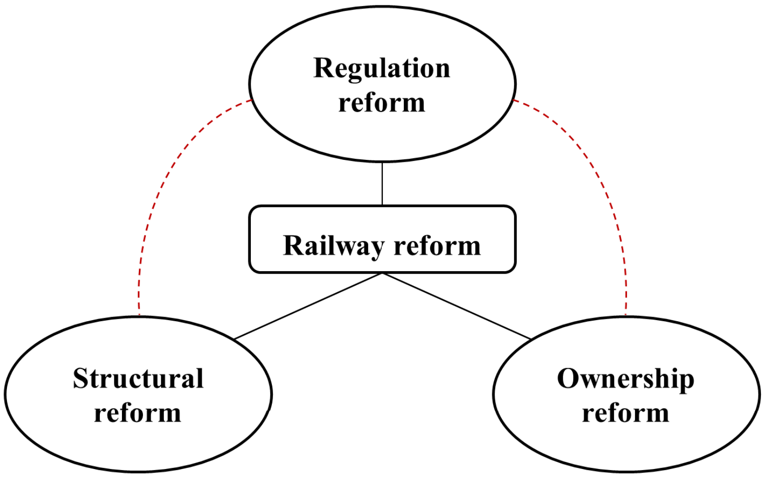


FIGURE 1.2: The three blocks of railway reforms implementation.

The regulation reform aims to liberalise the railway system to be more competitive and to improve its performance in terms of technical, financial and service quality. The structural reform aims to restructure the railway industry in the forms of vertical and/or horizontal separation. Vertical separation is a form of separating the rail infrastructure from train operations. Horizontal separation is a form of segregating passengers from freight rail services, as well as regional division. The ownership reform aims to transfer ownership of railway assets and operations from the government to the private sector. However, the reasons behind the railway reform can be summarised as follows:

- to revitalise passenger and freight rail services.
- to create a competitive railway market against other transport modes.
- to reshape the railway industry to be similar to other transport modes.
- to improve performance efficiency by adopting various rail policies.

1.1.3 Effects of Railway Reforms

The railway reform has a significant impact on performance. As discussed above, the railway system has been reformed via three forms, and each form has an impact on railway performance. For regulation reform, introducing competition in the passenger and freight rail markets has an impact by reducing costs and increasing technical efficiency. Additionally, economic regulation, as another form of regulation reform, has a positive effect on cost reduction, especially when this regulation is at a strong level (Smith et al., 2018). Besides regulation reform, structural reform has a significant impact on railway performance. Most forms of structural reform (e.g., vertical separation and horizontal separation) have positive impacts on the technical and financial performance of the railway industry (Cantos et al., 2012; Mizutani and Uranishi, 2013; Smith et al., 2018). Lastly, the ownership reform has a significant impact on railway performance. Moreover, the change in ownership of the railway systems improves technical and financial performance (Cantos et al., 2002; Cowie, 1999).

However, the impact of the railway reforms differs in terms of how the reforms are implemented. Although there is an improvement in railway performance due to different forms of reform, the way the reforms are implemented has a significant impact (Friebel et al., 2010). When the railway industry has been reformed as a one-package, there is a negative impact on railway performance. This means that when the railway system is reformed by introducing more than one form, the impact is negative. In contrast, when the railway industry has been reformed sequentially by introducing one form at a time, there has been a positive impact on the railway industry. This could be explained by governments having the opportunity to adjust the reform process based on their interests, and this benefit is absent in one package of railway reforms. A detailed discussion of the impact of railway reform on the performance of the railway market is provided in Chapter 5, 6 and 7 in the literature review section.

1.2 Research Scope

1.2.1 Railway Sector in Saudi Arabia

The railway industry in Saudi Arabia has not been fully developed with a lack of rail infrastructure, which was limited only to the eastern region. The first rail project which was constructed in the 1950s was the Dammam-Riyadh line for freight transportation, but passenger services were introduced in the 1980s. However, there was a lack of investment in the rail infrastructure in Saudi Arabia until the beginning of the 2000s when several projects were planned to be constructed. The first project is the North-South Railway (NSR), which is designed to transport passengers, freight and minerals. The second project is the Haramain High-Speed Rail (HHSR), which is aimed at serving passengers between Makkah and Madinah, passing through Jeddah. The third project is the land-bridge to connect the eastern with western regions. Currently, two projects have been launched to transport passengers and freight, which are the HHSR and NSR projects.

According to the Public Transport Authority (2017), the main players in the railway industry in Saudi Arabia are the following:

- **Ministry of Transport (MoT)**: as responsible for setting the strategic plans to enhance and expand the infrastructure development of the transport sector as a whole, including the road network, the railway infrastructure and the ports.
- **Public Transport Authority (PTA)**: as the regulator of public transport for passenger and freight services.
- **Saudi Railway Organisation (SRO)**: as a state-owned company under the MoT and responsible for infrastructure management and service operation for passenger and freight. SRO is the operator of the Riyadh Dammam line.
- **Saudi Railway Company (SAR)**: as a state-owned company under the Public Investment Fund (PIF) and responsible for the rail infrastructure and train operation of the north-south line (NSR).
- **Saudi Spanish Train Project Company (SSTPC)**: as a consortium between one Saudi private company and 12 Spanish company and responsible for operating and maintaining the HHSR project.

1.2.2 The changes

The railway industry in Saudi Arabia has a slow development process in terms of its infrastructure and organisation. Figure 1.3 shows a timeline of the railway

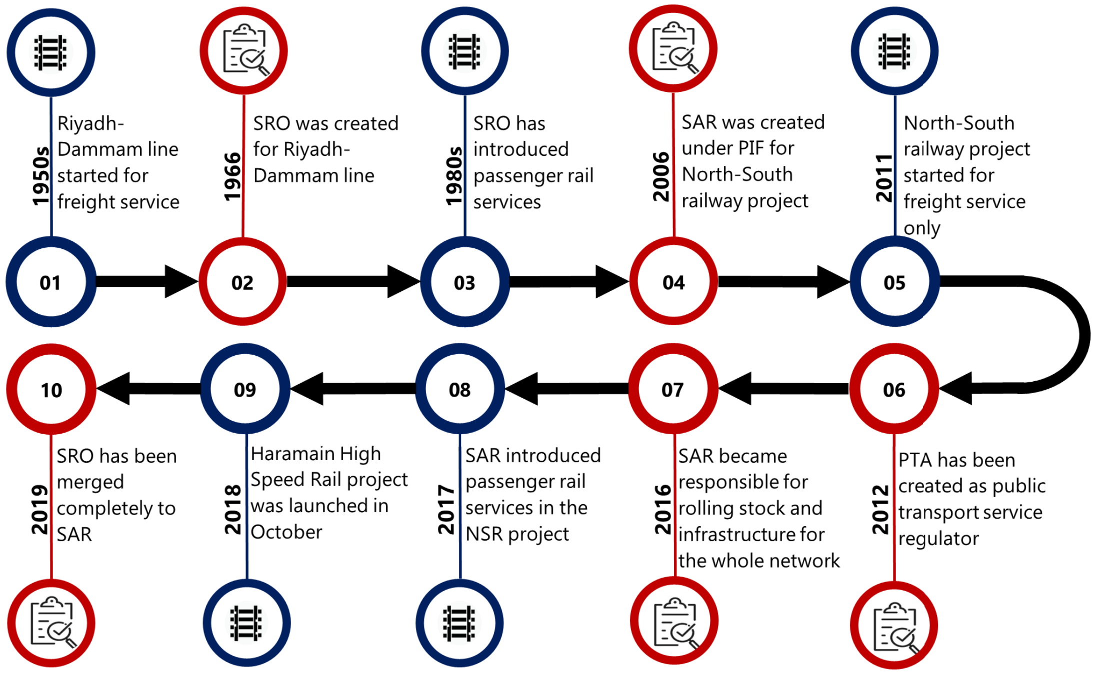


FIGURE 1.3: A time-line of railway development in Saudi Arabia.

development process in Saudi Arabia. As discussed above, the first railway project was the Riyadh-Dammam line, located in the Eastern region to transport freight only. In 1966, a Royal Decree addressed the creation of SRO as a railway company under MoT to be responsible for the Riyadh-Dammam line (Public Transport Authority, 2017).

Since the beginning of the 2000s, the Saudi government has announced several strategic plans to develop its railway industry. Several railway projects have been planned to connect the substantial locations that can support economic development and passenger and freight transportation. In addition, two entities have been created to support the development process. In 2006, a Royal Decree led to the creation of SAR under the Public Investment Fund (PIF), which is owned by the Saudi government, for the NSR project. This project was launched in 2011 to transport freight and minerals. Six years later, SAR has introduced passenger rail services on the NSR. Besides the NSR project, the HHSR project started in operation in 2018 to transport passengers from Makkah to Madinah.

For the railway organisation development, a railway reform has been addressed by a Royal Decree to transfer rail tracks of the whole network to SAR. This means that SAR is responsible for the rail infrastructure of the Riyadh-Dammam line, including the NSR. Another development is the creation of the PTA under the MoT as a public transport service regulator for passenger and freight sectors, and this has been addressed formally by a Royal Decree. The last railway reform was stated in a Royal

Decree in 2019 to merge SRO into SAR. This means that SAR is fully responsible for operating and managing the NSR and Riyadh-Dammam line.

The rail infrastructure of the Saudi railway system is shared for passenger and freight rail services except for some parts of the networks. For example, the HHSR is a high-speed rail that is mainly served passengers. The other parts of the network are running conventional trains, which in this way serve passengers and freight. The North-south rail project has some parts that run freight only. This part connects the mineral sites with the exporting port in the northern region. The major destinations for passenger and freight movements are the Riyadh-Dammam line.

1.3 Research Aim and Objectives

1.3.1 Research Problem and Questions

The Saudi government has started to reform its railway industry by following the international trends in railway reform. The initial stage of railway reform that occurred was the ownership reform when the government created a new railway company, SAR, for the NSR project rather than constructing and operating this project under SRO responsibility. The last stages of railway reform are privatising the HHSR under a build-operate-transfer (BOT) contract with SSTPC and merging SRO by transferring its assets to SAR. Thus, the research project will attempt to answer the following questions:

- What are the railway reforms and their impacts on performance?
- Which railway organisation can maximise technical, cost and welfare efficiency for the Saudi railway system?
- Do the research recommendations comply with the rail market needs in Saudi Arabia?

1.3.2 Aims and Objectives

The research project aims to determine the optimal organisation of rail services in a medium-sized country such as Saudi Arabia that has not been fully developed. This research attempts to fill the gap in the literature on the assessment of the railway organisation in the Saudi Arabian context. The assessment to optimise the organisation will be carried out by quantitative and qualitative analyses. In particular, the research project will perform several analytical techniques to evaluate different forms of railway organisation in terms of production, cost, and welfare efficiency. Thus, to carry out this aim, this research will address the following objectives:

- Defining the forms of railway reform that have occurred recently and their impacts on the performance of the railway industry.
- Reviewing the railway industry in Saudi Arabia and its projected plans and projects.
- Developing and testing econometric models and an analytical framework to estimate the technical, cost, and welfare of rail services.
- Quantifying the validity of the recommendations of this research via selected expert focus groups.

1.4 Thesis Structure

This section outlines the thesis structure and provides a brief description of each chapter. This thesis has eight further chapters as follows:

- **Chapter 2: Railway Reform**
  This chapter contains a review of the recent forms of railway reform and their impacts on performance. The chapter starts with an overview of railway reform and its forms. The discussion of regulation reform involves competition introduction, economic regulation, and the establishment of a rail regulator. Structural reform has been discussed in terms of the horizontal and vertical forms and the arguments for each form. A discussion of ownership reform has been addressed in terms of how the rail service provider has been changed, and some examples to support the discussion are given. In the end, the chapter reviews the literature regarding the estimation of the technical, cost, and welfare performance of the rail market.
- **Chapter 3: Research Sample**
  This chapter contains a detailed discussion of the railway industry of the main case study (Saudi Arabia). Additionally, the chapter states the criteria for selecting other railway industries as case studies for comparison purposes.
- **Chapter 4: Research Methodology**
  This chapter outlines the framework of this research. The chapter starts by discussing the analytical framework to assess the technical and cost efficiency and the methods that will be used to measure the efficiency. The following discussion is on welfare estimation.
- **Chapter 5: Technical Assessment**
  This chapter discusses the application of the technical assessment. This assessment aims to assess the impact of railway reform on technical efficiency.

This research estimates the technical efficiency via parametric and non-parametric approaches for 29 railway systems from 2000 to 2017. This estimation is then followed by a regression analysis to measure the impact of rail policies on the technical inefficiency levels.

- **Chapter 6: Cost Assessment**
  This chapter discusses the application of the cost assessment. This assessment aims to estimate the cost efficiency and to determine the impact of railway reform. The cost efficiency has been estimated via parametric and non-parametric approaches for 22 railway systems between 2000 and 2017. After the cost efficiency estimation, this research conducted a regression analysis to determine the impact of railway reform.

- **Chapter 7: Welfare Assessment**
  This chapter discusses the application of the welfare assessment. This research develops an approach based on the concept of welfare economics in this assessment. This approach develops demand and cost models to calculate welfare and its components. In each model, dummy variables are included to capture the impact of railway reform. The data sample for this assessment consists of 22 railway systems from 2001 to 2017.

- **Chapter 8: Policy Recommendation and Validation**
  This chapter discusses the results of three assessments in terms of the impact of railway reform. Based on these impacts, this research draws some recommendations for the Saudi railway market. This chapter also presents and discusses the results of the expert interviews. For the purpose of validation, this research provides a comparison between the findings and expert interviews.

- **Chapter 9: Conclusion and Future Work**
  This chapter draws a general conclusion about this research. Additionally, it outlines the main contributions of this research. In the end, the chapter discusses the limitations, future work, and opportunities for further investigators.

Chapter 2

Railway Market Structure

2.1 Overview

Traditionally, the railway industry was organised as vertical integration of rail infrastructure and train operations under public control. This form resulted in the industry being the most highly regulated. Economically, concerns about railway costs encouraged governments to exert tight control over the industry, including market entry and exit, fares, rail operations, and internal relationships. Due to these regulations, it was believed that the railway industry should be organised as a natural monopoly under government control, with a single owner accountable for the whole industry. As a result, the railway industry in most countries followed this form of market structure.

As stated by Nash and Preston (1992), railway economics has three fundamental aspects that may be considered constants. First, infrastructure costs are classified as fixed and sunk costs at high levels. Second, by its nature, the railway industry is a multi-product market. This implies that the industry provides rail services for a network with a diverse set of origins and destinations, as well as for the passenger and freight sectors. This characteristic creates certain challenges when it comes to cost allocation for each product. Last, the economies of density derive in part from these high fixed costs, which are aggravated by infrastructural indivisibilities. The effect of these constant factors led to the rail industry's sub-additive costs, which cannot be covered by marginal cost pricing. Following this finding, it was highly recommended, based on Berg and Tschirhart (1988) terminology, to organise the rail market as a natural monopoly with single ownership at the network level and either public management or ownership, to avoid any private monopoly abuses.

Despite the suggestion stated above, there are several arguments against this point of view, as highlighted by Preston (2012). First, based on econometric studies, it was found that rail operations have constant returns to scale with respect to operator size for the US railroads (Caves et al., 1985), which are small in size. This finding suggests that, unlike rail infrastructure, rail operations are not a natural monopoly. In another example, Pels and Rietveld (2003) found that there are increasing returns to scale in rail operations for European railways, which are considerably larger in size. Second, theoretical evidence presented by Leibenstein (1966) suggests that public ownership is associated with X-inefficiency. Third, the multi-product nature of the rail industry leads to diseconomies of scope. Economies of scope mean that the production of one service can lead to a cost reduction of producing another service. In other words, it is more cost effective for a firm to produce a variety of services or products rather than producing a single one. On the other hand, diseconomies of scope occurs when producing a service leads to increasing the cost of other services. For the railway system, this means that there are diseconomies of scope that occur by combining passengers with freight rail services.

Last, based on the theory of the contestable market, governmental control and ownership are not justified in the natural monopolistic market. Furthermore, the concept of contestability theory states that a firm, in a specific environment, will perform efficiently without regard to how many firms are in the market. This firm has an incentive to operate efficiently due to the threat of competition. To achieve this condition, the market should be free without any entry and exit barriers, and sunk costs should be at the lowest level. Obviously, this condition is not applicable to rail infrastructure, but it may be for rail operations.

2.2 Railway Reform

In recent decades, the railway industry has struggled to compete with other transport modes. The first reason is that the railway system was characterised by a state-owned monopoly that failed to meet market requirements due to a lack of incentives for managers (Drew and Ludewig, 2011). Different government departments, such as the ministries of transportation, finance, and labour, are also involved in operating the railway industry. The second reason is financial obstacles. Decision-makers seek to improve and expand railway services and infrastructure, but the railway industry, which is owned by the state, is faced with insufficient funds for railway development. Furthermore, decision-makers seek to provide railway services more appealing to rail users by lowering fares, but the railway, which is a vertically integrated monopoly owned by the government, is run at high fixed costs (Fröidh and Nelldal, 2015), which could have a negative impact on the railway market's financial performance. The limited technological developments in the railway industry might explain the high fixed costs.

To address these difficulties, the railway industry was reformed substantially across various countries. The railway reform can be described as any changes in the rail policies, investment plans and the structure of the rail industry (World Bank, 2017). The railway reform aims to reshape the industry in terms of several aspects (e.g., the regulations and legislation controlling the rail industry and the economic aspects of the rail market). Thus, it is essential to define the goals and objectives that are intended to be achieved by reforming the railway industry. World Bank (2017), for example, identified the following common railway reform objectives:

- Reduce government expenditures and liabilities associated with providing railway services;
- Provide more efficient railway services in terms of financial and sustainable aspects;
- Encourage the private sector to be involved in the railway industry in order to mitigate government responsibilities;
- Enhance railway services and customer responsiveness;
- Facilitate the railway system to be capable of promoting competition or accommodating new regulations;
- Specify the substantial goals that are needed to be achieved from the railway reform.

Based on the country's experiences, restructuring the rail market can be clustered into three blocks, as shown in Figure 2.1. The first block is how to create a competitive railway market among service providers. The railway industry in most countries was characterised as a state-owned monopoly with no competition, and this is unlike the other transport modes. Thus, the rail market has been reformed to introduce competition in line with other transport modes.

The second block is the main dimensions of separability, where the separation can be either horizontally or vertically. The horizontal dimension is to separate the rail market functions into passenger and freight sectors, which will be operated and managed separately. The railway industry can also be separated horizontally depending on a distinct geographic focus or market business core. This option is typically used for passenger rail services (e.g., in Germany and Great Britain). The horizontal separation intends to generate more manageable business units, increase transparency in financial performance and compare the efficiency of performance through benchmarking for business units (World Bank, 2017). Vertical separation, on the other hand, is a method of separating rail infrastructure from train operations into one or more entities. In

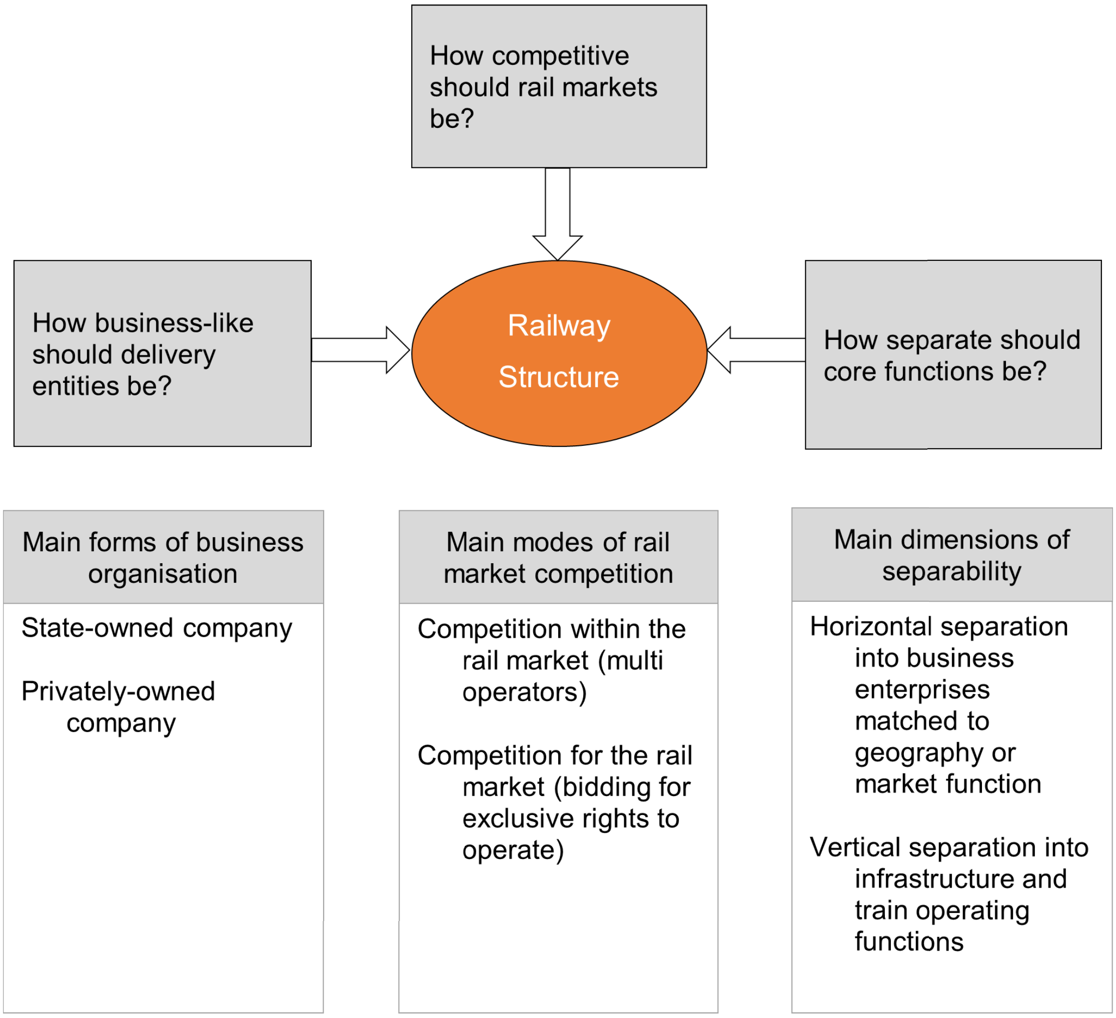


FIGURE 2.1: Main building blocks of rail industry structure. (Source: World Bank (2017)).

addition, this option can open the rail infrastructure of a vertically integrated rail operator to horizontally separated train operating companies for operation purposes.

The last block is the ownership of the railway industry. Moreover, this option means involving the private sector in the railway industry. This could reflect the previous two blocks (introducing competition and vertical and/or horizontal separation).

To achieve the three blocks, the railway industry across countries has been reformed in three aspects. The first aspect is regulation reform, which aims to create a competitive market, promote fair access to rail infrastructure, regulate rail services for the passenger sector, create an independent regulatory, etc. The second aspect is structural reform, where the industry is restructured horizontally and/or vertically. The third aspect is ownership reform, which is the change in how the rail market is managed and operated, whether by governments, private sectors, or some combinations, such as Public-Private Partnerships (PPPs).

2.2.1 Regulation Reform

Regulation may be considered as a substantial aspect of reforming the railway industry. As discussed above, railway reform has the potential to improve performance efficiency in terms of financial performance, long-term sustainability, and customer satisfaction. Thus, governments have enacted regulation reform to change the monopolistic behaviour of rail markets. This reform has different objectives, such as competition introduction, restructuring the railway organisation, improving financial sustainability, and creating regulatory bodies. For example, the European Union (EU) issued a series of directives to govern the railway sector, which were critical in reforming the railway systems in the European countries (Bošković and Bugarinović, 2015; Bougna and Crozet, 2016; Cantos et al., 2012; Grushevska et al., 2016; Johnson and Nash, 2012; Nash et al., 2014), and Table 2.1 summarises the four Railway Packages addressed by the EU.

2.2.1.1 Competition Regulation

The railway industries were reformed and regulated in a way to introduce competition. The expected outcomes of introducing competition are to revitalise the railway sector by increasing railway transport in the modal split and improving railway service efficiency in terms of production and financial performance (Lang et al., 2010). According to Gómez-Ibáñez and de Rus (2006), there are three forms to introduce competition in the railway market. The first form is separating the railway infrastructure from train operations, with implanting track access charges. This form could be explained by changing the railway organisation to be similar to the other transport modes such as road, air, and sea transports (Preston, 2002). The second form is to privatise the railway industry by involving the private sector in its management and operation. Because railway systems in most countries were characterised as state-owned monopolies, allowing the private sector to compete with one another might contribute to reductions in the cost of the railway services. The last form is to liberalise the railway services from government control. Practically, vertical separation is commonly used to introduce competition, but this form has been adopted through a combination of liberalising and privatising the railway industry.

Another form of competition that can be applied to the railway industry is yardstick competition, which was formalised initially by Shleifer (1985). The behaviour of natural monopolistic companies to set the prices on the basis of ”cost of services” is less influenced by incentives to improve efficiency because the prices are determined based on ”prices tracking costs”. When a particular market has incentives to improve service quality, reduce costs, and promote technological innovation, Schleifer highlighted that the appropriate costs and prices might be determined by a comparison between the

TABLE 2.1: EU Railway Packages. (Source: Ranghetti (2018)).

Package no.	**Purpose**
First Railway Package (2001)	This package aimed to introduce organisational separation, a full separation between rail infrastructure and train operation. Alongside the vertical separation, the package has stated the market opening for the rail freight market at an international level. This means that open access competition for the freight market has been introduced by this package. In addition, the package addressed accounting horizontal separation at the market function level. This means that a railway operator is required to create a separate balance sheet for each rail service for the passenger and freight sectors. This directive aimed to cross-finance rail services.
Second Railway Package (2004)	The package has a directive stating a further step toward freight rail deregulation. This directive aimed to remove all market entry barriers at a domestic level for railway operators who are licenced by the EU. This requires the railway network to be open for freight movements by 2007. Other directives are interoperability and safety requirements, and the European Railway Agency (ERA) has been created to monitor and manage these two requirements.
Third Railway Package (2007)	A directive stated in this package aimed to deregulate the international passenger rail market for railway operators from 2010. Another directive is an extended development of technical harmonization. A new system has been designed to specify the minimum requirements for train drivers. Finally, the package has stated regulations for passengers' rights and service quality for the freight sector.
Fourth Railway Package (2016)	This package contains several directives organised within two pillars. Firstly, the 'Market Pillar' has multiple directives for the governance of railways and market deregulation for the passenger sector. These directives aimed to enforce the organisational separation and provide market access for new entrants. Secondly, the 'Technical Pillar' has multiple directives for interoperability and safety.

current or historical performance of similarly regulated companies. This comparison could provide a useful benchmark or ”yardstick” to set the prices and cost of services.

The yardstick competition scheme is applied in a way to improve the efficiency of railway systems by using specific indicators as described by Mizutani et al. (2009). For example, in Japan, the regulator authority uses several indicators (e.g., operating cost) to improve the performance of rail operators. This scheme is applied to review rail fares based on operators' performance. When the regulator revised rail fares, the yardstick competition scheme is applied to act as a motivation for railway companies. If a railway company operates at higher levels of operating costs compared to the standard costs, it is seen as an inefficient company, and it is expected to reduce its

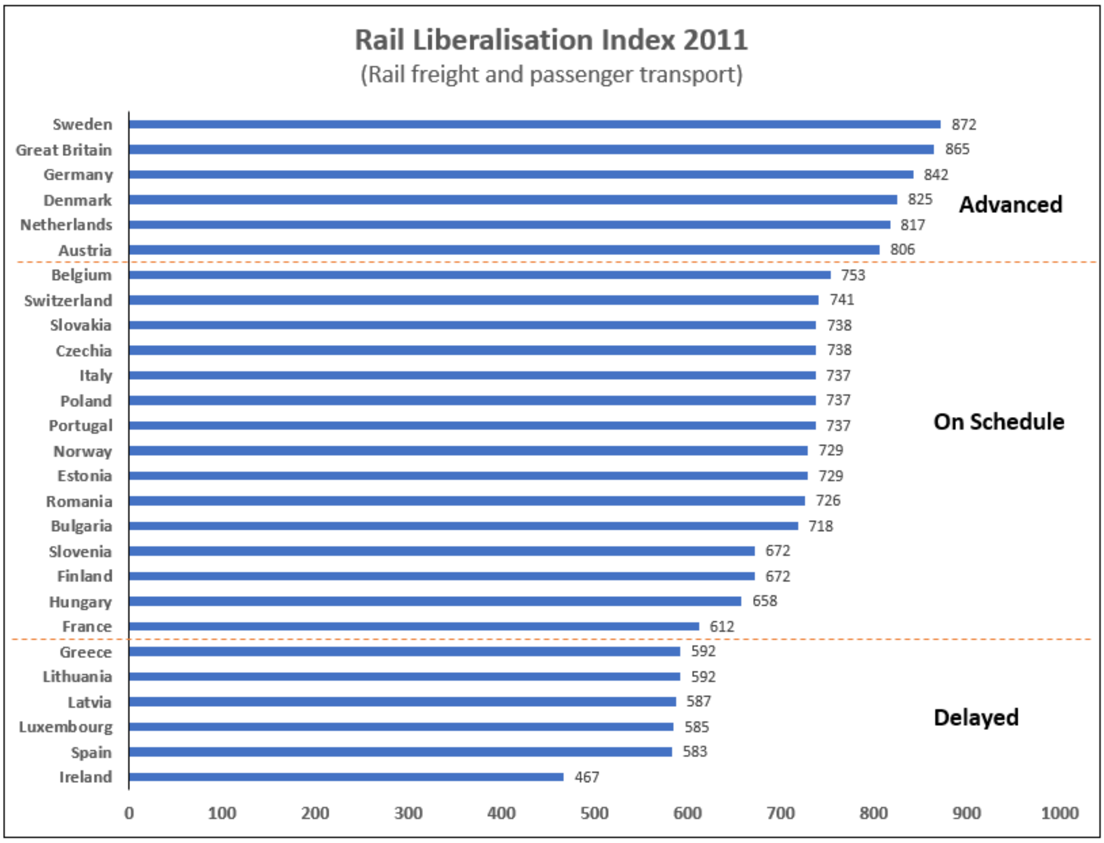


FIGURE 2.2: Rail Liberalisation Index (LIB) of the European rail market. (Source: IBM and Kirchner (2011)).

operating costs. In contrast, if a railway company operates at low levels of operating costs compared to the standard costs, it is seen as an efficient company, and it is not required for this company to take further action to reduce the cost. In fact, this efficient company is granted 50% of the difference between the operating and standard costs as an incentive.

In terms of liberalisation and market opening, IBM has published multiple reports (2002, 2004, 2007, and 2011) to measure the process of liberalising the European rail market. In the latest report, the European countries are classified into three groups based on the Rail Liberalisation Index, as shown in Figure 2.2. This classification is based on assigning the scores (from 1 to 10) to outline the progress in different areas of reform, which are clustered into: the legislative transposition of the European directives and regulations; the effective implementation of these policies; and the competitive characteristics of the markets (Smith et al., 2018).

However, the practical implementation of introducing competition in the railway industry can be clustered into three forms, as highlighted by Finger (2014) and as shown in Figure 2.3. The first form is on-track competition (or competition in the market), which means that two or more vertically integrated rail companies compete against each other to serve railway transport between the same designations. An

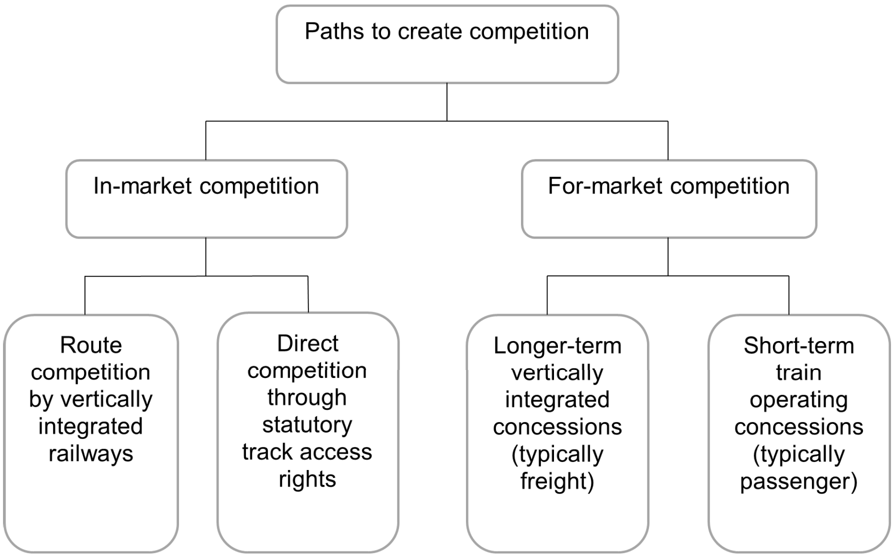


FIGURE 2.3: The main paths to introduce the competition in the rail market. (Source: World Bank (2017)).

example of this form is freight railway services in the United States. The second form is off-track competition (competition for the market), in which two or more train companies compete for train services under a regime of regulated access to the track infrastructure. In addition, this form is designed mainly for competitive tendering. Regional passenger rail services in some EU countries are an example of off-track competition. The last form of competition is open-access competition, in which two or more train companies compete against each other to serve passenger and freight transport under a regime of regulated access to the track infrastructure. Examples of this form are the freight rail services in EU countries, most rail services in Australia and some passenger rail services in Austria, the Czech Republic, Germany, Great Britain, Italy and Sweden.

2.2.1.2 Structure Regulation

Governments have enacted a railway reform to change the behaviour of monopolies in the form of restructuring the railway organisation. Furthermore, the structural reform is intended to facilitate the railway systems to introduce different forms of competition. The purpose of this reform is to restructure the railway organisation in two dimensions, horizontally and vertically, as discussed above. This vertical separation between rail infrastructure and train operations creates two bodies. Infrastructure managers (IMs) are often the owners of rail infrastructure represented by a government entity or a private company awarded a specified contract. IMs are responsible for infrastructure maintenance, investment in capacity expansion, safety, and providing a well-prepared track for train operations (Finger and Messulam, 2015). On the other hand, Railway Undertakings (RUs) (also called Train Operating Companies (TOCs)) are responsible for providing rail services for rail users. A detailed discussion of the structural reform is provided in Section 2.2.2.

2.2.1.3 Economic Regulation

Economic regulation aims to improve the financial sustainability of the railway sector. The best practice for a railway market is marginal social cost pricing (World Bank, 2017). This means that passenger and freight rail services are priced at a marginal social cost. This condition aims to achieve maximum overall economic welfare for the whole community. However, practically, this condition is hard to achieve by a single railway market for the following reasons, as stated by the World Bank (2017):

- In most railway markets, rail service prices do not include external charges. This case does not comply with the concept of economic welfare.
- In the short-run, the variable costs of a railway market are less than the total costs. Thus, marginal cost pricing will lead to a financial deficit. Even if the marginal cost pricing is applied in the long-run, the revenue is not sufficient to cover the operating costs.
- The term ‘margin’ refers to a relatively small unit of output. According to this definition, the margin of the railway market is a single passenger seat-km or wagon-km of freight. In practice, the output increments at which prices may be legitimately established are significantly larger, such as for a class of service, a class of trains, a regular commodity movement, or a specific freight shipper.

As a result of the structural reform, it is necessary to design a framework to regulate the economic relationship between IMs and RUs. This framework could help to classify areas of potential improvement in financial performance for IMs and RUs. However, income sources and expenses are the two essential components for IMs and RUs, and a detailed discussion of each component, as well as how each operator might maximise profit, is provided below.

For IMs, the major sources of income are Track Access Charges (TACs) paid by RUs and government subsidies (Finger and Messulam, 2015). TACs are regulated by policymakers and/or rail regulators. This charge is regulated in a way to utilise the rail network more efficiently and to provide fair access for RUs. TACs, however, are structured into two forms (Lang et al., 2010). The first form is a charge based on network usage (e.g., train-kilometre or gross-ton-kilometre). The second form is the charge calculated by a mixture of fixed and variable prices (two part tariffs). According to Nash (2012), EU policy outlines that TACs are computed based on short-run marginal societal cost, and TACs should avoid any discriminatory mark-ups to fulfil financial obligations. Moreover, TACs should include congestion, environmental costs, marginal tear and wear costs, and scarcity costs. Congestion costs are added to represent the cost of running an extra train, which may create additional delays for others. Marginal wear and tear costs are calculated based on the specifications of track and rolling stock. Scarcity costs are included as a cost to prevent RUs from running an extra train when the network capacity reaches its maximum. In addition, scarcity charges aim to improve the usage and management of rail infrastructure. Scarcity costs are charged as a reservation fee when a rail service provider reserves a route, and this route will be unavailable to other providers. This charge is intended to eliminate any route reservations made by service providers just in case they need them. In Germany, for example, the TACs are calculated using the following criteria (Link, 2004; World Bank, 2017):

- A base charge that is associated with the category and utilisation of track;
- A product factor charge which reflects the degree of track allocation in terms of priority;
- A regional factor charge;
- Several multiplicative or additive surcharges for higher weights, special trains, etc.

Governments, on the other hand, subsidise IMs to account for financial losses that are not covered by TACs. Generally, the purpose of subsidies is to fund capacity expansions and building projects, as well as infrastructure operations and performance enhancement.

Besides income sources, the expenses of a typical IM could be outlined into four categories, which are operating costs, renewal work, infrastructure maintenance, and depreciation. The operating costs involved, for example, traffic control centres and traffic controllers. The cost of renewal works includes upgrading tracks and signalling systems, etc. However, IMs can achieve profit maximisation in several ways, as highlighted by Finger and Messulam (2015). First, IMs can charge high TACs on routes that have high levels of traffic. Second, IMs can cut maintenance and operating expenses by using a large-scale maintenance contract and an automated control system. Last, IMs might stop operations on lines that are either unprofitable or require high operating costs.

Unlike IMs, RUs are often regulated by safety and competition authorities rather than directly by rail regulators. The major sources of income for RUs vary depending on whether they provide passenger or freight services. For the passenger sector, RUs earn revenue from rail users (e.g., commuters, regional and long-distance passengers) and government subsidies. On the other hand, RUs are normally not subsidised in the freight market, which indicates that income sources mainly depend on rail users (i.e., shippers). This requires RUs to be more efficient in their operations (e.g., input sources should be allocated efficiently to minimise operating costs). Thus, RUs that provide freight rail services take full responsibility for their success.

The essential components in the RUs' economic model are the cost of TACs and rolling stock. As mentioned above, TACs are regulated, and in certain countries, they contribute 25% to 30% of the total cost of RUs for long-distance services (Finger and Messulam, 2015). Although rolling stocks are not usually regulated, they do have technical standards depending on a number of factors. The technical requirements of rolling stock are specified by rail infrastructure and safety regulations. Rail infrastructure determines the technical specifications of rolling stock (e.g., dimensions, speed, acceleration, and power voltage). However, rolling stocks can be considered the highest source of expenditure for RUs (20% to 30% of fixed costs) (Finger and Messulam, 2015). Due to the technical specifications, the railway industry is unlike other transport modes where transport stock, such as cars and trucks, can be sold or replaced, whereas rolling stock is difficult to sell or replace. This means that RUs are solely responsible for supplying rolling stocks. However, forming a Rolling Stock Company (ROSCO) as a rolling stock provider is an alternate approach for transferring risk. The rolling stock can then be leased from a ROSCO or an infrastructure operator for a set period of time (e.g., contract length).

2.2.1.4 Industry Regulation

Alongside the enactment of the introduction of competition and other forms of railway reform, different regulatory bodies were created. The essential objectives of these bodies are to monitor railway performance and ensure fair competition and access to rail infrastructure. The regulatory bodies, as stated in EU directives, for example, should be independent of the government. This could be explained by the protection of the railway industry from 'change in mind' in government actions. For example, the Office of the Rail Regulator (ORR) (recently changed to the Office of Rail and Roads (ORR)) was established at the start of the privatisation of British Rail to control the infrastructure operator, while the Office of Passenger Rail Franchising (OPRAF) was established to regulate passenger rail franchises (Nash and Smith, 2011). In France, the Regulatory Authority for Railway Activities (ARAF) was created to regulate the coordination between rail infrastructure and train operators. The role of ARAF is to monitor traffic on the railway network and provide suggestions to the Ministry of Transport. However, ARAF's responsibility is restricted to making decisions because the state is involved in the financial decisions (Quinet, 2011). This means that ARAF is not fully independent of the state to regulate the railway system. Recently, ARAF is known as ARAFER (Autorité de régulation des activités ferroviaires et routiéres).

According to Benedetto et al. (2017), there are a number of reasons for establishing a rail regulator. The first reason is to minimise transaction costs. It is argued that vertically integrated organisations are more likely to have lower levels of transaction costs than vertically separated organisations. This could be explained by the fact that IMs and RUs within a vertically integrated organisation share the same interests. In contrast, each party in vertically separated organisations has different objectives and interests. Thus, the rail regulator can play a role in minimising the transaction costs within vertically separated organisations, and the regulator should be independent of the government (political actions), especially when the government is involved in the industry as an IM or RU. The second reason is to improve railway efficiency. It is argued that IMs within vertically separated organisations have less motivation to be more efficient, but IMs within vertically integrated organisations may be more effective since they have the same interests within the parent group. Thus, the rail regulator can play a role in enforcing IMs to be more efficient in vertically separated organisations.

The third reason is promoting competition and fair access to rail infrastructure. The essential aim of the vertical separation is to create a competitive environment, and it is critical to provide fair access for RUs to the rail infrastructure. Thus, the role of the rail regulator is to monitor the competition conditions and non-discriminatory access for RUs. The last reason is to monitor and improve safety standards. For instance, the EU Railway Packages state that RUs should have a safety certificate to operate in the European railway industry. This certificate is issued by the rail regulator or other organisations that are in charge of safety control.

2.2.2 Structural Reform

The railway organisation could be restructured into two dimensions. The horizontal reformation is dealing with the rail services, and there are two forms of this reformation, as stated by Fitzová (2017). First, the core functions of railway systems can be integrated or separated horizontally, meaning that passenger and freight rail services are provided by one or different entities. The later form is called market function separation. Second, rail services are separated based on a regional division. This horizontal form has been applied to various railway systems in Europe and East Asia. This form is mainly designed for passenger rail services, and passenger rail services in Germany, Great Britain, and Japan are examples of this form (Mizutani and Nakamura, 1997; Preston, 1996; Link and Merkert, 2011).

The arguments about separating or integrating the market function separation are concerned with the profitability of the rail market. For instance, Fitzová (2017) describes freight rail services as more profitable compared to passenger rail services. Indeed, passenger services are often operated at a loss because of the need to be affordable to rail users, and the passenger services are required to be subsidised to mitigate the losses. Having said that, separating passenger services from freight services could contribute to beneficially allocating the subsidy.

Besides the horizontal reformation, the vertical dimension represents the degree of coordination between IMs and RUs in terms of operation and management, as shown in Figure 2.4. This figure shows the generic structure of the railway system. This structure presents the breakdown of the railway activities at different levels (i.e., train, sale, station and track activities). This structure does not involve the other activities that can be operated by railway companies (e.g., road transport, ferries, hotels, property development, etc). The vertical separation, for example, is the degree of the separation between IMs and RUs. According to Ranghetti (2018), vertical separation can take three forms, which are the following:

- **Accounting separation:** the IM and RU have separate accounts but are organised within the same entity.
- **Organisational separation:** the IM and RU are organised with a separation of subsidies and decision-making procedures but under a holding company (e.g., the railway organisation in Germany, Italy and France).
- **Institutional separation:** the IM and RU are managed and operated by separate entities (e.g., the railway organisation in Great Britain, Sweden, Spain, and the Netherlands).

Besides the vertical separation, the railway organisation can take the form of integration between the IM and RU. The organisation can be described as vertically integrated when the IM and RU are managed and operated by one entity. The railway systems in Japan (excluding new Shinkansen lines, which are vertically separated (Kurosaki, 2018)), Canada, and the US are examples of this organisation.

Several works of literature attempted to demonstrate the probable causes that might have led to the creation of different vertical structural forms. van de Velde et al. (2012), for instance, delivered arguments to evaluate the vertically integrated and separated organisation forms. The authors argue that separating IMs from RUs will promote competition in the railway market, which might contribute to improving performance efficiency. Thus, the vertical separation could develop a competitive rail market that has a positive impact on performance efficiency in return. Vertical integration, on the other hand, is seen to have the potential to simplify the intricate technical functioning of the rail industry by promoting close collaboration between IMs and RUs. Despite this assumption, vertical integration can result in higher total costs, less performance, and less competition in the rail industry.

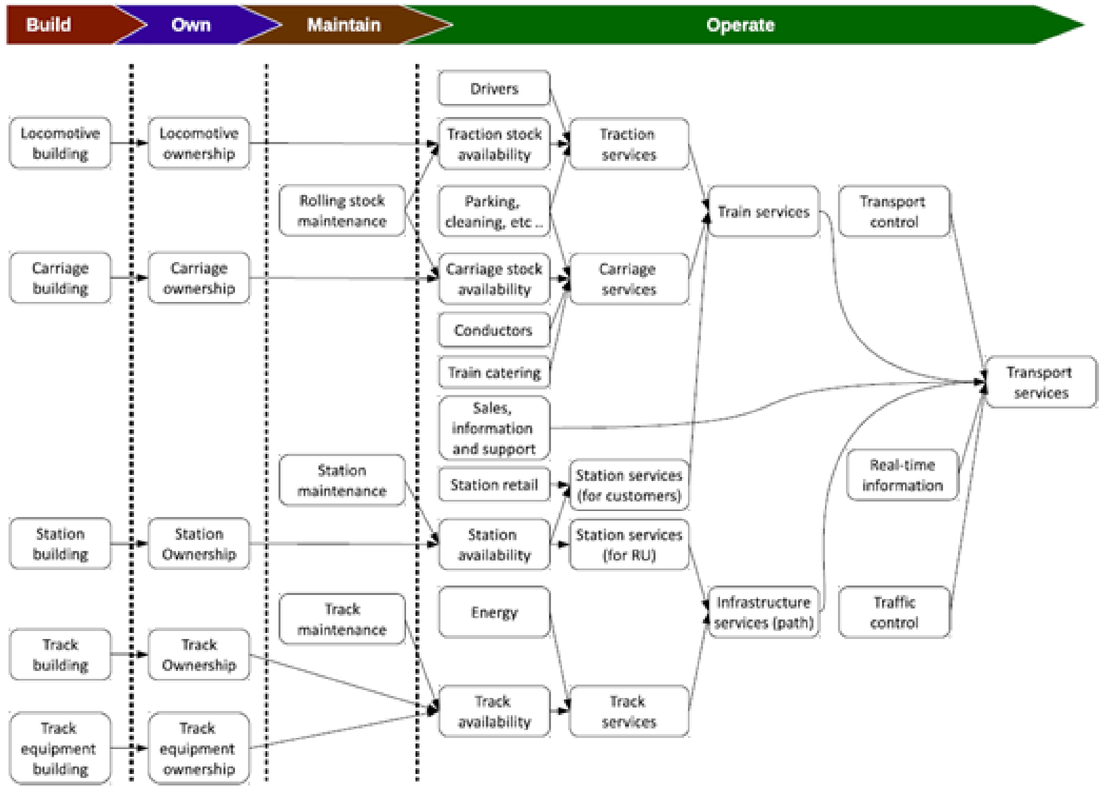


FIGURE 2.4: The structure of the railway sector. (Source: van de Velde et al. (2012)).

Another argument related to the vertical reformation is the transaction costs, as described by Growitsch and Wetzel (2009). The vertically integrated organisation might help with potential transaction cost reductions as there is a high level of cooperation between IMs and RUs in the areas of capacity allocation, security management, traffic and timetable coordination. In contrast, the vertically separated organisation with several independent operators may contribute to increasing the number of contract negotiations, which in turn raises the cost of transactions. Having said that, this argument differs in two ways. First, if real-time traffic coordination is taken into account, it is less likely that transaction costs would rise as real-time traffic is more dependent on the movement of trains within the network than on the number of operators. Second, it might be difficult to compare vertically integrated and separated organisations when long-term capacity allocation is taken into account. Moreover, the development of the rail network and train operations are both taken into account in the investment plans. For instance, the investment in high-speed rail passenger services requires the rail infrastructure to be able to run high-speed trains and to be aware of the movement of the trains on the network. Therefore, whether the organisation is vertically integrated or separated, the transaction costs at the production stage should not differ considerably, but the costs could escalate at the operation stage. For the operation stage, Merkert et al. (2012) assessed the transaction costs for three vertically separated systems. They applied a bottom-up model to estimate the transaction costs that are directly related to the interaction between IMs and RUs. They found that the transaction costs are relatively small compared to the total costs (around 1% to 2%)[1].

According to Preston (2002), vertical separation has various advantages. The first advantage is that it facilitates the political process of privatising the railway industry. Vertical separation, for example, helped the British government in defining the economic level of each component of the rail industry. Furthermore, the passenger services were chosen to be subsidised and separated into 25 TOCs, while the rest of the railway sector was determined to be managed and operated commercially. The second advantage is that the railway sector may be restructured to be more like other modes of transportation. The similarity is to provide competitive operators in the railway sector with much the same opportunity as buses and coaches in road transport.

2.2.3 Ownership Reform

Besides reforming the railway industry structurally, as discussed above, the railway reform was implemented in order to break up the state ownership of the railway industry. This breakup aims to divide the railway ownership between the public and private sectors. The breakup framework of the railway industry involves IMs remaining under state ownership in some countries, whereas the other assets and services are transferred to the private sector through privatisation. Governments apply the phenomenon of privatisation in order to enhance economic efficiency, eliminate government intervention, promote competition and market discipline, and ensure high revenue generation. With regard to the railway sector, the objectives that could be achieved from privatising the railway industry, as stated by Preston (1996), are to maximise the use of the railway system; to provide higher customer satisfaction; to provide a higher quality of rail services; and to procure value for money in the rail market.

To involve the private sector, the governments privatised the railway industry in two forms, as shown in Figure 2.5. The first form is deregulating the railway market to be free for new railway operators. For example, open-access competition requires the railway market to be liberalised and deregulated for a new entrant. The second form is explicit privatisation, where governments transfer the whole, or part of, the railway industry to the private sector. Practically, the common procedure of railway privatisation is either selling the whole or parts of the industry to the private sector or franchising the rail services to TOCs, and there are two approaches to franchising the rail services. The first approach is franchising rail services through a direct award to

[1] Merkert (2012) applied a top-down model to measure the transaction costs, and he found these costs ranged from 10% to 12% of the total costs. He pointed out that the main difference between the two approaches is that the top-down model captures the transaction costs that are not directly related to the interactions between IMs and RUs (e.g., the interaction with a ROSCO).

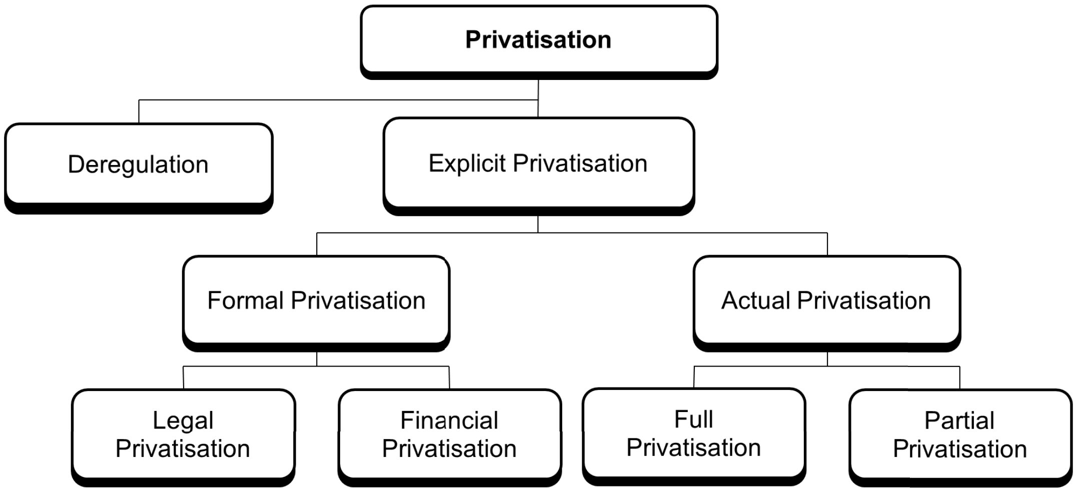


FIGURE 2.5: Forms of privatisation. (Source: Obermauer (2001)).

TOCs, which means that the regulatory authority awards railway services directly to a specific TOC without competition. This form, for example, was implemented in the Netherlands and the majority of regional passenger railway services in Germany (Link and Merkert, 2011). The second form is franchising railway services through a competitive tendering process. Moreover, the regulatory authority receives an unlimited number of bids from TOCs to be the service provider. This form, for example, is adopted in Great Britain, Sweden and parts of regional passenger rail services in Germany (Nash and Smith, 2011; Preston, 2002).

However, there are several arguments for and against privatisation. According to Bognetti and Obermann (2008), the arguments in favour of the privatisation of state-owned sectors were split into two directions. It is assumed that the entities that are owned by the state are naturally inefficient based on ineffective resource allocation and high costs. Second, liberalising the state-owned entities could create competition, which enhances cost reduction and higher service quality. On the other hand, the potential negative outcomes that could result from privatisation are a failure in controlling sectors that are strategically critical to the country, employment reductions, and involvement of foreign investors, which could contribute to transferring the profit outside the country.

Besides the above arguments, Preston (1996) highlighted a number of possible issues with franchising railway services. First, the bidder who won the contract may be ineffective. This may happen if the winner overestimates the demand for rail service or underestimates the cost. Second, it would be difficult to monitor and assess the winner's performance if the franchise agreement could not be described simply or comprehensively. Therefore, privatising the state-owned entities could revitalise the technical and financial performance of these entities, but there is a high risk of failure.

There have been both positive and negative experiences with railway privatisation. For instance, the railway industry in New Zealand experienced negative outcomes from privatisation between 1993 and 2008. In order to improve the service quality and financial performance of the railway services, New Zealand Rail was sold fully to a foreign company in 1993 (Laird, 2013; Heatley and Schwass, 2014). A few years later, some of the objectives were achieved; the share of freight rail services in the modal split increased, which in turn increased the revenue gained from rail services. Although the railway market was improving, the debt of the railway system was growing at the same time as its operating revenue was insufficient to cover its total costs. Insufficient investment also caused the rolling stock and rail infrastructure to deteriorate. The railway operator was therefore requesting subsidies from the government to maintain the rail infrastructure. However, in 2004, the New Zealand government decided to regain control of the rail network by purchasing it back rather than subsidising (Heatley and Schwass, 2014). As a result of this decision, the rail industry was reorganised into a vertically separated organisation. In 2008, the rail service was purchased back from the railway operator by the New Zealand government. According to Heatley and Schwass (2014), the government believes that the railway industry is highly important for the economic development and environmental performance of New Zealand. In addition, it was preferred to subsidise a state-owned entity rather than a foreign private company. Therefore, the responsibility of the railway industry was returned to the New Zealand government.

Besides the negative experience in New Zealand, the railway sector in Estonia also had a negative consequence due to the privatisation from 1999 to 2007. In order to privatise the railway market, the Estonian Railways was reorganised between 1997 and 1999 and divided into state-owned enterprises in the form of vertical and horizontal separation (Lust, 2017; Vare, 2011). In 1999, the railway sector was privatised and sold to international and Estonian investors. Because the government was seeking foreign strategic investors who might be able to improve the railway market to be more productive and profitable, it was preferred for the railway privatisation to be processed through international tenders (Lust, 2017). However, the Estonian government took over the responsibility for the railway industry in 2007. This decision was taken due to the deterioration in the rail services and financial performance under the management of the private sector (Lust, 2017).

In contrast to the negative experience, the following examples might be considered successful, or at least did not collapse compared to the previous two examples. According to Nash and Smith (2011) and Preston (1996), British Rail was privatised in 1997 in the form of selling and franchising the railway services and assets. As British Rail was reformed in the form of vertical and horizontal separation, the rolling stock and freight rail services were divided and sold to three and six private companies, respectively. Passenger rail services were franchised into 25 TOCs, but, currently, this number has decreased to 16. For the rail infrastructure, the current rail infrastructure operator is Network Rail, as a state-owned company. Another example is the break-up and privatisation of the Japanese National Railway (JNR) (Ishida, 2011; Kurosaki, 2018; Mizutani and Nakamura, 1997). The JNR was divided into six passenger train companies and an independent freight train company. According to Kurosaki (2018), the Railway Business Act of 1987 listed three classes of enterprise licenses, which are:

- **Class I:** Firms provide railway services by using their own rail infrastructure, for example, the passenger train companies.
- **Class II:** Firms provide railway services by using rail infrastructure owned by other organisation, for example, JR Freight Railway Company.
- **Class III:** Firms provide rail infrastructure to be handed over to Class I or to be rented by Class II.

In Germany, after merging the two state-owned railway companies in 1996, 'Deutsche Bundesbahn' and 'Deutsche Reichsbahn', Deutsche Bahn AG (DB AG) was created as holding business enterprise (Brenck and Peter, 2007). The operation and management of DB AG were divided into local and regional passenger services, long-distance passenger services, freight services and infrastructure. However, a railway reform was implemented in order to specify the profitable and non-profitable rail services, as described by Brenck and Peter (2007). It was decided for long-distance passenger and freight services to be profitable services and not subsidised, and these services remain to be provided by DB AG. On the other hand, local and regional passenger services were classified as non-profitable and subsidised services. In addition, the responsibility for these services was transferred to 33 regional authorities in order to franchise them (Brenck and Peter, 2007; Link and Merkert, 2011).

2.3 Literature Review

This section reviews the literature on estimating railway efficiency. Table 2.2 on page 33 summarises the studies that estimate the efficiency of railway systems. It should be noted that not all these studies aimed to assess the impact of railway reform on efficiency, and they are included in this research to evaluate the approaches to efficiency estimation. This table outlines the sample size of each study; the method that is used to estimate railway efficiency; the functional form that might be required based on the estimation method; and the variables that are used for the estimation. The relevant findings of these studies are discussed in Chapter 5, 6 and 7 in the literature review section. In this research, it has been found that three metrics are generally applied by the literature to estimate railway efficiency, which are technical efficiency, cost efficiency, and welfare economics.

The sample of these studies mainly targets the European railway systems. This is not surprising because these systems are facing different levels of railway reform. Some studies include railway systems from the Organisation for Economic Co-operation and Development (OECD) countries, particularly Japan and South Korea, and some European countries. There is a lack of studies that include railway systems, for example, from the Middle East and Central Asia. Thus, this research has the opportunity to assess the impact of railway reform by involving railway systems that have never been examined before.

Efficiency is a performance metric that may be measured by the percentages of producible output to input bundle. To identify if a firm is efficient or inefficient, the actual measurement of production or cost will be measured and compared to an estimated or calculated frontier to determine the efficiency, and two types of frontiers are commonly used. Firstly, the production frontier illustrates the minimum inputs required to produce outputs or the maximum outputs producible based on the given inputs. Secondly, the cost frontier illustrates the production of given outputs with respect to minimising the expenditure, the given prices of inputs, and the production technology.

However, the major purpose of these frontiers is to measure a firm's inefficiency in order to improve its performance. This optimisation can address the maximum or lowest values that can be attained given particular pricing and technology-imposed limitations, which are referred to as a limit or frontier. When the production frontier is seen as optimal, the efficiency measured is referred to as technical efficiency. If, on the other hand, the optimum is determined by the economic purpose of minimising costs, the efficiency evaluated is known as cost efficiency.

Technical efficiency (TE) is measured with respect to production technology. This technology can be described as the process of transforming inputs into outputs. Having said that, the TE can be employed to determine the possible reduction in the input bundle consumed to produce given outputs, which is called input-oriented. The TE can also be implemented to determine the possible maximisation of given outputs produced from the input bundle, and this is called output-oriented TE. Both cases are obtained by comparing the actual production with the production frontier. Thus, the technical efficiency (TE) is determined as the quotient of the maximum production determined by the production frontier ($P^{\star}$) and the actual production, and it is limited between 0 and 1. If a firm operates on the production frontier, ($TE = 1$), this firm is described as technical efficient, while a firm that operates below the production frontier ($0 \leq TE < 1$) is described as technical inefficient.

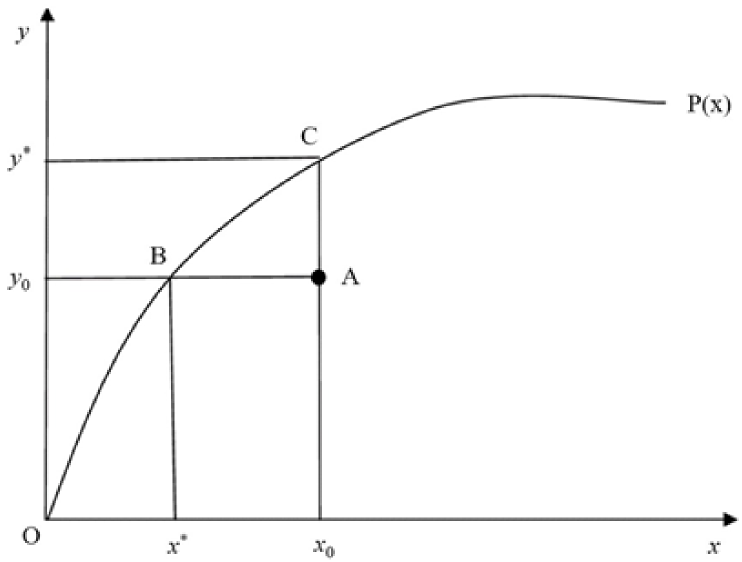


FIGURE 2.6: Technical efficiency.

Let us assume a firm consumes a single input (x_0) to produce a single output (y_0) under the production frontier $(P(x))$, as shown in Figure 2.6. The firm can reduce the input (x_0) up to $(x^\star)$ with keeping the produced output (y_0) at the same level. Thus, the input-oriented technical efficiency (TE_I) is obtained by

$$TE_I = \frac{O\,x^\star}{O\,x_0} \tag{2.1}$$

On the other hand, the firm can maximise the output (y_0) up to $(y^\star)$ at the same level of input consumption. Thus, the output-oriented technical efficiency (TE_O) is measured by

$$TE_O = \frac{O\,y^\star}{O\,y_0} \tag{2.2}$$

Cost efficiency (CE) is obtained as the quotient of the minimum cost determined by the cost frontier $(C^\star)$ and the actual cost (C), and it is limited between 0 and 1. The cost frontier is obtained to determine the minimum cost required to produce outputs. If a firm operates on the cost frontier, this firm is described as cost efficient, while a firm that operates above the cost frontier $(0 \leq CE < 1)$ is described as cost inefficient. Once the firm is determined as cost inefficient, the following step is to investigate the reasons why the firm cannot achieve the cost minimisation. The reasons could be that the firm consumes inputs more than necessary, it is technically inefficient, or the input prices are used in other than cost-minimising properties. To assess that, first the input-oriented technical efficiency (TE_I) will be measured, and then the cost allocative efficiency will be obtained as the quotient of the CE and TE_I.

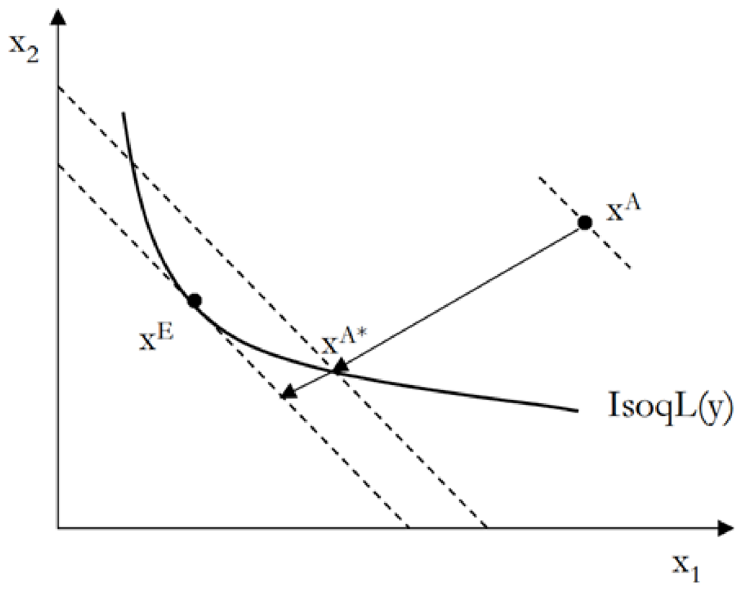


FIGURE 2.7: Cost efficiency. (Source: Cantos et al. (2002))

Let us assume firm A consumes a vector of inputs, $x = (x_1, x_2)$, priced with, $w = (w_1, w_2)$, to produce a vector of outputs, $y = (y_1, y_2)$, as shown in Figure 2.7. The observed cost of this firm is calculated as $(C^A = w_1 x_1^A + w_2 x_2^A)$ while the minimum cost is measured as $(CE = w_1 x_1^E + w_2 x_2^E)$. Thus, the cost efficiency of this firm is the ratio of CE over C^A. The distance between x^A and $x^{A\star}$ is the technical inefficiency, which represents the maximum reduction achievable in the input bundle to produce the same outputs. Technical cost inefficiency is the ratio between $C^{A\star}$, which is calculated as $(C^{A\star} = w_1 x_1^{A\star} + w_2 x_2^{A\star})$, and C^A. Allocative efficiency in terms of costs (CAE) is obtained as the quotient of the cost efficiency and technical cost inefficiency, $CAE = CE/C^{A\star}$.

Welfare, or total social surplus, is defined as the benefits for society obtained from a market or industry, which is the sum of producer and consumer surplus and externalities. Producer surplus is defined as the difference between the price producers actually receive for their goods and the price at which they are willing to sell them. Consumer surplus is defined as the difference between the amount consumers would be willing to pay for a good or service and the amount they actually pay. Externalities could be negative or positive based on their impacts on society.

To assess the impact of the change in market structure on welfare economics, Figure 2.8 outlines two situations of market conditions that can be faced by a firm under similar cost circumstances. In a monopolistic market, the situation is quite similar to the idea of producer sovereignty, which means that consumers have limited choices of what firms produce. Thus, firms have a high level of market power, and the firm can achieve profit maximisation at output Q_m, where the average revenue is always higher

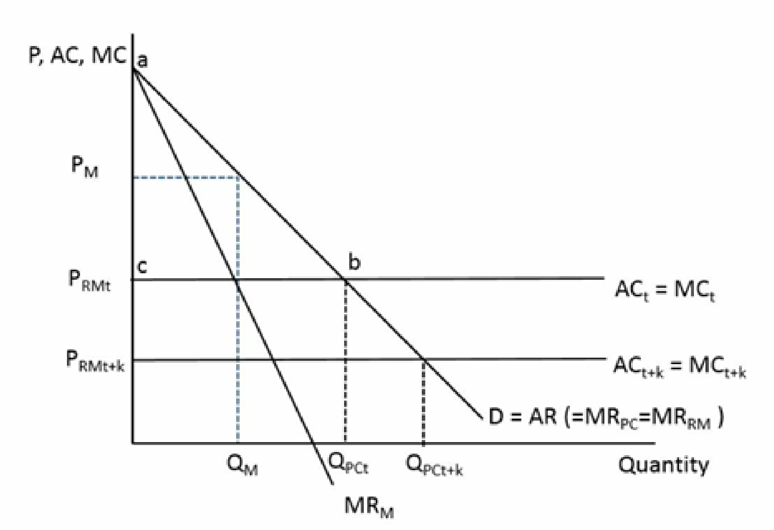


Figure 2.8: The difference between the free and regulated markets. (Source: Cowie (2017))

than the marginal revenue. In contrast, in a competitive market, the market power is transferred from the producers to the consumers, granting consumer sovereignty. In addition, the producers are forced to set the level of outputs at Q_{PCt} priced at P_{RMt}, which means that the producer surplus is equal to zero.

The methods for estimating the technical and cost efficiency of railway performance are quite debatable to identify the appropriate method, while welfare economics is the only approach to estimate welfare. For these studies, Data Envelopment Analysis (DEA) is applied to estimate the efficiency as non-parametric and deterministic. The DEA is quite demanding to estimate railway efficiency with the physical inputs and outputs, and it is less popular to estimate cost efficiency. Stochastic Frontier Analysis (SFA) is a parametric and stochastic approach to estimating railway efficiency. The application of the SFA to estimate railway efficiency is average. In fact, the SFA is more commonly applied in the literature to estimate the technical efficiency rather than the cost efficiency of railway performance. Corrected Ordinary Least Squares (COLS) is a parametric and deterministic approach. The COLS is less popular for estimating technical and cost efficiency. Seemingly Unrelated Regression (SUR) is a parametric and deterministic approach. The SUR is a generalisation approach to linear regression that estimates a set of several equations. Each equation consists of its dependent variable and, most likely, a different set of independent variables. This method is quite popular in the literature to estimate the cost function. The reason for this popularity is the capability of the SUR to accommodate and impose the input cost share equations (Shepherd's lemma equation) to estimate the cost function. These input cost share equations are presented in Figure 2.7 as the dashed lines.

Table 2.2: Summary of literature review on railway efficiency estimation.

Authors	**Sample**	**Functional form**	**Method**	**Variables**
Oum and Yu (1994)	19 OECD railway systems (1978-1989)	N.A.	DEA	Number of passengers per train; passenger traffic density; average passenger trip length; percentage of passenger train-km in total train-km; number of tonnes per train; freight traffic density; average length of haul of freight traffic; passenger train density; freight train density.
Coelli and Perelman (1996)	17 European railway companies (1979-1983)	Translog distance function and translog production	COLS	passenger-km; tonne-km; labour; energy consumption; network-line length.
Cantos et al. (1999)	17 European countries (1970-1995)	N.A.	DEA	passenger-km; tonne-km; labour; energy and material consumption; track length; coaches; wagons; locomotives.
Cowie (1999)	Swiss railway insudtry (1990-1997)	N.A.	DEA	train-km; locomotives; gradient; railcars; labour.

Table 2.2 continued from previous page

Authors	Sample	Functional form	Method	Variables
Cantos et al. (2000)	17 European railway companies (1970-1995)	N.A.	DEA	passenger-km; tonne-km; passenger train-km; freight train-km; labour; energy consumption; network-track length; locomotives; coaches; wagons
Coelli and Perelman (2000)	17 European railway companies (1988-1993)	Translog distance function and translog production	COLS	passenger-km; tonne-km; labour; rolling stocks; network-line length
Cantos and Maudos (2001)	16 European railway companies (1970-1990)	Translog cost function	SFA	total costs; pass-km; tonne-km; labour price; energy price; material price
Cantos et al. (2002)	17 European railways companies (1970-1995)	N.A.	DEA	passenger-km; tonne-km; track length; labour price; energy and materials prices
Cowie (2002)	23 of the 25 TOCs in Britain	Translog cost function	SUR	Total costs; train-km, labour costs; rolling stock leasing charges; infrastructure access charges.
Christopoulos et al. (2001)	10 railway systems (1969-1992)	Generalised McFadden flexible cost function		Total cost; train-km; capital price; labour price; energy price
Loizides and Tsionas (2002)	10 railway systems (1969-1992)	Required		Total costs; pass-km, ton-km; labour price; energy price; capital stock

Table 2.2 continued from previous page

Authors	Sample	Functional form	Method	Variables
Mizutani (2004)	Japanese railway system	Translog cost function	FIML	Total variable costs; train-km; labour price; energy price; material and maintenance price; network-line length; station spacing; number of lines; non-underground ratio.
Farsi et al. (2005)	50 Swiss railway companies (1985-1997)	Cobb-Douglas cost function	SFA	Total costs; pass-km; tonne-km; network-length; energy price; labour price; capital prices; average cost measured per pass-km.
Lan and Lin (2006)	39 worldwide railway systems (1995-2002)	Translog distance function	SFA	passenger-km; tonne-km; passenger-train-km; freight-train-km; labour; coaches; wagons.
Jensen and Stelling (2007)	Swedish rail system (1970-1999)	Translog cost function	LEA	passenger-km; tonne-km; labour cost; producer price; energy price.
Kumbhakar et al. (2007)	17 railway companies (1971-1994)	Translog distance function	SFA	passenger-km; tonne-km; labour; energy consumption; rolling stock.
Asmild et al. (2009)	23 European countries (1995-2001)	N.A.	DEA	train-km; network-line length; labour cost; material purchases.

Table 2.2 continued from previous page

Authors	Sample	Functional form	Method	Variables
Boardman et al. (2009)	Canadian National Railway (1993-2003)	Negative exponential demand function	CBA	Revenue-tonne-km; fuel price; staff cost; staff; operating expenditure; operating expenses; annual and hourly compensation; privatisation transaction cost; cost per share; no. of shares sold at privatisation.
Cantos et al. (2010)	16 European countries (1985-2005)	Distance function	DEA	passenger-km; tonne-km; labour; rolling stock (four categories); track length.
Friebel et al. (2010)	11 European countries (1980-2003)	Cobb-Douglas production function	SFA	passenger-km; tonne-km; network-line length; labour.
Cantos et al. (2012)	23 European countries (2001-2008)	Translog distance function	DEA & SFA	passenger-km; tonne-km; labour; rolling stock (four categories); track length.
Mizutani and Uranishi (2013)	23 OECD countries from Europe and East Asia (1994-2007)	Translog cost function	SUR	Total cost; train-km; passenger-km; tonne-km; wage rate; energy price; material price; capital price; route length; technology index; passenger revenue share; train density; freight cars;

Table 2.2 continued from previous page

Authors	Sample	Functional form	Method	Variables
Smith (2012)	13 European rail infrastructure managers (1996-2006)	Translog cost function	SFA	Total costs; maintenance costs; renewal costs; network-track length; passenger-train-km per route-km; freight-train-km per route-km; single-track km divided by route-km; electrified track-km divided by track-km; average salary.
van de Velde et al. (2012)	26 OECD countries (1994-2010)	Translog cost function	SUR	Passenger-km; tonne-km; route length; technology index; wage rate; energy price; material price; capital price.
Preston and Robins (2013)	British rail system (1979/80-2008/09)	Negative exponential demand function	CBA	Total costs, total revenue; pass-km; revenue per pass-km; train-km.
Mizutani et al. (2015)	26 European and East Asian countries (1994-2007)	Translog cost function	SUR	Total cost; train-km; passenger-km; tonne-km; wage rate; energy price; material price; capital price; route length; technology index; passenger revenue share; train density; freight cars.
Bougna and Crozet (2016)	27 European rail systems (1997-2011)	Translog distance function	SFA	passenger-km; tonne-km; labour; coaches, wagons; locomotive; network-track length; energy consumption.

Table 2.2 continued from previous page

Authors	Sample	Functional form	Method	Variables
Preston (2018)	British rail system (1979/80-2014/15)	Negative exponential demand function	CBA	Total costs, total revenue; pass-km; revenue per pass-km; train-km.
Smith et al. (2018)	17 European railways (2002-2010)	Translog cost function	SUR	Train-km; passenger-km; tonne-km; passenger revenue share; passenger load factor; passenger trip length; freight train length; passenger train-km; freight train-km; no. of cars per passenger train; rout length; technology index; wage rate; energy price; material price.
Lerida-Navarro et al. (2019)	27 European rail systems (1998-2012)	Cobb-Douglas production function and translog production function	DEA & SFA	passenger-km; tonne-km; labour; rolling stock (four categories); track length.
Preston and Bickel (2020)	British rail system (1979/80-2017/18)	Negative exponential demand function	CBA	Total costs, total revenue; pass-km; revenue per pass-km; train-km.

Chapter 3

Understanding the Railway System in Saudi Arabia

3.1 Introduction

This chapter reviews the railway system in Saudi Arabia and outlines the selected railway systems for delivering the research assessments. As this research aims to deliver an assessment of reforms to optimise railway organisation, this assessment involves countries from Europe, the Middle East, North Africa, and East and Central Asia. Before conducting the assessment, it is necessary to review the Saudi railway industry, which is the main case study. This review involves the current structure of the railway organisation and any future plans to develop the railway industry.

3.2 Saudi Railway Market

3.2.1 Railway Sector

The structure of the Saudi railway system is divided into three categories, as shown in Figure 3.1. The first category is the sector which is responsible for the strategic plans to develop and improve the transport infrastructure, and the MoT is responsible for this role (Public Transport Authority, 2017). The role of the MoT is to establish strategic plans to enhance and expand the infrastructure development of the transport sector as a whole, including the road network, the railway infrastructure, and the ports.

The second category is the sector who responsible for the regulation and enforcement, and the PTA is designed to be accountable to regulate and define the transport system (not only the railway system) and facilitate the public transport services for passengers within and between cities (Public Transport Authority, 2017). This means that PTA

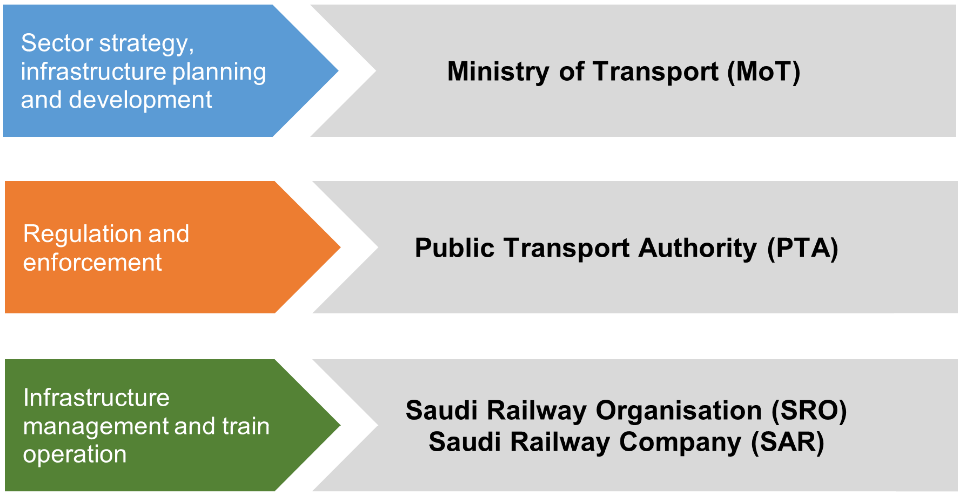


FIGURE 3.1: The current structure of the rail market in Saudi Arabia

has a major role in addressing policies for public transport. The process of addressing these policies can be explained by that PTA identifies and prepares which policies are needed for public transport, then these policies are sent for the Council of Ministers to roll out as royal decrees. These royal decrees are subsequently circulated to the relevant establishments for implementation. Moreover, PTA was founded by MoT in 2012 to support the strategic development related to the public transport within and between cities, and this support is targeting the passenger services by regulating the public transport to provide highly comfortable and reliable transport services (MoT, 2017).

The last category is infrastructure management and train operation for passengers and freight, and there are two independent bodies, which are SRO and SAR. The SRO is a public railway body owned by the MoT, and this body was established in the 1950s as the Riyadh Dammam line was constructed (Public Transport Authority, 2017). The SRO was responsible for operating and maintaining the rail infrastructure, train operations and rolling stock. This indicates that the SRO was a vertically integrated organisation. On the other hand, the SAR is a public-owned company that is owned by the PIF. The creation of the SAR occurred in 2006 in response to the strategic plan to construct the North-South rail project (Public Transport Authority, 2017).

3.2.2 Major developments

Prior to 2000, there was limited development in the railway industry in Saudi Arabia. In terms of railway construction, the Riyadh-Dammam line is the first railway project that has been constructed by the Saudi government. When the project was opened in the 1950s, it served freight transport only between Riyadh and the eastern region, and passenger rail services were introduced in the 1980s. To operate this project, SRO has been created by the government to be fully responsible for the Riyadh-Dammam line.

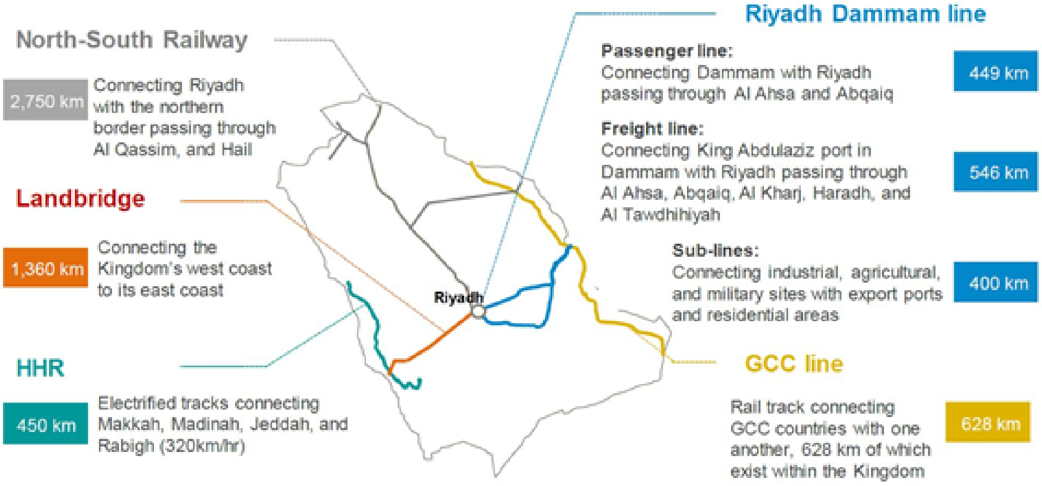


FIGURE 3.2: The current and planned projects for the railway network in Saudi Arabia. (Source: Public Transport Authority (2017))

At the beginning of the 2000s, there were some challenges faced by SRO, as summarised by the Riyadh Chamber of Commerce (2005). The first challenge is the short-haul distance for passenger and freight rail services. Furthermore, SRO operates 470 km of freight rail service. The breakeven distance could exceed the operated freight distance, which means that the operating cost is not absorbed. For example, SRO in 2004 earned nearly SR 240 million in total, while the operation of rail services cost SRO around SR 390 million. This indicates that there was a financial deficit of SR 150 million. The second challenge is the relatively low traffic density. SRO had a density of one million tonne-km per route km. This compares poorly with North America, where the density of Class 1 railways is around 22.9 million tonne-km per route km. Traffic density is important since much of the cost burden associated with rail operations is fixed. The SRO traffic base does not include high-density point-to-point movements except for container business moving to and from its captive dry port. The third challenge is high unit operating costs. The SRO's unit operating costs of SR 0.20 per tonne-km are roughly double the costs in Australia and North America, which average SR 0.09 per tonne-km (Riyadh Chamber of Commerce, 2005).

However, during this period, the Saudi government has launched a railway expansion plan. This plan consists of different railway projects to connect the substantial locations for freight and passenger transportation, and these projects are shown in Figure 3.2. The first project is the North-South Railway (NSR) project. This project was planned to transport mineral ore from new mines planned in the northeast of the country. The NSR also connects the Hazm Al Jalamid phosphate and the Az Zabairah bauxite mines to mineral processing and export facilities on the Arabian Gulf at Ras Azur, north of Al Jubail.

The second project is the East-West Railway (Saudi Land-bridge), which will provide a connection between the east and west regions. Moreover, the objectives of this project are to construct a rail line between Riyadh and Jeddah (950 km line length) and between Dammam and Jubail (115 km line length) and to update the existing infrastructure between Riyadh and Dammam (Riyadh International Convention and Centre, 2018). The project will be built by the private sector under the Public-Private Partnership (PPP) (Public Transport Authority, 2017). The last project is HHSR, which connects Makkah with Madinah, passing through Jeddah. This project is designed to transport only passengers, with a total length of 450 km.

The Saudi railway industry has been developing rapidly since 2010. These developments can be clustered into four categories, which are the expanding railway industry; regulation reforms; organisational reforms; and privatisation. Two projects have started in operation to expand the railway infrastructure. The NSR project has been launched by SAR in two stages. The first stage was the opening of the mineral line in 2011, which spans from the Al Jalamid mine to the facilities and port in Ras Al Khair with a 1392 km network length (Briginshaw, 2016). SAR has transported 23.68 million tonnes of different types of minerals, such as phosphate and bauxite, from mines located at Hazm Al-Jalameed to export ports at Ras Al-Khair between 2011 and 2017 (Briginshaw, 2017b). In February 2017, SAR introduced passenger rail services from Riyadh to Al Qassim, and these services have been expanded to Hail, eight months later (Briginshaw, 2017a).

Besides the NSR project, the HHSR project was launched in 2018 by the Saudi Spanish Train Project Company Ltd., which was awarded by SRO in 2011 (International Railway Journal, 2011). In fact, SRO has announced the contract for equipping, operating, and maintaining the HHSR, and two consortia were bidding for this contract. The first consortium was the Al-Shoula Group, which consists of one Saudi company and 12 Spanish companies. The second consortium was the Al Rajhi Group, which involves one Saudi company, SNCF and Alstom. After several negotiations with the two consortia, the final decision made by SRO was to select the Al Shoula Group for the HHSR project. The contract includes providing rolling stock and maintaining and operating the HHSR project for 12 years (International Railway Journal, 2011). Currently, the Al Shoula Group is known as the Saudi Spanish Train Project Company (SSTPC).

Alongside launching several railway projects, the Saudi government took action to regulate its railway industry in order to improve safety and performance. The Railway Sector at the PTA is responsible for regulating the railway industry. The PTA was created in 2012 by a Royal Decree to be responsible for regulating public transport (Public Transport Authority, 2018). The PTA has Road, Railway and Sea Transport Sectors, and each sector is responsible for regulating the public services of each mode. The role of the Railway Sector is summarised as follows:

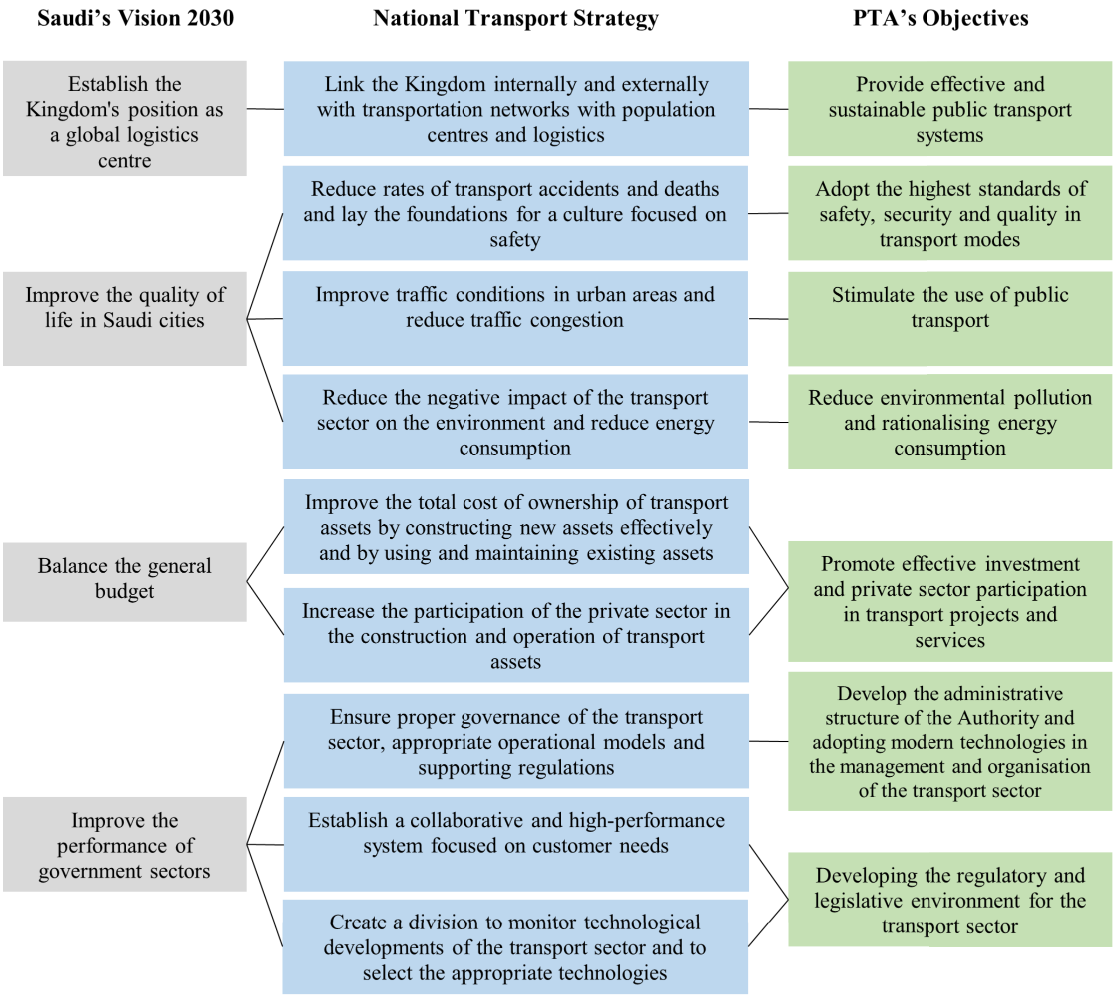


FIGURE 3.3: The national framework to develop the transport sector in Saudi Arabia. (Source: Public Transport Authority (2018)).

- Monitoring the quality, efficiency and safety of rail transport in the Kingdom in general.
- Issuing safety certificates and railway licenses for rail transport service providers.
- Developing rail services and managing the privatisation of the railway industry in partnership with the private sector internationally and locally.
- Developing the railway network within the framework of the national transport strategy to achieve the Saudi's Vision 2030, which is summarised in Figure 3.3.

With regards to safety and performance control, the Railway Sector has issued Safety Certificates (SC) and Railway Licences (RL) for incumbent operators. Table 3.1 contains details about the current licences issued to incumbent operators. It should be noted that each operator cannot provide rail services until it obtains valid SC and RL from the Railway Sector.

TABLE 3.1: The licences issued for the incumbent operators. (Source: Railway Sector (2018)).

Grantee	Scope	Type	Issued	Expires
SRO	Passenger and freight rail services	SC	15/10/2014	14/10/2019
		RL	15/10/2014	14/10/2019
SAR	Passenger and mineral rail services	SC	30/09/2017	29/09/2020
		RL	30/09/2017	29/09/2020
SSTPC	Passenger rail services	SC	11/10/2018	31/12/2018
		RL	11/10/2018	31/12/2018

The last development to discuss is the organisational reform. There are two actions that have been taken by the Saudi government for the owned railway assets. Firstly, the constructed rail infrastructure and planned railway projects have been transferred to SAR by a Royal Decree addressed in 2016 (Briginshaw, 2016; Public Transport Authority, 2017). This means that SAR became formally responsible for the full network. This decision pushes forward to privatising the railway industry, as the Saudi government considers SAR as a private company. Secondly, the Saudi government had addressed a major organisational reform in 2019 by merging SRO with SAR, which means that all of SRO's assets have been transferred formally to SAR. This action is planned to take up to three years for a complete merge (Briginshaw, 2019). In April 2021, the Saudi government announced its approval to abolish SRO and to replace it with SAR (Asharq Al-Awsat, 2021).

Besides the government actions for organisational reform, the PTA's Railway Sector is preparing a regulatory framework for the railway industry. This framework aims to facilitate the railway industry to follow the international trends in railway regulation. Moreover, the framework will introduce competition, rail service franchising, new regulations for infrastructure managers, etc. The PTA's Railway Sector has published the framework as a draft, and there is no planned date to publish the formal framework and implement it.

3.3 Other Railway Systems

The research project aims to deliver an assessment of reforms to optimise the railway organisation in the Saudi Arabian context. Different forms of railway organisation will be assumed for conducting the analysis. However, the research project performs benchmark techniques, which require several railway industries with different organisational forms to conduct the analysis. Additionally, having a large number of railway industries could contribute to the strength of the analysis. Thus, the research project selects railway industries based on the following criteria:

- **Data availability** This is the significant criteria for selecting countries because insufficient data is not adequate to conduct the analysis.
- **Countries under state-owned monopoly** The research project will take into consideration the state-owned monopolistic form of railway organisation. This requires the research project to involve countries following this organisational form.
- **Countries with vertically separated organisation** One of the assumptions of the railway organisation is the vertical separation. This leads this research to select countries with a vertically separated organisation.
- **Countries with a vertically separated organisation within a holding company** Another assumption is the vertical separation within a holding company. This means that the research project will involve countries with a vertically separated organisation within a holding company.
- **Countries with vertically integrated organisation** The last assumption is vertical integration. This requires this research to select countries with a vertically integrated organisation.

This research did not consider the difference between developed and developing countries. The variables that can be considered are related to service quality (e.g., accident rates, service reliability, etc.). This could promote extra further obstacles in selecting railway systems due to data availability. This research considers other variables that could have impacts on railway efficiency based on country characteristics (e.g., GDP per capita, traffic density, network-line density and population density). These variables are easily obtainable from online sources (e.g., the World Bank).

3.3.1 Structural Form

This section presents the structural form of the selected railway systems. As discussed in Section 2.2.2, the structural form can take vertical and horizontal dimensions. Table 3.2 contains information about the vertical and horizontal structures of the selected railway systems. According to this table, the vertical structure represents the relationship between IMs and RUs, and this relationship is structured in three forms. First, the vertical integration (VI) form means that IMs and RUs are managed and operated by one entity (e.g., Algeria, Azerbaijan, Ireland, Morocco and Turkmenistan). Second, the organisational separation (OS) form represents the vertical separation between IMs and RUs within a holding company (e.g., Austria, Belgium, Germany, Italy and Poland). Last, the institutional separation (IS) form indicates the vertical separation between IMs and RUs (e.g., Bulgaria, the Czech Republic, Slovakia, Uzbekistan). The vertical separation forms are more popular in the European

TABLE 3.2: The structural form of the selected countries. (Source: Mehmood Alam (2017), Ranghetti (2018) and van de Velde et al. (2012).

Country	Structural form					
	Horizontal structure			Vertical structure		
	HI	RS	MFS	OS	IS	VI
Algeria	1976 -	:	:	:	:	1976 -
Austria	-2002	:	2003 -	2005 -	:	-2004
Azerbaijan	1991 -	:	:	:	:	1991 -
Belgium	-2001	:	2002 -	2005 -	:	-2004
Bulgaria	-2004	:	2005 -	:	2001 -	-2001
Czech Rep.	-1994	:	1995 -	:	2003 -	-2002
Finland	-2011	:	2012 -	:	1995 -	-1994
France	-2004	:	2005 -	2015 -	1997 - 2014	-1996
Germany	-1999	2000 -	2000 -	2000 -	:	-1999
Greece	:	:	:	2007 - 2009	2010 -	-2006
Iran	1914 -	:	:	:	:	1914 -
Ireland	1994 -	:	:	:	:	1994 -
Italy	-2000		2001 -	2001 -	:	-2000
Japan	-1987	1987 -	1987 -	:	:	1987 -
Kazakhstan	1997 -	:	:	2001 -	:	-2000
Latvia	-2000	:	2001 -	2007 -		-2006
Morocco	1963 -	:	:	:	:	1963 -
Pakistan	1861 -	:	:	:	:	1861 -
Poland	-2002	:	2003 -	2001 -	:	-2000
Portugal	:	:	:	:	1997 -	-1996
Saudi Arabia	- 2010	2011 -	:	:	2016	-2015
Slovakia	2005 -	:	2006 -	:	2002 -	-2001
Slovenia	-2008	:	2009 -	:	1991 -	:
South Korea	1994 -	:	:	:	:	1994 -
Switzerland	2005 -	1999	2006 -	:	2009 -	-2008
Turkey	1929 -	:	:	:	2016 -	-2015
Turkmenistan	:	:	:	:	:	:
Uzbekistan	1994 -	:	:	:	2001 -	-2000

Notes: HI = Horizontal Integration. RS = Regional Separation. MFS = Market Function Separation. OS = Organisational Separation. IS = Institutional separation. VI = Vertical Integration. : = not applicable. n.d. = no date.

countries due to the EU Railway Packages, which aim to facilitate the railway systems to promote competition.

Besides the vertical structure, the horizontal structure represents the organisation of rail services, which can be found in three forms. First, the regional separation (RS) form indicates that the rail services are regionally separated (e.g., Germany, Japan, Saudi Arabia and Switzerland). Second, the market function separation (MFS) form means that passenger and freight rail services are separated (e.g., Austria, Belgium, Germany, Poland and Slovakia). Last, the horizontal integration (HI) form means that none of the previous two forms is adopted (e.g., Algeria, Azerbaijan, Morocco and Turkmenistan).

3.3.2 Competition Form

In this section, the competition forms of the selected railway systems are presented. As discussed in Section 2.2.1.1, there are three forms to create a competitive rail market, and Table 3.3 contains the start date of implementing each form in the selected railway systems. First, the on-track competition form represents the market deregulation for free entry for freight rail services (e.g., Austria, Belgium, Finland, Germany and Poland). It appears that Germany is the first country that introduced this competition form for the freight sector. In addition, the remaining European countries have introduced this form due to the first EU Railway package. Second, the off-track competition form denotes the franchise system for passenger rail services (e.g., Austria, Belgium, France, Germany and Italy). Last, the open-access competition form represents the market deregulation for free entry for passenger services (e.g., Austria, the Czech Republic, and Italy). It should be noted that this form is not considered in the assessments of this research because it is not popular in this sample.

It should be noted that this research mentions the yardstick competition alongside other forms of competition that can be implemented in the railway industry. This competition form is not popular widely, whereas the indirect competition forms (e.g., on-track and off-track competition) are commonly adopted for the railway systems.

3.4 Conclusion and comments

The Saudi government aims to develop its railway industry with strategic plans and a series of railway reforms. As discussed above, the government has a strategic plan to expand its rail infrastructure, and two projects have already been launched. In the beginning, all railway projects were planned to be under SRO responsibility, except the NSR project, where the government created SAR. It should be noted that the Saudi government aimed to build railway projects with the participation of the private sector. This indicates that the government started with ownership reform to improve the functionality of its railway industry.

With regards to the railway organisation, the government addressed several reforms to its rail market. Before creating SAR, the railway industry was organised in the form of vertical integration without any form of competition, where SRO was the monopolistic rail operator. When SAR was created, the railway organisation was restructured in the form of horizontal separation, but this separation was on the basis of regional divisions

Table 3.3: The competition form of the selected countries. (Source: Mehmood Alam (2017), Casullo (2016), Ranghetti (2018) and van de Velde et al. (2012).)

Country	**Competition form**		
	ONTC	**OFTC**	**OAC**
Algeria	:	:	:
Austria	2003 -	2003 -	2011 -
Azerbaijan	:	:	:
Belgium	2002 -	2010 -	:
Bulgaria	2005 -	:	:
Czech Rep.	1995 -	:	2011 -
Finland	2012 -	:	:
France	2005 -	2011 -	:
Germany	1995 -	1997 -	:
Greece	:	:	:
Iran	:	:	:
Ireland	:	:	:
Italy	2001 -	2009 -	2011 -
Japan	:	:	:
Kazakhstan	:	:	:
Latvia	2007 -	2001 -	:
Morocco	:	:	:
Pakistan	:	:	:
Poland	2003 -	2004 -	:
Portugal	:	1999 -	:
Saudi Arabia	:	:	:
Slovakia	2006 -	2012 -	:
Slovenia	2009 -	:	:
Spain	2007 -	:	:
South Korea	:	:	:
Switzerland	1999 -	:	:
Turkey	:	:	:
Turkmenistan	:	:	:
Uzbekistan	:	:	:

Notes: ONTC = On-track competition for freight sector. OFTC = Off-track competition for passenger sector. OAC = Open access competition for passenger sector.

rather than market function. This means that the railway industry is separated horizontally between SAR and SRO without any form of competition. The first form of competition that has been introduced by the government is competition for the market. This form was introduced for the HHSR project as competitive tendering under the BOT scheme. When SRO was merged with SAR, the organisation changed to be regionally separated, as the HHSR is operated by SSPTC, without any form of vertical separation.

The last comment to mention is the railway industry regulator. The government has created the PTA as the public transport regulator. One of the PTA's roles is to issue railway licences and safety certificates for incumbent railway operators. This indicates that the government follows the international trend to monitor and improve railway performance and safety standards.

As the railway system in Saudi Arabia faces a series of changes in its organisation (e.g., regional and institutional separation), this research aims to create a sample that can incorporate railway systems with different organisational forms. The sample consists of several railway systems from Europe, the Middle East, North Africa, and Central and East Asia. The availability of the data played a major role in selecting this sample. There are some railway systems that are included because they have some degree of similarity with the Saudi railway systems, which are relatively comparable. For example, the railway systems (e.g., Algeria, Iran and Morocco) are included to evaluate the state-owned monopoly before the creation of SAR. Railway systems that are separated on a regional basis (e.g., Germany and Japan) are included to assess the regional division between SRO and SAR before the merger. The institutionally separated railway systems (e.g., Bulgaria, Portugal, Slovakia, and Uzbekistan) are included to examine the Saudi railway system when SAR became responsible for the rail infrastructure of the Riyadh-Dammam line. On the other hand, the organisationally separated railway systems (e.g., Austria, Belgium, Germany, Italy and Poland) are included to evaluate the second form of vertical separation. The European railway systems (e.g., Austria, Belgium, Germany, France, Spain and Slovakia) are included to represent the competition forms. Lastly, this research will not be limited to European railway systems but will also include systems from other regions (e.g., Africa and Asia).

Chapter 4

Methodology

4.1 Introduction

This chapter outlines the framework of the research methodology. As one of this research's objectives is to determine the technical, cost, and welfare efficiency of rail services, Figure 4.1 outlines the framework for delivering this research. The procedure for conducting the technical and cost assessments is relatively similar, but the objective of each assessment is different. As stated in Section 1.3.2, different organisational forms will be taken into consideration to achieve optimisation. The technical and cost efficiency will be assessed to select the optimum reform option. Furthermore, the assessment will identify the most efficient railway reform option that can achieve maximum performance in terms of technical efficiency. On the other hand, cost efficiency will be measured in order to identify the optimum railway reform that can achieve cost minimisation. This research employs econometric techniques to estimate railway efficiency in delivering these two assessments.

For the welfare assessment, this research defines the welfare of rail rail services by calculating the welfare and its components of the railway system. Moreover, this research develops a framework built on the concept of welfare economics. Through these calculations, different forms of railway organisations will be examined by assessing impact of the change in the railway market with respect to railway reform. Then, this research will selects the optimal railway organisation that can maximise the welfare.

4.2 Efficiency Measurement Techniques

This section reviews the econometric methods that can be applied to estimate the efficiency of railway systems. These methods can be classified and clustered at different

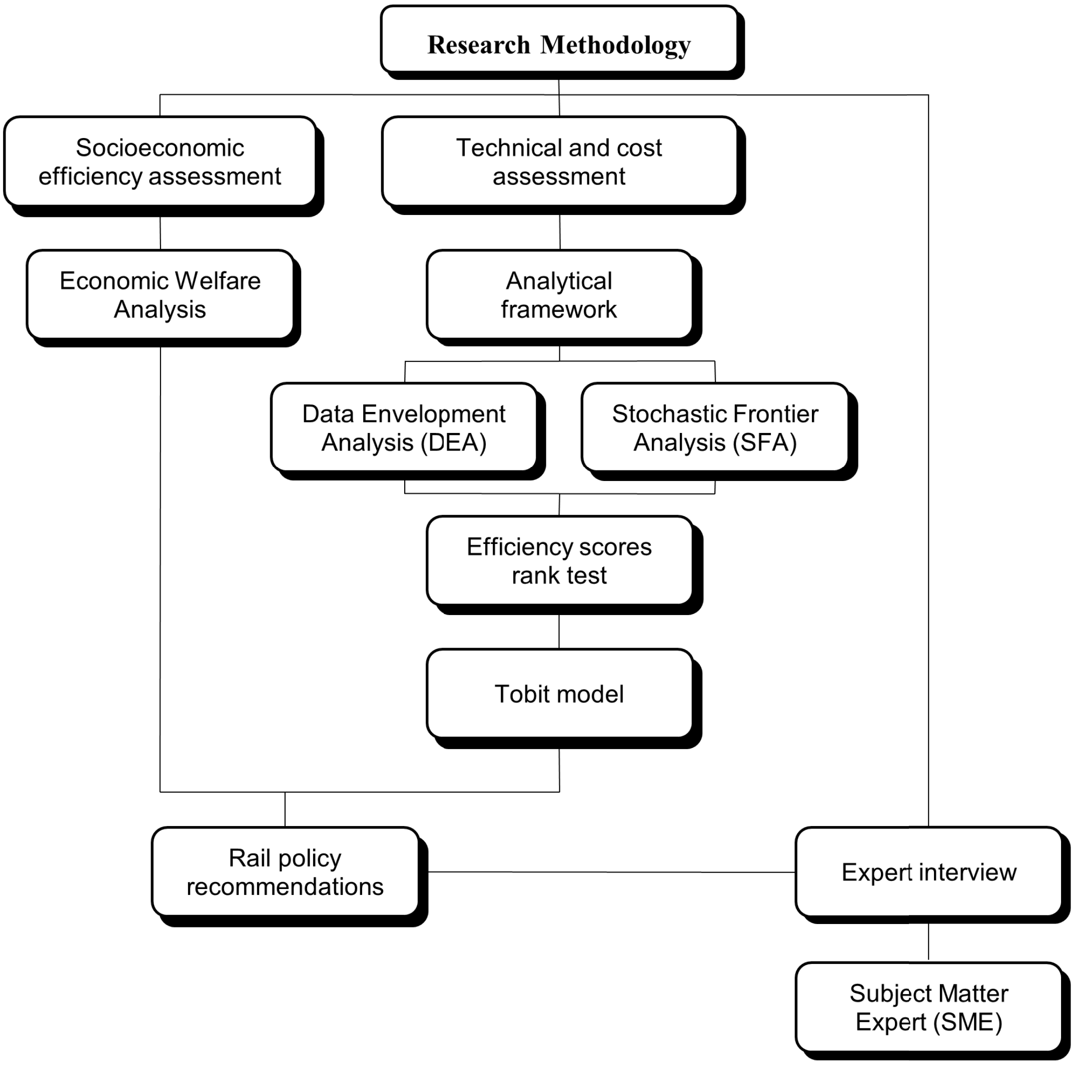


FIGURE 4.1: The framework of the research methodology

levels. First, the econometric methods can be described as parametric, which employs several assumptions for the functional form to estimate the frontiers, while the non-parametric approaches rely on fewer assumptions for the functional form to estimate the frontier. Second, the econometric methods can be defined as deterministic, which measures the deviation between the frontier and the observed measurement (e.g., production or cost) as pure efficiency or inefficiencies, while the stochastic methods allow for data randomness and noise that could have an impact on the performance.

However, there are various approaches that can measure performance efficiency, as summarised by Oum et al. (1999). The simple methods to determine efficiency are Partial Factor Productivity (PFP) and Total Factor Productivity (TFP). PFP refers to the ratio of a firm's output to a single input. The issue of using a single input and output does not reflect the overall performance efficiency as the railway is a multi-product industry. In addition, PFP is not capable of estimating the production or cost frontier. Besides PFP, TFP is measured as the quotient of a total output quantity and a total input quantity. This can reflect the nature of the railway as a multi-product industry, but TFP is not applicable to estimate the production or cost frontier to measure the technical and cost efficiency. Additionally, TFP is not capable of accommodating multiple firms for comparison purposes. Therefore, the simple methods to determine the efficiency are not appropriate for the research project.

Besides the simple methods, econometric methods can be used to estimate a cost or production function. The further step is to determine the change in performance efficiency through the estimated cost or production functions. However, the econometric methods can accommodate multiple inputs and outputs and multiple firms, but these methods assume that all firms are successful in reaching the frontier, which is not always true. This indicates that the econometric methods are not applicable to estimate the production and cost frontiers to determine the technical and cost efficiency. Therefore, the research will rely on methods that can estimate the production and cost frontier, fit multiple inputs and outputs, and accommodate multiple firms for comparison purposes.

Before discussing the frontier methods, it should be mentioned that there are two approaches that are widely used to obtain the production and cost frontiers, as stated by Brage-Ardao (2017). The first approach is a non-parametric approach, which relies on fewer assumptions about the functional form to calculate the frontier. The second approach is a parametric approach, which employs several assumptions for the functional form to estimate the frontiers. However, another classification of these techniques is deterministic and stochastic methods. For deterministic methods, the deviation between the frontier and the observed measurement (e.g., production or cost) is interpreted as pure efficiency or inefficiencies, while stochastic methods allow for data randomness and noise that could have an impact on the performance.

This research considers several aspects in selecting the appropriate methods to determine performance efficiency. The first aspect is the capability to accommodate multiple inputs and outputs. As the railway is a multi-product industry, this research applies approaches that can accommodate multiple inputs and outputs. The second aspect is estimating the production and cost frontiers. As discussed in the previous section, technical and cost efficiency are measured as a comparison between the actual production and cost measurements of firms with the production and cost frontier. This means that this research employs techniques that can estimate the frontier in order to deliver the comparisons. The last aspect is involving multiple firms to deliver a comparison. This research covers railway systems at international level, and it is necessary to implement methods that are designed to compare multiple firms. Having said that, this research will apply Data Envelopment Analysis (DEA), as a non-parametric and deterministic approach; Stochastic Frontier Analysis (SFA), as a parametric and stochastic approach; and Corrected Ordinary Least Squares (COLS), as a parametric and deterministic approach.

4.3 Production Technology

4.3.1 Production Function

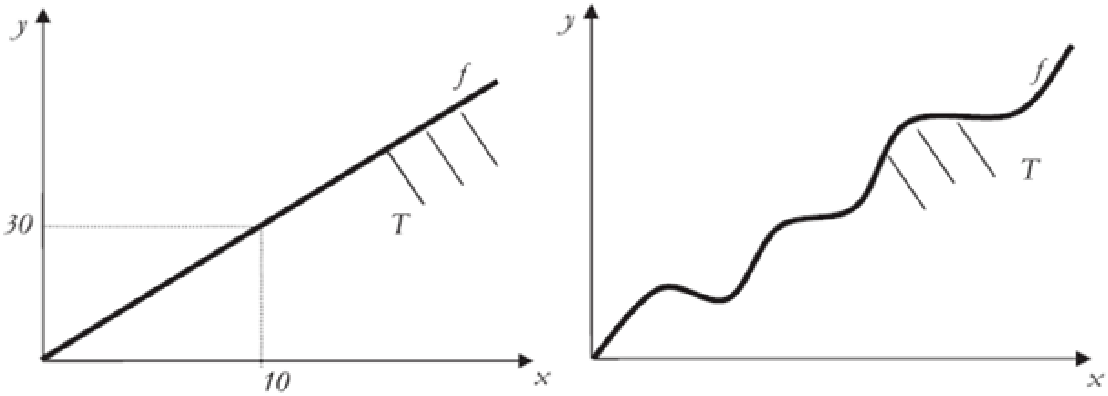
The production function describes the technology used by a firm to produce outputs. This technology can be defined as a process of transforming inputs (x) into outputs (y). Moreover, all firms in a particular environment consume one or more inputs (i.e. K-inputs) to produce one or more outputs (i.e. M-outputs). To represent this process in a multidimensional framework, it assumes that inputs (x) are a vector in the nonnegative real space of some finite dimension K i.e. denoted as $x = (x_1, x_2, \ldots, x_K) \in R_+^K$, and outputs (y) are a vector in the nonnegative real space of some finite dimension M i.e. denoted as $y = (y_1, y_2, \ldots, y_M) \in R_+^M$. Schematically, Figure 4.2 shows a process of typical production technology that can be analysed.

FIGURE 4.2: A typical production process. (Source: Sickles and Zelenyuk (2019)).

From the economics point of view, the key question is how to represent the production technology of a typical firm that can also fit other firms. Thus, this technology is assumed and characterised as the technology set denoted as T, and it can be written in a general form as

$$T = \Big\{(xy) \in R_+^K \times R_+^M : y \text{ is producible from } x\Big\} \tag{4.1}$$

That is, T is a set composed of all possible pairs (x, y) such that the second vector of the pair, y, is producible from the first vector of the pair, x. This technology, denoted by the expression 'y is producible from x', may vary in different environments, firms, or even within the same firm at different times or circumstances. When conducting empirical research, these conditions could be defined in detail; however, for the purposes of theory illustration, it is more practical to maintain this technology as general as possible. Consider a simple example of a production technology that consume a single unit of input to produce three units of outputs; the corresponding



(a) Example of a technology set.

(b) Example of a technology set.

FIGURE 4.3: A representation of a technology set. (Source: Sickles and Zelenyuk (2019)).

technology set can be expressed as $T = \left\{(xy) \in R_+^K \times R_+^M : y \leq 3\right\}$ and is shown in Figure 4.3a. Meanwhile, Figure 4.3b depicts an alternative technology set that can take any form. Figures 4.3a and 4.3b illustrate the technology set with all possible combinations of inputs and outputs (x, y) that lie on or below the f curve (technology frontier). Outside of this technology set, all points will be characterised as technologically impossible.

The production technology has another characterisation, which can be given via the output set $(P(x))$ corresponding to any particular level of inputs. To formulate this, it can be written as

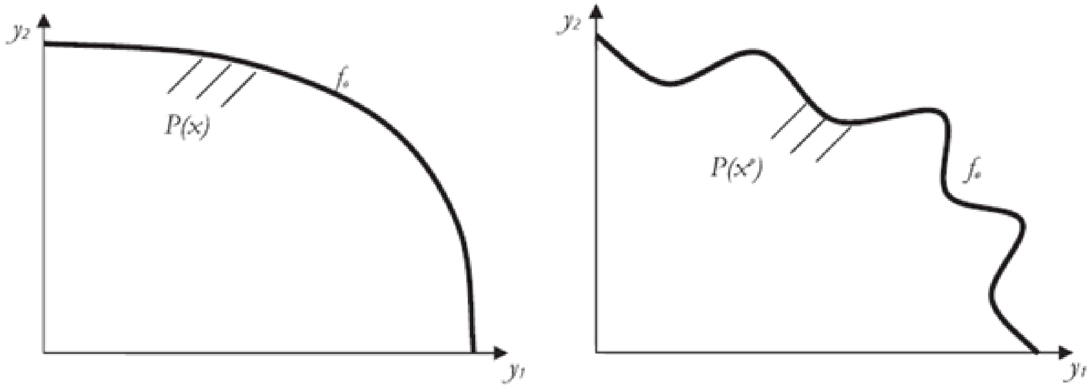
A second description of the production technology is provided by the output set $(P(x))$ corresponding to any distinct level of inputs. To express this, it could be written as

$$P(x) = \left\{y \in R_+^M : y \text{ is producible from } x\right\}, x \in R_+^K \tag{4.2}$$

$P(x)$ provides all possible output combinations that can be generated by $x \in R_+^K$. Figure 4.4a and 4.4b depict two distinct output sets, each containing all combinations of outputs (y) that are located on or below the output-technology frontier (f_O) curve. Again, all points inside the set are characterised as technologically feasible, while all points outside the set are defined as technologically impossible.

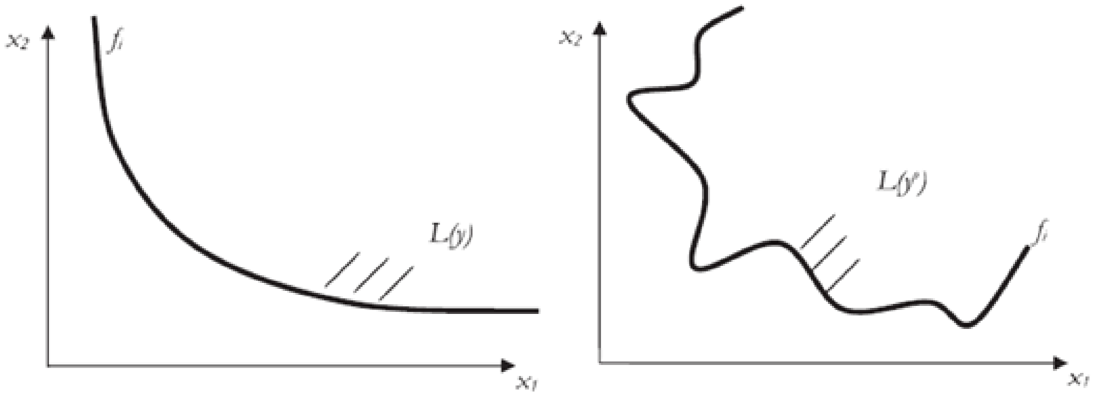
In addition, the production technology has an additional input-related characteristic, which is the input requirement sets corresponding to any distinct output level. This is expressed by the input-correspondence $L(y)$, which represents the input requirement sets as

$$L(y) = \left\{x \in R_+^K : y \text{ is producible from } x\right\}, y \in R_+^M \tag{4.3}$$



(a) Example of a technology set. (b) Example of a technology set.

FIGURE 4.4: A representation of an output set. (Source: Sickles and Zelenyuk (2019)).



(a) Example of an output set. (b) Example of an output set.

FIGURE 4.5: A representation of a technology set. (Source: Sickles and Zelenyuk (2019)).

In words, for each level of outputs $y \in R_{+}^{M}$ requires an input set of all possible combinations of inputs that can achieve this particular level of outputs, and this is shown in Figure 4.5a and 4.5b.

4.3.2 Distance Function

To model a production technology with multiple outputs, econometric studies often employ several techniques, as highlighted by Coelli and Perelman (2000). The possible approach is to create a single output index by combining multiple outputs (e.g., aggregate revenue). On the other hand, the technology can be modelled with a dual cost function, which incorporates both the production function and the cost function within a minimisation framework. However, these approaches require a number of assumptions and additional data that must be taken into consideration. In the first approach, it is assumed that the firm's behaviour aims to maximise revenue, while in the second, the firm aims for cost minimisation. The distance function is an alternative function for the multiple output case. This function is simply the transformation function or the Production Possibility Function (PPF). The distance function can be applied without price data or explicit behavioural assumptions. This benefit makes the distance function more applicable to the multiple output production cases than other approaches, which require behavioural assumptions and price data to estimate the production function by modelling both technical and allocative efficiency (Coelli and Perelman, 2000; Kumbhakar et al., 2015).

Let us start by defining the production technology of a firm using the output set $(P(x))$, which is stated in Equation (4.2). $P(x)$ is assumed to satisfy the axioms listed in Färe and Primont (1995), which are as follows:

1. $0 \in P(x)$.
2. Non-zero output levels cannot be produced from a zero level of inputs.
3. $P(x)$ satisfies strong disposability of outputs; that is, if $y \in P(x)$ and $y^* \leq y$, then $y_* \in P(x)$.
4. $P(x)$ satisfies strong disposability of inputs; that is, if y can be produced from x then y can be produced from any $x^* \geq x$.
5. $P(x)$ is closed;
6. $P(x)$ is bounded; that is, $P(x)$ has an upper boundary, which called the frontier, and this belongs to $P(x)$.
7. $P(x)$ is convex.

Shephard (1970) defines the output-distance function on $P(x)$ as

$$D_O = \min\{\theta : (y/\theta) \in P(x)\} \tag{4.4}$$

where θ is the ratio to minimise the distance between the observed outputs and the frontier, as shown in Figure 4.6. Lovell et al. (1994) describe the output distance function as follows:

1. $D_O(x, y)$ is non-decreasing in y and non-increasing in x;
2. $D_O(x, y)$ is linearly homogeneous and convex in y;
3. $D_O(x, y) \leq 1$, if $y \in P(x)$;
4. $D_O(x, y) = 1$, if $y \in Isoq\,P(x) = \{y : y \in, \omega \cdot y \notin P(x), \omega > 1\}$.

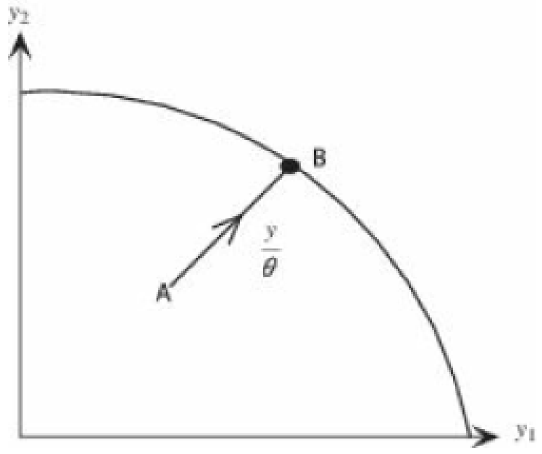


FIGURE 4.6: Output Distance Function: Single Input and Two Outputs.

In summary, a producer is described as efficient if it is located on the frontier or isoquant. On the contrary, a producer is described as inefficient if it is located inside the frontier. From linear homogeneity, we can obtain $D_O(x, \omega \cdot y) = \omega \cdot D_O(x, y)$, for any $\omega > 0$. One can arbitrarily select one of the outputs (say the Mth output) and set $\omega = 1/y_M$. Then, $D_O(x, y/y_M) = D_O(x, y)/y_M$. Thus, if we adopt the standard flexible translog form, the deterministic output distance function can be written as

$$\begin{aligned} \ln(D_{Oi}/y_{Mi}) = \alpha_0 &+ \sum_{m=1}^{M-1} \alpha_m \ln(y^*_{mi}) + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{n=1}^{M-1} \alpha_{mn} \ln(y^*_{mi}) \ln(y^*_{ni}) \\ &+ \sum_{k=1}^{K} \beta_k \ln(x_{ki}) + \frac{1}{2}\sum_{k=1}^{K}\sum_{l=1}^{K} \beta_{kl} \ln(x_{ki}) \ln(x_{li}) \\ &+ \sum_{m=1}^{M-1}\sum_{k=1}^{K} \delta_{mk} \ln(y^*_{mi}) \ln(x_{ki}) \end{aligned} \tag{4.5}$$

where $y^*_{mi} = y_{mi}/y_{Mi}$ and $i = (1, 2, \dots, N)$ is the number of observations. Let

$$\ln(D_{Oi}/y_{Mi}) = TL(x_i, y_{mi}/y_{Mi}; \alpha, \beta, \delta) \tag{4.6}$$

Or,

$$\ln(D_{Oi}) - \ln(y_{Mi}) = TL(x_i, y_{mi}/y_{Mi}; \alpha, \beta, \delta) \tag{4.7}$$

Hence,

$$-\ln(y_{Mi}) = TL(x_i, y_{mi}/y_{Mi}; \alpha, \beta, \delta) - \ln(D_{Oi}) \tag{4.8}$$

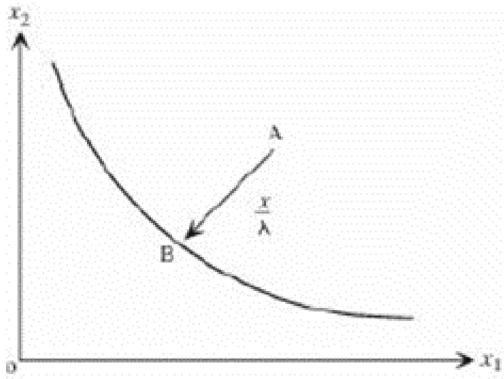


FIGURE 4.7: Input Distance Function: Two Inputs and a Single Output.

In a similar way, the input distance function can be defined on the input set $(L(y))$, that is stated above in Equation (4.3), as

$$D_I = \max\{\lambda : (x/\lambda) \in L(y)\} \tag{4.9}$$

where λ is the ratio to minimise the distance between the observed inputs and the frontier, as shown in Figure 4.7. Lovell et al. (1994) also describes the input distance function as follows

1. $D_I(x, y)$ is non-decreasing in x and non-increasing in y;
2. $D_I(x, y)$ is positively linearly homogeneous and concave in x;
3. $D_I(x, y) \geq 1$, if $x \in L(y)$;
4. $D_I(x, y) = 1$, if $x \in IsoqL(y) = x : x \in L(y)$.

The concept is similar to the output distance function to describe if the producer is efficient or not. From linear homogeneity, we obtain $D_I(\omega \cdot x, y) = \omega \cdot D_I(x, y)$, for any $\omega > 0$. One can arbitrarily select one of the inputs (for example, the Kth input), and set $\omega = 1/x_K$, then $D_I(x/x_K, y) = D_I(x, y)/x_K$. Thus, the deterministic input distance function can be written in a translog form as

$$\begin{aligned}\ln(D_{Ii}/x_K) = \alpha_0 &+ \sum_{m=1}^{M} \alpha_m \ln(y_{mi}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln(y_{mi}) \ln(y_{ni}) \\ &+ \sum_{k=1}^{K-1} \beta_k \ln(x^*_{ki}) + \frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1} \beta_{kl} \ln(x^*_{ki}) \ln(x^*_{li}) \\ &+ \sum_{m=1}^{M}\sum_{k=1}^{K-1} \delta_{mk} \ln(y_{mi}) \ln(x^*_{ki})\end{aligned} \tag{4.10}$$

where $x^*_{ki} = x_{ki}/x_{Ki}$ and $i = (1, 2, \ldots, N)$ is the number of firms. Let

$$\ln(D_{Ii}/x_{Ki}) = TL(y_{mi}, x_{ki}/x_{Ki}; \alpha, \beta, \delta) \tag{4.11}$$

or,

$$\ln(D_{Ii}) - \ln(x_{Ki}) = TL(y_{mi}, x_{ki}/x_{Ki}; \alpha, \beta, \delta) \tag{4.12}$$

Hence,

$$-\ln(x_{Ki}) = TL(y_{mi}, x_{ki}/x_{Ki}; \alpha, \beta, \delta) - \ln(D_{Ii}) \tag{4.13}$$

The first description of the output and input distance function can be checked via the monotonicity condition (Last and Wetzel, 2010). To be monotone, the output distance function should be non-decreasing in outputs and non-increasing in inputs at each observation, while the input distance function should be non-increasing in outputs and non-decreasing in inputs at each observation. To check this condition for the outputs, the distance function is derived with respect to the outputs as follows

$$\epsilon_m = \frac{\partial \ln(D)}{\partial \ln(y_m)} = \alpha_m + \sum_{m=1}^{M} \alpha_{mn} \ln(y_m) + \sum_{k=1}^{K} \delta_{mk} \ln(x_k) \tag{4.14}$$

The monotonicity condition for the output distance function is satisfied when Equation (4.14) is non-negative for all outputs $\epsilon_m > 0$, while the input distance function satisfies the condition when Equation (4.14) is non-positive for all outputs $\epsilon_m < 0$. For the inputs, the distance function is derived with respect to the inputs as follows

$$\epsilon_k = \frac{\partial \ln(D)}{\partial \ln(x_k)} = \beta_k + \sum_{k=1}^{K} \beta_{kl} \ln(x_k) + \sum_{m=1}^{M} \delta_{mk} \ln(y_m) \tag{4.15}$$

The monotonicity condition for the output distance function is satisfied when Equation (4.15) is non-positive for all inputs $\epsilon_k > 0$, while the input distance function satisfies the condition when Equation (4.15) is non-negative for all inputs $\epsilon_m < 0$.

The concavity condition, which is the second description of the output and input distance function, can be checked via the Hessian matrix ($\hat{H}$). For the output distance function, it requires the Hessian matrix of second order derivatives of the distance function with respect to the outputs to be negative semi-definite. $\hat{H}$ is defined as

$$\hat{H} = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{k1} & \dots & \alpha_{kl} \end{bmatrix} - \begin{bmatrix} \epsilon_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \epsilon_m \end{bmatrix} + \begin{bmatrix} \epsilon_1\epsilon_1 & \dots & \epsilon_1\epsilon_n \\ \vdots & \ddots & \vdots \\ \epsilon_m\epsilon_n & \dots & \epsilon_m\epsilon_m \end{bmatrix} \tag{4.16}$$

where α_{mn} are the second-order coefficients of the outputs and ϵ_m are are the output elasticities, which is written in Equation (4.14).

For the input distance function, it requires the Hessian matrix of second order derivatives of the distance function with respect to the inputs to be negative semi-definite. $\hat{H}$ is defined as

$$\hat{H} = \begin{bmatrix} \beta_{11} & \dots & \beta_{1l} \\ \vdots & \ddots & \vdots \\ \beta_{k1} & \dots & \beta_{kl} \end{bmatrix} - \begin{bmatrix} \epsilon_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \epsilon_k \end{bmatrix} + \begin{bmatrix} \epsilon_1\epsilon_1 & \dots & \epsilon_1\epsilon_l \\ \vdots & \ddots & \vdots \\ \epsilon_k\epsilon_l & \dots & \epsilon_k\epsilon_k \end{bmatrix} \tag{4.17}$$

where β_{kl} are are the second-order coefficients of the inputs and ϵ_k are the input elasticities, which is written in Equation (4.15).

The second description of the output and input distance function also states to be linearly homogeneous in outputs and inputs, respectively. That is, the linear homogeneity in outputs is given if

$$\sum_{m=1}^{M} \alpha_m = 1; \qquad \sum_{m=1}^{M} \alpha_{mn} = 0; \qquad \sum_{m=1}^{M} \delta_{mk} = 0, \tag{4.18}$$

and linear homogeneity in inputs is given if

$$\sum_{k=1}^{K} \beta_k = 1; \qquad \sum_{k=1}^{K} \beta_{kl} = 0; \qquad \sum_{k=1}^{K} \delta_{mk} = 0. \tag{4.19}$$

This distance function will be employed in the technical assessment for the following reasons. First, the objective of this assessment is to estimate the technical efficiency. Second, the railway industry is a multi-product market, which requires a function that can accommodate multiple outputs. Last, this assessment will estimate the technical efficiency via a parametric approach, which requires a functional form to represent the relationship between the inputs and outputs. Thus, the distance function fits the purpose of this assessment.

4.3.3 Cost Function

It has been discussed that the production function cannot accommodate multiple outputs. One of the solutions to dealing with multiple outputs is to assume the cost minimisation behaviour. Thus, the cost function is a tool to calculate the cost of a firm by including a set of input quantities and prices with an assumption of cost minimisation behaviour. In general, a cost function is a mathematical description of the relationship between outputs, input prices and other explanatory variables and can be written as follows:

$$C = f(y, w, z) + \varepsilon \tag{4.20}$$

where C is total cost; y is a vector of outputs; w is a vector of input prices; and z is a vector of explanatory variables. Equation (4.20) can be re-written to formulate the cost minimisation problem as follows

$$C(x, y) = \min_{x} \{wx : x \in L(y)\} \tag{4.21}$$

where x is a vector of inputs; and $L(y)$ is input-correspondence stated in Equation (4.3). The explanatory variables (z) are dropped for simplicity. It is important to highlight that not every relationship represents a proper cost function, which justifies the hypothesis of cost minimisation. Therefore, a proper cost function is characterised as follows

- $C(y, w) \geq 0$; This means that the total cost (C) is nonnegative. This is true because w and x vectors are nonnegative.
- Homogeneous of degree one in input prices (also unknown as linear homogeneity). This characteristic is explained below in detail.
- Monotonic in outputs (y). This condition means that an increase (decrease) in output levels should always lead to increase (decrease) in total cost. To represent this condition mathematically, this is given by

$$\frac{\partial C}{\partial y} > 0. \tag{4.22}$$

- Monotonic in input prices. Let w and w‘ denote vectors of input prices such that $w \geq w'$. Nondecreasing in input prices implies that

$$C(y, w') \geq C(y, w), \tag{4.23}$$

that is, an increase in the price of an input cannot cause the total production cost to fall. In addition, because the derivatives of this cost function result in optimal factor demands, this restriction is further warranted; that is, it is the result of Shephard's lemma, which states the optimal factor demands, x_i, are derived from

$$\frac{\partial C_i}{\partial w_i} = x_i(w, y) \geq 0. \tag{4.24}$$

- Concavity in input prices. This characteristic is explained below in detail.

To explain these characterisations, recall Equation (4.21) and rewritten as follows

$$\ln(C_i^A) = \ln(C^*(w_i, y_i)) + \varepsilon_i \tag{4.25}$$

where C_i^A is the actual cost; $\ln(C^*(w_i, y_i))$ is the frontier cost function; and ε_i is the error term. If the cost function is assumed to take a translog form, Equation (4.25) can be rewritten as follows

$$\begin{aligned}\ln(C_i^A) = {} & \alpha_0 + \sum_{m=1}^{M} \alpha_m \ln(y_{mi}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln(y_{mi})\ln(y_{ni}) \\ & + \sum_{j=1}^{J} \beta_j \ln(w_{ji}) + \frac{1}{2}\sum_{j=1}^{J}\sum_{l=1}^{J} \beta_{jl} \ln(w_{ji})\ln(w_{li}) \\ & + \sum_{m=1}^{M}\sum_{j=1}^{J} \delta_{mj} \ln(y_{mi})\ln(w_{ji}) + \varepsilon_i\end{aligned} \tag{4.26}$$

One of the cost function's characterisations is price homogeneity. This means that the cost function is homogenous of degree 1 in the input prices, and the cost function should satisfy the following additional parameter restrictions

$$\sum_{j=1}^{J} \beta_j = 1; \qquad \sum_{j=1}^{J} \beta jl = 0; \qquad \sum_{j=1}^{J} \delta_{mj} = 0. \tag{4.27}$$

To formulate the price homogeneity, let us assume the dataset consists of a single-output and two input prices. The cost function can be written in a translog form as follows

$$\begin{aligned}\ln(C_i^A) = {} & \alpha_0 + \alpha_y \ln(y_i) + \frac{1}{2}\ln(y_i)^2 + \beta_1 \ln(w_{1i}) + \beta_2 \ln(w_{2i}) + \frac{1}{2}\beta_{11}\ln(w_{1i})^2 \\ & + \beta_{12}\ln(w_{1i})\ln(w_{2i}) + \frac{1}{2}\beta_{22}\ln(w_{2i})^2 + \delta_{y1}\ln(y_i)\ln(w_{1i}) \\ & + \delta_{y2}\ln(y_i)\ln(w_{2i}) + \varepsilon_i\end{aligned} \tag{4.28}$$

The constraints to satisfy the price homogeneity are as follows

$$\begin{aligned} \beta_1 + \beta_2 =& 1; \\ \beta_{11} + \beta_{12} =& 0; \\ \beta_{22} + \beta_{12} =& 0; \\ \delta_{y1} + \delta_{y2} =& 0; \end{aligned} \tag{4.29}$$

Equivalently, these constraints could be re-arranged as follows

$$\begin{aligned} \beta_1 =& 1 - \beta_2; \\ \beta_{11} =& - \beta_{12}; \\ \beta_{22} =& - \beta_{12}; \\ \delta_{y1} =& - \delta_{y2}; \end{aligned} \tag{4.30}$$

If these constraints are substituted into Equation (4.28), the price homogeneity restrictions will be built into the model, and the equation can be rewritten as follows

$$\begin{aligned} \ln\left(\frac{C_i^A}{w_{1i}}\right) =& \alpha_0 + \alpha_y \ln(y_i) + \frac{1}{2}\alpha_{yy} \ln(y_1)^2 + \beta_2 \ln\left(\frac{w_{2i}}{w_{1i}}\right) + \beta_{22} \ln\left(\frac{w_{2i}}{w_{1i}}\right)^2 \\ &+ \delta_{y2} \ln(y_i) \ln\left(\frac{w_{2i}}{w_{1i}}\right) + \varepsilon_i \end{aligned} \tag{4.31}$$

Other characterisations of the cost function are monotonicity and concavity. In production theory, the cost function should be monotonic and concave in outputs and input prices. To satisfy the monotonicity condition, the cost is required to be nondecreasing in input prices and outputs, as shown in Equation (4.32) and Equation (4.33), respectively.

$$C_i^*(w_i^1, y_i) \geq C_i^*(w_i^0, y_i) \quad \text{if} \quad w_i^1 \geq w_i^0 \tag{4.32}$$

$$C_i^*(w_i, y_i^1) \geq C_i^*(w_i, y_i^0) \quad \text{if} \quad y_i^1 \geq y_i^0 \tag{4.33}$$

Having said that, the monotonicity condition can be checked by deriving the cost function with respect to outputs and input prices. Equation (4.34) and Equation (4.35) show the partial derivative of the cost function with respect to input prices and outputs, respectively.

$$\epsilon_j = \frac{\partial \ln(C_i^*)}{\partial \ln(w_i)} = \beta_j + \sum_{j=1}^{J} \beta_{jl} \ln(w_j) + \sum_{m=1}^{M} \delta_{mj} \ln(y_m) \tag{4.34}$$

$$\epsilon_m = \frac{\partial \ln(C_i^*)}{\partial \ln(y_i)} = \alpha_m + \sum_{m=1}^{M} \alpha_{mn} \ln(y_m) + \sum_{j=1}^{J} \delta_{mj} \ln(w_j) \tag{4.35}$$

The monotonicity condition is satisfied when Equation (4.34) and Equation (4.35) are non-negative for all input prices ($\epsilon_j > 0$) and outputs ($\epsilon_m > 0$), respectively.

The second condition is the concavity on input prices. This condition is satisfied only if the Hessian matrix ($\hat{H}$) of the second-order of the cost function with respect to input prices is negative semidefinite (Diewert and Wales, 1987). $\hat{H}$ is defined as

$$\hat{H} = \begin{bmatrix} \beta_{11} & \dots & \beta_{1l} \\ \vdots & \ddots & \vdots \\ \beta_{j1} & \dots & \beta_{jl} \end{bmatrix} - \begin{bmatrix} \epsilon_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \epsilon_j \end{bmatrix} + \begin{bmatrix} \epsilon_1\epsilon_1 & \dots & \epsilon_1\epsilon_l \\ \vdots & \ddots & \vdots \\ \epsilon_j\epsilon_l & \dots & \epsilon_j\epsilon_j \end{bmatrix} \tag{4.36}$$

where β_{jl} are the second-order coefficients of the input price and ϵ_j are the elasticities of input price, which is written in Equation (4.34).

As the cost efficiency will be estimated via a parametric approach, this approach requires a functional form that is capable of representing the cost minimisation framework. Moreover, in the cost assessment, a cost model will be developed to estimate the costs of railway systems by including the railway outputs and input prices. In a similar way, a cost model will be created for the welfare assessment to determine the impact of railway policies on railway costs.

4.4 Parametric Approaches

4.4.1 Corrected Ordinary Least Squares

The Corrected Ordinary Least Squares (COLS) is built on the estimation of the Ordinary Least Squares (OLS). Thus, it should be discussing the procedure of the OLS estimation and its assumptions. Let us assume the model to be estimated is constructed as follows

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \varepsilon_i \tag{4.37}$$

where y_i is the dependent variable; x_{in} is the independent variables; ε_i is the error term; and β_N are the model parameters to be estimated. To perform the OLS estimation, the following assumptions should be satisfied:

- There is a linear relationship between y_i and x_{in}.
- ε_i have a population mean with an expected value of zero (i.e., E(ε_i)=0).
- There is no correlation between ε_i and x_{iN}.
- ε_i have a constant variance (i.e., E(ε_i)= σ^2 homoskedasticity).
- ε_i have no serial correlation (i.e., E(ε_i ε_s) $= 0$ for all $i \neq s$ no autocorrelation).
- There is no perfect linear relationship between x_{iN} (i.e., no multicollinearity).

For the econometric estimation, the first assumption can be satisfied via adapting a functional form (e.g., Cobb-Douglas or translog).

The COLS technique is an early approach to estimate the production frontier and proposed by Winsten (1957). This approach is deterministic, which means the composed error includes only the inefficiency and does not capture the statistical noise. The procedure of COLS consists of three stages to estimate the production function in the Cobb-Douglas form, which are as follows:

1. The OLS is used to estimate the relationship of production technology based on prespecified or assumed functional forms such as Cobb-Douglas or Translog.
2. The intercept of the OLS is adjusted and corrected. This adjustment will shift the estimated average production function to be above all the observations, as shown in Figure 4.8. The estimated function is treated as a technology frontier.
3. The inefficiency for each observation is estimated as the distance from the observed measure to the estimated frontier.

To formulate these stages mathematically, let us assume that the production function takes the Cobb-Douglas form, which can be written as

$$\ln y_i = \beta_0 + \beta_1 \ln x_{i1} + \beta_2 \ln x_{i2} + \cdots + \beta_N \ln x_{iN} + \varepsilon_i \tag{4.38}$$

where y_i is a single output, $x_i = (x_{i1}, x_{i2}, \ldots, x_{iN})$ is a vector of inputs for production unit i, $\beta = (\beta_0, \beta_1, \ldots, \beta_N)$ is the vector of parameters (common for all i) that need to be estimated, $i = 1, 2, \ldots, N$ is the number of observation and ε_i is statistical noise. The first step is to estimate the parameters by OLS, and the parameters will be

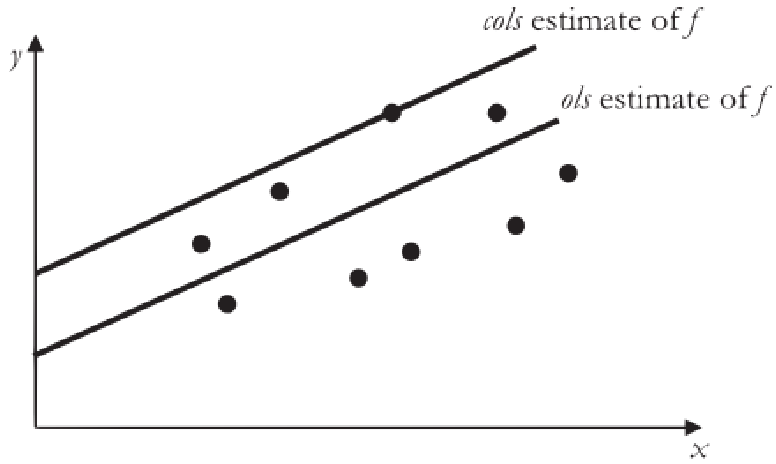


FIGURE 4.8: Average and frontier production function for a linear technology. (Source: Sickles and Zelenyuk (2019)

denoted as $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_N)$. The second step is to adjust the parameter of OLS intercept, which will be the correct intercept as

$$\hat{\alpha}_0^{COLS} = \hat{\alpha}_0 + \max_i\{\hat{\varepsilon}_i\} \tag{4.39}$$

where $\hat{\varepsilon}_i$ are OLS residuals. The third step is to estimate the Technical Inefficiency (TIN) for each observation as the deviation from the frontier or equivalently as the difference between the maximal OLS residual and the OLS residual corresponding to a particular observation as

$$TIN = \hat{\varepsilon}_0^{COLS} = \max_i\{\hat{\varepsilon}_i\} - \varepsilon_i \tag{4.40}$$

The last step is to calculate the technical efficiency for each observation as the exponent of the corrected OLS residual $(\hat{\varepsilon}_0^{COLS})$ as

$$TE = \exp\{\hat{\varepsilon}_0^{COLS}\} \tag{4.41}$$

As the technical and cost assessments will estimate the efficiency of railway systems via a parametric approach, the COLS method will be applied as an alternative method for the parametric approaches in the case that the SFA does not fit the data.

4.4.2 Stochastic Frontier Analysis

This section reviews the main models of the Stochastic Frontier Analysis (SFA). The SFA is becoming an increasingly common method to estimate firm efficiency. This method was first introduced by Meeusen and van den Broeck (1977) and Aigner et al. (1977). The SFA is a parametric and stochastic method, which allows for randomness and noise in the data or other external factors affecting the production process. Like other parametric methods, the SFA requires a functional form (e.g., production or cost functions) to represent the technology process.

The discussion of the SFA starts with the estimation of the production function. The models of the SFA are divided based on the data structure, whether cross-sectional or panel-data, as presented in Table 4.1. For the cross-sectional models, the production function can be written in Cobb-Douglas form as follows:

$$\ln y_i = \alpha_0 + \sum_{i=1}^{k} \beta_i x_{ki} + \varepsilon_i \tag{4.42}$$

$$\varepsilon_i = v_i - u_i \tag{4.43}$$

where y_i is the output; x_{ki} is the vector of inputs; and α and β are model parameters to be estimated. The composed error (ε) in Equation (4.42) consists of statistical noise (v) and inefficiency (u), as shown in Equation (4.43). Each element of the composed error has specific assumptions in terms of statistical distribution. These assumptions are required to estimate v and u by several methods, such as maximum likelihood (ML), which requires that the density function of ε is known. The ε density function can be determined through distributional assumptions that are assigned for v and u. For example, the half-normal specification of the Aigner et al. (1977) model and the exponential specification of the Meeusen and van den Broeck (1977) model (along with the normal assumption for v), generate a density function for ε, which can be applied for the maximum likelihood estimation. To conclude, v and u can be assumed to follow a specific statistical distribution. Kumbhakar et al. (2017) highlighted that the statistical noise component (v) is widely accepted to follow a normal distribution in both applied and theoretical work. On the other hand, the distributional assumption for the inefficiency component (u) is widely debatable. Kumbhakar and Lovell (2000, p.90) found evidence that supports the findings of Ritter and Simar (1997), who discuss that half-normal or exponential distribution should be assigned for the inefficiency component rather than a more truncated-normal or gamma.

The estimation of the SFA started with cross-sectional data. When repeated measures of inputs and outputs are available, then more useful information about inefficiency can be obtained, mainly by measuring the changes in efficiency or productivity through time. This feature is absent in cross-sectional models (Kumbhakar et al., 2017; Sickles and Zelenyuk, 2019). However, there are some studies that apply these models on panel-data (e.g., Coelli and Perelman (1996), Farsi et al. (2005), Lan and

Lin (2006) and Bougna and Crozet (2016)). These studies neglect the panel effect and estimate the inefficiency as a pooled model.

When the SFA was designed to make use of repeated observations, there were concerns regarding how these models coped with inefficiency over time. The models proposed by Pitt and Lee (1981) and Schmidt and Sickles (1984) to accommodate the panel-data are time-invariant. These models assume that inefficiency varies between firms but remains constant over time. It is difficult to accept the assumption that efficiency remains constant over time, especially in competitive markets (Kumbhakar and Lovell, 2000, p.90). However, Cornwell et al. (1990), Kumbhakar (1990), Battese and Coelli (1992) and Lee and Schmidt (1993) addressed this shortcoming by developing stochastic frontier models that produce time-varying inefficiency.

Another classification of the stochastic panel-data models is how to treat the heterogeneity. For the panel-data models, two methods can be performed to estimate Equation (4.44) based on the assumption of how to treat the relationship between α_0 and x_{kit}. The fixed-effects model assumes a correlation between α_0 and x_{kit}. On the other hand, the random-effects model assumes no correlation between α_0 and x_{kit}.

$$\ln(y_{it}) = \alpha_0 + \sum_{i=1}^{K} \beta_k \ln(x_{kit}) + v_{it} - u_{it} \tag{4.44}$$

Table 4.1: Summary of stochastic frontier models.

Authors	Model type[1]	Time effect[2]	Model specification	Estimation method[3]
Aigner et al. (1977)	CSM	-	1) $v_i \sim iidN(0, \sigma_v^2)$. 2) $u_i \sim iidN^2(0, \sigma_u^2)$, nonnegative half-normal. 3) v_i and u_i distributed independently of each other, and of the regressors.	ML
Meeusen and van den Broeck (1977)	CSM	-	1) $v_i \sim iidN(0, \sigma_v^2)$. 2) $u_i \sim iid$ exponential. 3) v_i and u_i distributed independently of each other, and of the regressors.	ML

Table 4.1 Summary of stochastic frontier models (continued).

Authors	Model type[1]	Time effect[2]	Model specification	Estimation method[3]
Stevenson (1980)	CSM	-	1) $v_i \sim iidN(0, \sigma_v^2)$. 2) $u_i \sim iidN^2(\mu, \sigma_u^2)$, nonnegative half-normal. 3) v_i and u_i distributed independently of each other, and of the regressors.	ML
Pitt and Lee (1981)	PDM	TIM	1) $\varepsilon_{it} = v_{it} - u_i$. 2) $v_{it} \sim iidN(0, \sigma_v^2)$. 3) $u_i \sim iidN^2(0, \sigma_u^2)$, nonnegative half-normal.	ML
Schmidt and Sickles (1984)	PDM	TIM	1) $\hat{u}_i = \max \alpha_i^* - \alpha_i^*$. 2) $TE = \exp(-\hat{u}_i)$.	GLS
Battese and Coelli (1988)	PDM	TIM	1) $\varepsilon_{it} = v_{it} - u_i$. 2) $v_{it} \sim iidN(0, \sigma_v^2)$. 3) $u_i \sim iidN^+(\mu, \sigma_u^2)$, nonnegative truncated-normal.	ML
Cornwell et al. (1990)	PDM	TVM	1) $u_{it} = \alpha_t - \alpha_{it}$. 2) $\alpha_{it} = \theta_{0i} + \theta_{1i}t + \theta_{2i}t^2$. 3) $\alpha_t = \max_j(\alpha_{jt})$.	GLS
Kumbhakar (1990)	PDM	TVM	1) $\varepsilon_{it} = v_{it} - u_{it}$. 2) $v_{it} \sim iidN(0, \sigma_v^2)$. 3) $u_{it} = G(t)u_i$. 4) $G(t) = \exp(-\eta(t - T))$. 5) $u_i \sim iidN^+(\mu, \sigma_u^2)$, nonnegative truncated-normal. 6) η is an unknown scalar parameter to be estimated.	ML
Battese and Coelli (1992)	PDM	TVM	1) $\varepsilon_{it} = v_{it} - u_{it}$. 2) $v_{it} \sim iidN(0, \sigma_v^2)$. 3) $u_{it} = G(t)u_i$. 4) $G(t) = \exp(-\eta(t - T))$. 5) $u_i \sim iidN^+(\mu, \sigma_u^2)$, nonnegative truncated-normal. 6) η is an unknown scalar parameter to be estimated.	ML

Table 4.1 Summary of stochastic frontier models (continued).

Authors	Model type[1]	Time effect[2]	Model specification	Estimation method[3]
Lee and Schmidt (1993)	PDM	TVM	1) $\alpha_{it} = \alpha_t - u_{it}$. 2) $\alpha_{it} = \eta_t \delta_i$.	ILS
Battese and Coelli (1995)	PDM	TVM	1) $\varepsilon_{it} = v_{it} - u_{it}$. 2) $v_{it} \sim iidN(0, \sigma_v^2)$. 3) $u_{it} \sim iidN^+(\theta z_{it}, \sigma_v^2)$, nonnegative truncated-normal. 4) $u_i t = z_{it} + W_{it}$. 5) θ is a vector of the parameters to be estimated. 6) z_{it} is a vector of the explanatory variables. 7) W_{it} is a random variable defined by the truncation of the normal distribution with zero μ and σ. 8) $TE_{it} = \exp(-u_{it}) = \exp(-z_{it}\theta - W_{it})$.	ML
Greene (2003)	CSM	-	1) $v_i \sim iidN(0, \sigma_v^2)$. 2) $u_i \sim iid$ gamma. 3) v_i and u_i distributed independently of each other, and of the regressors.	ML
Greene (2005)	PDM	TVM	1) α_0 is assumed to be normally distributed $N(0, \sigma_\alpha^2$ 2) $v_i t \sim iidN(0, \sigma_v^2)$ 3) $u_i t \sim iidN(0, \sigma_u^2)$	SML

[1] Model Type: CSM = Cross-sectional Model. PDM = Panel-data Model.
[2] Time effect: TIM = Time-invariant Model. TVM = Time-varying Model.
[3] ML = Maximum likelihood. MSL = Maximum simulated likelihood. GLS = Generalised Least Squares. ILS = Iterative Least Squares.

Once the distribution assumption of the composed error is specified, the SFA can be estimated by the maximum likelihood estimator. According to Kumbhakar et al. (2017), the log-likelihood function for each distribution is written as follows:

- Half-normal distribution:

$$\ln(\mathcal{L}) = \text{constant} - N\ln(\sigma) + \sum_{i=1}^{n} \ln\Phi(-\frac{\varepsilon_i\lambda}{\sigma}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2 \tag{4.45}$$

where $\varepsilon_i = y_i - m(x_i;\beta)$; $\sigma = (\sigma_u^2 + \sigma_v^2)^{1/2}$; $\lambda = \sigma_u/\sigma_v$; and $\Phi(.)$ is the standard normal cumulative distribution.

- Normal-exponential distribution:

$$\ln(\mathcal{L}) = const - N\ln(\sigma_u) + N\left(\frac{\sigma_v^2}{2\sigma_u^2}\right) + \sum_{i=1}^{n}\ln\Phi(-A) + \sum_{i=1}^{n}\frac{\varepsilon_i}{\sigma_u} \tag{4.46}$$

where $A = -\tilde{\mu}/\sigma_v$; and $\tilde{\mu} = -\varepsilon - (\sigma_v^2/\sigma_v)$.

- Normal-truncated-normal distribution:

$$\begin{aligned}\ln(\mathcal{L}) = & const - N\ln\sigma - N\ln\Phi\left(\frac{\mu}{\sigma_u}\right) + \sum_{i=1}^{2}\ln\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon_i\lambda}{\sigma}\right) \\ & - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2\end{aligned} \tag{4.47}$$

These log-likelihood functions estimate the stochastic cross-sectional models, and these functions can be more complicated when they are extended to estimate the stochastic panel-data models. These functions are explained in detail in other sources (e.g., Kumbhakar and Lovell (2000), Kumbhakar et al. (2015) and Sickles and Zelenyuk (2019)).

It is important to check the SFA specification via two statistical tests. First, the skewness test on OLS residuals is essential before conducting the maximum likelihood estimation for the cross-sectional models. This test is ideal to check for the validity of the model's stochastic frontier specification (Kumbhakar et al., 2015). This test was proposed initially by Schmidt and Sickles (1984). The concept of this test is that when the stochastic frontier models are used for the production function with the composed error ($\varepsilon_i = v_i - u_i$), where u_i is more than or equal to zero and v_i is distributed symmetrically around zero, the residuals obtained by OLS estimation should skew to the left (the skewness has a negative sign). On the other hand, when stochastic frontier models are used for the cost function with the composed error ($\varepsilon_i = v_i + u_i$), the residual should skew to the right (the skewness has a positive sign). The importance of the skewness test is that the positive sign means that a large number of firms have a production level that is close to the frontier (Almanidis and Sickles, 2012). When the OLS residuals skew to the right (positive sign) in the case of the production function, the model estimation concludes that all firms are super-efficient rather than inefficient (Green and Mayes, 1991).

Second, it is crucial to check the presence of the one-sided error for the stochastic frontier model. The reason for this check is that if there is no evidence supporting the presence of the one-sided error, the estimated model is reduced to a standard regression model, the OLS in this case, which will be sufficient (Kumbhakar et al., 2015, p.65). The likelihood ratio (LR) test can be used to compare the log-likelihood value of the OLS model with the log-likelihood value of the stochastic frontier model. The test statistic of the LR is written as follows:

$$-2[L(H_0) - L(H_1)] \tag{4.48}$$

where $L(H_0)$ is the log-likelihood value of the OLS model; and $L(H_1)$ is the log-likelihood value of the stochastic frontier model. The degree of freedom for this test is the number of restrictions. For example, when the stochastic frontier model is assumed to be half-normal, the number of restrictions equals one because the null hypothesis is stated as $\sigma_u^2 = 0$. On the other hand, when the stochastic frontier model is assumed to be truncated-normal, the number of restrictions increases to two because the null hypothesis is stated as $\sigma_u^2 = 0$ and $\mu = 0$. The LR test is considered a mixture of chi-square distributions (Coelli, 1995). The critical values of the mixed distributions for hypothesis testing are tabulated in Kodde and Palm (1986).

The SFA will serve as the primary parametric method for estimating efficiency in technical and cost assessments. In the SFA estimation, the stochastic frontier models that are mentioned in Table 1 will be used to determine the technical and cost efficiency of railway systems. Each of the stochastic frontier models will be tested via the LR test to assess the SFA specifications. If the SFA fails in the LR test, the technical and cost assessments will be based on the COLS method, as discussed above.

4.5 Non-parametric Approaches

4.5.1 Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a non-parametric and linear programming approach used to analyse the efficiency of a Decision Making Unit (DMU) in order to identify the most efficient DMU. Furthermore, DEA works more effectively for organisations that produce a wide range of outputs, such as products or services, based on using a variety of inputs. The concept of DEA, as stated by Sherman (1992), is 'used to compare a set of Decision Making Units (DMU's) providing a similar set of services or products and using a similar set of resources to evaluate their productivity'. This indicates the capability of the DEA to identify the most efficient and inefficient groups of DMUs. The concept of DEA was first introduced by Charnes et al. (1978)

and named as CCR-DEA model, and their model relied on constant returns to scale. An upgrade was developed by Banker et al. (1984), named as BCC-DEA model, to DEA to be capable of analysing the efficiency with variable returns to scale.

With regard to return to scale, it describes the impact of changes in inputs bundle on changes in outputs produced. When there is a constant change in outputs producible with respect to the change in input bundle, this is called Constant Return to Scale (CRS). In contrast, Variable Return to Scale (VRS) describes the variable change in outputs producible with respect to the change in input bundle, and there are two descriptions of VRS. Firstly, when the change in outputs producible is more than the change in input bundle, this is called Increasing Return to Scale (IRS). Secondly, Decreasing Return to Scale (DRS) describes that the change in outputs producible is less than the change in inputs bundle.

To estimate the technical efficiency, let us assume that there are N DMUs. Each $DMU_i (i = 1, 2, \ldots, N)$ consumes a vector of inputs, $x_i = (x_{1i}, x_{2i}, \ldots, x_{ki})$ to produce a vector of outputs, $y_i = (y_{1i}, y_{2i}, \ldots, y_{mj})$. The DMU to be evaluated is treated as DMU_0, and its input-output vector is donated as (x_0, y_0). The technical efficiency of a firm can be measured by solving the linear programming of the input-oriented BCC-DEA models as follows:

$$
\begin{aligned}
TE_I^V &= \min \theta \\
s.t. \sum_{i}^{N} \lambda_i x^i &\leqslant \theta x_0; \\
\sum_{i=1}^{N} \lambda_i y^j &\geq y_0; \\
\sum_{i=1}^{N} \lambda_i &= 1; \\
\lambda_j \geq 0(i = 1, 2, ..., N)
\end{aligned}
\tag{4.49}
$$

On the other hand, the technical efficiency of a firm can be measured by solving the linear programming of the output-oriented BCC-DEA model as follows:

$$
\begin{aligned}
TE_O^V &= \frac{1}{\min\theta} \\
s.t. \sum_{i}^{N} \lambda_i x^i &\leqslant x_0; \\
\sum_{i=1}^{N} \lambda_i y^j &\geq \theta y_0; \\
\sum_{i=1}^{N} \lambda_i &= 1; \\
\lambda_j \geq 0(i &= 1, 2, ..., N)
\end{aligned}
\tag{4.50}
$$

To estimate the cost efficiency via the DEA, let us assume that there are N DMUs. Each $DMU_i(i = 1, 2, \ldots, N)$ consumes a vector of inputs, $x_i = (x_{1i}, x_{2ji}, \ldots, x_{ki})$ priced with, $w_i = (w_{1i}, w_{2i}, \ldots, w_{li})$ to produce a vector of outputs, $y_i = (y_{1i}, y_{2i}, \ldots, y_{mj})$. The DMU to be evaluated is treated as DMU_0, and its inputs, prices and outputs are donated as (x_0, w_0, y_0). The minimum cost can be driven through solving the DEA linear programming problem as follows:

$$
\begin{aligned}
CE_0^* &= \min \sum_{i=1}^{N} w_{i0} x_{i0} \\
s.t. \sum_{i}^{N} \lambda_i x^i &\leqslant x_0; \\
\sum_{i=1}^{N} \lambda_i y^j &\geq y_0; \\
\sum_{i=1}^{N} \lambda_i &= 1; \\
\lambda_j \geq 0(i &= 1, 2, ..., N)
\end{aligned}
\tag{4.51}
$$

Having the minimum cost obtained from solving the linear problem, the cost efficiency of DMU_0 can then be measured as follows:

$$
CE_0 = \frac{C_0^*}{C_0} \tag{4.52}
$$

The main advantages of DEA are as follows, as highlighted by Ramanathan (2003):

- DEA is capable of incorporating multiple firms with multiple inputs and outputs, and it requires the quantity of inputs and outputs only to calculate technical efficiency. This advantage is quite helpful when the price information is not available or hard to obtain.

- DEA is capable of determining possible sources of efficiency as well as efficiency levels. This method can decompose economic inefficiency into technical and allocative inefficiencies.
- DEA is non-parametric and it does not require the specification of an explicit functional form relating inputs to outputs.
- One can add as many constraints to the original formulation as desired such as assigning some bounds on the input and output weights that the DMUs can take.
- DEA focuses on the computed best-practice frontiers rather than on finding patterns of central tendency.

Besides these advantages, DEA has limitations and drawbacks, which are as follows:

- DEA requires a separate linear program to be solved for each DMU. When there are many DMUs, the computation can be unwieldy. This limitation has been minimized with the development of software that specifically deals with DEA problems.
- Since DEA is a non-parametric method, statistical hypothesis tests are difficult to implement to assess the reliability of results.
- DEA scores are highly sensitive to the sample size and the characteristics of inputs and outputs. Increasing the sample size will tend to reduce the average efficiency score because including more DMUs provides greater scope for DEA to find similar comparison partners. In contrast, when the sample size is relatively small compared to the number of inputs and outputs, this situation has an impact by inflating the efficiency scores. Thus, there are several rules to specify the minimum number of DMUs in the sample.
- The DEA only determines efficiency levels relative to best practice within the selected sample. Thus, it is not meaningful to compare the scores between two different studies because differences in best practice between the samples are unknown.
- DEA produces results that are particularly sensitive to measurement error. This means that if one DMU has understated or overstated inputs or outputs, this DMU becomes an outlier and affects the shape of the production frontier that is used to determine the efficiency levels of other DMUs.

4.6 Tobit Model

This research will conduct a regression analysis to assess the impact of rail policies on technical and cost efficiency. As discussed in Section 1.3.2, the research project assumes different forms of rail policies to select the optimum railway organisation. This task will perform a regression analysis approach after the efficiency estimation, and a limited dependent variable model is more appropriate to conduct the analysis, which is widely used in the literature. The Tobit model was first proposed in the econometrics literature by Tobin (1958). The Tobit model (also known as a censored regression model) is intended to estimate linear relationships between variables when the dependent variable is limited. In this case, the estimation of the OLS might be biase because the OLS assumes a normal and homoscedastic distribution of the disturbance and the dependent variable (Amemiya, 1984; Maddala, 1983).

The standard Tobit model can be defined as follows:

$$y_i^* = \beta x_i + u_i, i = 1, 2, \ldots, n \tag{4.53}$$

$$\begin{aligned} y_i &= y_i^* if < 0 \\ \text{Otherwise}, y_i &= 0 \end{aligned} \tag{4.54}$$

where $u_i \sim N(0, \sigma^2)$; x and β are vectors of explanatory variables and unknown parameters, respectively. The y_i^* is a latent variable and y_i is the efficiency scores.

4.7 Welfare economics

This research considers welfare efficiency to assess the impact of railway reform. As the first part of the analysis is to optimise the railway reforms from the producer perspective, the customer perspective will be taken into consideration in the assessment. Moreover, the research project will seek to enhance the customer experience by identifying which railway organisation can provide better customer satisfaction, and this will be calculated through the concept of welfare economics. The welfare, or total social surplus, is defined as the benefits for society obtained from a market or industry, which is the sum of producer and consumer surplus and externalities. Producer surplus is defined as the difference between the price producers actually receive for their goods and the price at which they are willing to sell them. Consumer surplus is defined as the difference between the amount consumers would be willing to pay for a good or service and the amount they actually pay. The externalities component in the welfare calculations can be positive or negative based on their impacts. For example, congestion costs are seen as negative costs because there are losses in railway efficiency. Also, accident costs are seen as negative costs that can lead to an increase in railway costs. On the other hand, the shift from road transport to the railway system can be seen as positive because it can lead to cost savings that can be generated from road congestion and accidents.

This research will assess the impact of the change in the market structure on welfare and its components. Moreover, the research project will assume that consumers will seek to maximise their utility through obtaining products or services at the lowest prices, and that producers will be required to be more efficient and productive to meet consumer preferences. The assessment will be addressed through the welfare analysis of rail services. As discussed above, calculating the consumer and producer surplus relies on the demand and supply of products, which are rail services in the research case. This means that the research project will develop demand and supply models for rail services.

Welfare economics is used to measure the socioeconomic efficiency of the rail market. The analysis of welfare economics consists of three elements, consumer surplus, producer surplus, and externalities (Mankiw, 2016). Welfare economics (w) is calculated as follows:

$$W = CS + PS + E \tag{4.55}$$

Consumer surplus will be calculated with respect to rail demand models, as stated above. Producer surplus will be calculated as total revenue minus total cost. As the cost assessment examines different forms of railway organisations, the outcome of this assessment will be the effect of each organisation on cost efficiency by increasing or reducing the total cost. Thus, the assessment outcomes will be used to calculate the producer surplus of each organisation form.

4.8 Conclusion and Comments

The objective of this research is to estimate railway efficiency using technical, cost, and welfare assessments. Although each assessment has a specific objective, the technical and cost assessments follow a similar procedure. To estimate technical and cost efficiency, this research employs the frontier methods rather than the regular methods that assume that all firms succeed in reaching the frontier. Thus, this research has selected several methods for efficiency estimation (e.g., DEA, SFA, and COLS). On the other hand, welfare will be estimated based on the concept of welfare economics. After conducting the three assessments, this research will provide a comparison between the results of each assessment.

Chapter 5

Technical Assessment

5.1 Introduction

This chapter evaluates railway reforms by conducting a technical assessment.This assessment focuses on the physical inputs and outputs of railway systems. Furthermore, it is assumed that a railway system has a primary objective of transforming inputs into outputs. This is usually represented by a functional relationship between the input bundle and producible outputs, which is called the production technology. However, the railway system occasionally fails to maximise production technology, resulting in outputs that are not maximised at the given inputs or inputs that are not minimised at the same producible outputs. To measure this effectiveness, this research employs the concept of technical efficiency. Following the calculation of technical efficiency, the assessment attempts to examine the effect of railway reform on performance.

5.2 Literature Review

5.2.1 Competition

The impacts of railway reform on efficiency have attracted the interest of scholars, and numerous papers have been published to investigate these impacts. These papers employed a variety of estimating methods as well as empirical data. The impact caused by introducing different forms of competition on technical efficiency will be reviewed, and off-track competition is the first type of competition to be discussed. Cantos et al. (2010) found that competitive tendering for passenger rail services lowers technical efficiency, although this impact is not statistically significant. Friebel et al. (2010) indicated that the introduction of competitive tendering increases technical efficiency, but this achievement can be gained only if it is properly phased. This finding suggests that as the railway system is gradually reformed by introducing competitive tendering, there will be a positive impact on technical efficiency. In contrast, when competitive tendering is implemented alongside other types of railway reform as part of a single package (the so-called 'big bang' approach), there is a negative impact on technical efficiency. This may be explained if the railway industry is restructured in stages. This can assist governments in addressing any necessary changes in the intermediate stages of appropriate reforms, but this option is missed when several reforms are introduced as one package. Cantos et al. (2012) concluded that competitive tendering has a positive effect on technical efficiency. This finding indicates that countries that implemented competitive bidding for the passenger sector gained an increase in technical efficiency. However, when the authors evaluated the railway systems that completed all three forms (i.e., off- and on-track competition, and vertical separation), there was a significant increase in technical efficiency. In other words, there is a considerable increase in technical efficiency when the railway sector is reformed by introducing competitive tendering with the other forms as a whole package. It should be highlighted that the key difference between Friebel et al. (2010) and Cantos et al. (2012) findings is how they addressed the railway reform process in their assessments. Moreover, Friebel et al. (2010) created a model to evaluate the railway reform process by incorporating two variables: one for sequential reforms and the other for a one-package reform. In contrast, Cantos et al. (2012) built their model by including a dummy variable for railway systems that have completed all three forms (vertical separation, one-track, and off-track competition), in addition to the specific dummy variables for the three policies.

For non-econometric studies to evaluate the off-track competition, Alexandersson and Hultén (2007) assessed the competitive tendering in Sweden. They found that the franchise system in Sweden has significantly reduced public subsidies for passenger rail services. In a number of instances, the first franchise phase resulted in 20 to 30 percent savings. Additionally, there have been advancements in management, ticketing, and rolling stock, and some of them could be linked directly to franchising and the entry of new players into the market. The authors also found a reduction in the direct subsidies, which are paid to mitigate the deficits in train operations, but there is an increase in the total subsidy for the whole system paid to the infrastructure manager to maintain the rail infrastructure.

The second competition form is the on-track competition. Mulder et al. (2005) indicated that the technical efficiency of the Dutch railway system has improved as a result of market deregulation of freight services for new entrants. Cantos et al. (2010) found that on-track competition has a positive impact on technical efficiency. Moreover, there is an increase in efficiency gained by the freight market deregulation for new entrants. Friebel et al. (2010) conclude that there is an increase in technical efficiency due to free entry, but this increase is associated with the process of reforming the railway industry, as discussed above. Cantos et al. (2012) showed that the impact of on-track competition on technical efficiency is positive and statistically significant, and this impact is maximised when the authors assessed the combined railway reforms, as discussed above.

5.2.2 Organisation

Besides the introduction of competition, the organisational reform has an impact on technical efficiency. The first form of organisational reform is restructuring the railway organisation vertically. Cantos et al. (2010) showed that vertical separation significantly increases technical efficiency, but this increase is maximised when the vertical separation is combined with on-track competition. Additionally, the full potential impact of the vertical separation could be gained when combined with horizontal separation. Friebel et al. (2010) concluded that if vertical separation is appropriately phased, it has a positive effect on technical efficiency. This finding means that the process of reforming the railway industry is controlling the effect of vertical separation, as stated above. Cantos et al. (2012) found that restructuring the railway industry in a vertical separation form only does not significantly increase technical efficiency. However, this finding changed when the railway industry completed all three types of reforms (competition introduction, vertical and horizontal separation). In other words, there is a significant increase in technical efficiency caused by the vertical separation when combined with other forms.

The second form is restructuring the railway organisation horizontally. According to Cantos et al. (2010), the impact of horizontal separation varies depending on the process of reforming the railway industry. Firstly, separating passenger from freight rail services without vertical separation tends to reduce technical efficiency. Secondly, when a railway industry completes all three types of reforms (competition, vertical and horizontal separation), the impact of horizontal separation is not statistically significant. Cantos et al. (2012) concluded that the impact of horizontal separation on technical efficiency is statistically significant. Furthermore, the separation between passenger and freight rail services may only improve technical efficiency, but this improvement is maximised significantly by combining horizontal and vertical reforms.

5.2.3 Ownership and liberalisation

Along with the structural reform, the ownership reform has an influence on technical efficiency. Cantos et al. (1999) concluded that the degree of management autonomy has an impact on technical efficiency. Moreover, a higher degree of autonomy indicated more technically efficient railway performance. Cowie (1999) found that the difference in technical efficiency between public and private companies is statistically significant. The average technical efficiency of private companies is 89%, which is much greater than the 76% of public companies. This difference can be explained by the difference in management levels between public and private companies. Furthermore, private companies have more effective management than public companies, with a 13% difference in management efficiency scores between the two groups.

The impact of rail liberalisation on efficiency has been examined by the use of IBM metrics, which are discussed in Section 2.2.1.1. Bougna and Crozet (2016) employed the competition index and overall liberalisation metrics to measure the competition level of the railway market. The authors indicated that promoting a more competitive environment has a positive impact on technical efficiency. This finding has been concluded when these metrics are evaluated individually. However, when the two metrics are combined into a single model, the effect of overall liberalisation has changed to be negative, while the levels of the competition index remain positive. Lerida-Navarro et al. (2019) relied on the IBM metrics to measure the impact of competition levels on technical efficiency. The authors found that the European railway systems with high measures of the IBM metrics are more efficient, but this relationship is only statistically significant at a 90% level. However, the competition index appeared statistically significant at the 1% level. This indicates that railway systems with high scores on the competition index are more efficient.

To conclude, the IBM's measurements of rail liberalisation have been incorporated into the recent studies. The purpose of these studies is to evaluate the effect of rail liberalisation on railway performance. There is a gap, however, in evaluating the impact of railway reform by incorporating railway systems from outside Europe. In order to evaluate the impact of railway reform, this study includes railway systems that are still organised as state-owned monopolies or have not been examined previously.

5.3 Methodology

5.3.1 Non-parametric approach

This research selects Data Envelopment Analysis (DEA) as a non-parametric approach. DEA does not require a functional form to represent the relationship between the inputs and outputs, as discussed in Section 4.5.1. Recall the linear programming problem stated in Equation (4.49), there is one modification applied to the standard DEA problem by adding the constraint $\sum_{i=1}^{N} \lambda_i LL^j \leqslant \theta LL_0$. This modification aims to compare the railway system (DMU_0) with others with an equal or lesser network-line length Cantos et al. (2002, 2012). This means that the DEA analysis compares railway industries with similar network and infrastructure characteristics. Thus, the technical efficiency of a railway system can be measured by solving the linear programming of the input-oriented BCC-DEA models as follows:

$$\begin{aligned} TE_I^V &= min\theta \\ s.t. \sum_{i}^{N} \lambda_j x^j &\leqslant \theta x_0; \\ \sum_{j=i}^{N} \lambda_i y^j &\geq y_0; \\ \sum_{j=i}^{N} \lambda_i LL^j &\leqslant \theta LL_0; \\ \sum_{j=i}^{N} \lambda_i &= 1; \\ \lambda_i &\geq 0 (i = 1, 2, ..., N) \end{aligned} \tag{5.1}$$

It should be noted that the above equation represents the DEA estimation under VRS. To estimate the technical efficiency under CRS, the constrain ($\sum_{i=1}^{N} \lambda_i = 1$) is removed from the DEA problem. Moreover, if the constrain is removed from the DEA problem, the efficiency then will be estimated under CRS. On the other hand, if the constrain is kept in the DEA problem, the efficiency then will be estimated under VRS.

5.3.2 Parametric approach

This research project considers Corrected Ordinary Least Square (COLS) and Stochastic Frontier Analysis (SFA) as parametric approaches. The primary objective is to estimate the technical efficiency via the SFA approach. As these approaches require specifying the functional form, this research selects the distance function to represent the nature of the railway industry as a multi-output industry. In the Research Methodology chapter, it has been discussed that the production function cannot accommodate multi-outputs. As a solution, the distance function is commonly applied in the literature when firms have multiple outputs.

The distance function can be written in input- or output-oriented forms. Each form can be selected based on which variables the firms have more control over. Coelli and Perelman (2000) highlighted that selecting the orientation form of the distance function is not particularly significant for the railway industry. Thus, this research also chooses the input-distance function as a functional form. Recall the input distance function from Section 4.3.2, which is written as follows:

$$-\ln(x_{Ki}) = TL(x_{ki}/x_{Ki}, y_{mi}, \alpha, \beta, \delta) - \ln(D_{Ii}) \tag{5.2}$$

Equation (5.2) will be slightly modified by introducing a term of the technological progress (T). Thus, the input-distance function can be rewritten as follows

$$\begin{aligned}-\ln(x_{Kit}) = \alpha_0 &+ \sum_{m=1}^{M}\alpha_m \ln(y_{mit}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M}\alpha_{mn}\ln(y_{mit})\ln(y_{nit}) \\ &+ \sum_{k=1}^{K-1}\beta_k \ln(x^*_{kit}) + \frac{1}{2}\sum_{k=1}^{K-1}\sum_{l=1}^{K-1}\beta_{kl}\ln(x^*_{kit})\ln(x^*_{lit}) + \theta_t T + \frac{1}{2}\theta_{tt}T^2 \\ &+ \sum_{m=1}^{M}\sum_{k=1}^{K-1}\delta_{mk}\ln(y_{mit})\ln(x^*_{kit}) + \sum_{m=1}^{M}\varphi_{mt}\ln(y_{mit})T \\ &+ \sum_{k=1}^{K-1}\phi_{kt}\ln(x^*_{kit})T - \ln(D_{Iit})\end{aligned} \tag{5.3}$$

To estimate the input-distance function via COLS, the procedure is similar to what stated in Section 4.4.1 with slight changes. Coelli and Perelman (1996) highlighted that the intercept of the OLS is adjusted and corrected. The adjustment for the input-distance function will be replacing the unobservable term "$-\ln(D_{Ii})$" in Equation (5.2) by a random error term (ε_i). Then, the intercept will be corrected by adding the largest positive OLS residual. Lastly, the technical efficiency for each observation is calculated as follows:

$$TE = \exp(\max \hat{\varepsilon}^+ - \hat{\varepsilon}_i) \tag{5.4}$$

where $\hat{\varepsilon}^+$ is the maximum positive OLS residual.

Besides the COLS estimation, the SFA can be applied to estimate the input distance function. Recall Equation (5.2), the unobservable term "$-\ln(D_{Ii})$" will be replaced with statistical noise (v) and inefficiency (u). The inefficiency term u will be restricted as non-positive in the case of the input distance function (i.e., $u_i = -u_i$) (Kumbhakar et al., 2007). Lan and Lin (2006) explain this restriction as follows due to excess inputs. Thus, Equation (5.2) can be rewritten as follows

$$-\ln(x_{Ki}) = TL(x_{ki}/x_{Ki}, y_{mi}; \alpha, \beta, \delta) + v_i - u_i \tag{5.5}$$

This research has attempted to estimate the input-distance function via the SFA method by using the Stata written commands by Kumbhakar et al. (2015), but there are some practical issues with the SFA estimation, which lead to this method not being applicable for the research dataset and not fitting the purpose of this research. First, the Schmidt and Sickles (1984) and Pitt and Lee (1981) models estimate the efficiency as time-invariant. The invariant efficiency over time does not capture the change in efficiency due to the railway reform. This invariant measure does not support the effect estimation of railway reform on efficiency. Thus, these two models are rejected.

Second, several stochastic frontier models have been rejected because of the LR test. The Aigner et al. (1977), Kumbhakar (1990) and Battese and Coelli (1992) models are rejected because they fail to reject the null hypothesis of the LR test. In other words, there is no difference between the estimation of the half-normal model and the OLS. On the other hand, for stochastic panel-data models, the proposed parameters to measure time-varying efficiency are not statistically different from zero. Thus, this model collapses to the Pitt and Lee (1981) model (Kumbhakar et al., 2015).

Besides the rejected models, this research project has attempted to estimate more stochastic frontier models. However, these models do not achieve convergence to obtain the optimal solution for maximising the log likelihood function, such as the true fixed- and random-effects models proposed by Greene (2005). To conclude, the stochastic frontier approach does not fit the data sample of this research project. The possible explanation for this reason is that the data has the wrong skewness based on the OLS estimation, as explained in Section 4.4.2. Therefore, this project will rely on COLS as a parametric approach.

5.4 Data Sample

This research estimates the multiple-outputs and -inputs technology of the railway systems. The outputs of the railway system can be defined at two levels. The intermediate output level is related to train services, and this level can be clustered into three measures as follows:

- Train output measured as train-km, which is the total distance travelled by each train.
- Passenger-train output measured as passenger-train-km, which is the total distance travelled by each passenger-train.
- Freight-train output measured as freight-train-km, which is the total distance travelled by each freight-train.

Based on the passenger-train and freight-train, the train-km can be calculated as the sum of these two measures. The final output level is related to passenger and freight rail services, and this level can be grouped into two levels as follows:

- Passenger output measured as pass-km, which is the total distance transported by each passenger.

- Freight output measured as tonne-km, which is the total distance transported by each tonne of freight.

Passenger and freight outputs are commonly used in the literature as the final output measures of rail services. Although these measures are affected by policies and fare rates, Oum and Yu (1994) highlighted that these measures fit the econometric studies when they aim to examine government policies. Because the purpose of this research project is to assess different rail policies, the final output measures are selected.

The primary inputs used by a railway system to produce rail services are labour, capital and energy (Coelli and Perelman, 1996). The labour input is the total number of staff that are involved in the production of rail services. This input could be defined at different levels such as part-time and full-time employees, but this requires more detailed data which is often not available. Coelli and Perelman (1996) define the capital input like the length of network-lines, length of tracks and the number of rolling stock (locomotives, coaches, wagons and railcars). The length of track can be defined as single-, double-track or network-line length. In addition, the percentage of electrified lines can be used to measure the technological progress of the railway system. The energy variable is defined as the energy consumed by the rolling stock (locomotives, railcars and multiple-unit trailers). Diesel and electricity are the main two sources of energy used by the rolling stock. Including all these variables in the estimation could be problematic. For example, energy consumption and the number of rolling stocks can be closely correlated. More rolling stock used by a railway system means more energy consumption.

This research is using annual data on 29 railway systems from Europe, North Africa and Asia from 2000 to 2017. This sample has been selected based on the criteria that are mentioned in Section 3.3. The data is collected from the International Union of Railways (UIC) and completed from the annual reports of each railway company and statistical reports of each country. The dataset consists of two outputs and three inputs. The two outputs are passenger output (y_1) and freight output (y_2) measured as pass-km and tonne-km, respectively. The selection of these measures is discussed above. The inputs are the total number of rolling stock (x_1), the network-line length in km (x_2) and number of employees (x_3). Due to issues with data availability, there is not sufficient data for the energy consumption measure.

The average values of the outputs and inputs and the number of observations for each railway system are provided in Table 5.1. There is a large variation in the input and output ratio. For example, the railway systems in Japan, Germany, and France are much bigger in terms of inputs and outputs compared to the smallest railway systems in Ireland, Slovenia, and Saudi Arabia. Some railway systems exhibit a balance between passenger and freight railway services (i.e, Algeria, Belgium, Germany, and Morocco), where the variation is less than 10%. On the other hand, several railway

TABLE 5.1: The average values of the outputs and inputs at the country level.

Country	No obser.	y_1^a	y_2^a	x_1	x_2	x_3
Algeria	18	1,065	1,451	10,969	3,814	11,556
Austria	14	9,470	16,526	23,689	5,310	44,764
Azerbaijan	16	696	7,539	20,225	2,097	14,087
Belgium	10	9,055	7,520	15,543	3,500	39,147
Bulgaria	18	2,218	3,467	12,196	4,146	25,560
Czech Rep.	18	6,888	13,357	36,746	9,468	56,790
Finland	18	3,748	9,797	11,925	5,888	9,246
France	14	79,306	38,024	48,095	29,530	163,812
Germany	18	75,854	81,578	127,963	34,273	251,200
Greece	7	1761	610	4,217	2,490	8,064
Iran	16	13,367	20,974	23,092	7,603	13,494
Ireland	11	1,670	274	1,772	1,830	4,807
Italy	15	43,821	17,192	52,837	16,377	91,537
Japan	16	252,594	21,681	23,595	20,163	131,257
Kazakhstan	15	15,490	196,057	91,195	14,489	107,229
Latvia	12	623	15,086	6,215	2,174	13,794
Morocco	18	3,837	5,145	6,468	2,017	8,547
Pakistan	18	21,714	4,030	21,122	7,791	71,807
Poland	15	16,644	40,818	84,982	19,814	123,810
Portugal	9	3,575	2,317	4,163	2,837	9,633
Saudi Arabia	18	340	1,505	2,324	1,248	1,735
Slovakia	18	2,556	8,473	19,335	3,635	18,336
Slovenia	18	755	3,352	4,366	1,219	8,217
South Korea	18	27,358	10,104	15,297	3,499	29,321
Spain	18	21,412	9,159	17,277	13,702	29,106
Switzerland	18	15,321	8,996	16,562	3,110	28,780
Turkey	17	5,133	9,617	19,196	9,189	30,681
Turkmenistan	5	1,525	10,082	14,160	2,874	17,713
Uzbekistan	8	2,371	19,378	22,643	4,073	50,029
Total	434					

[a] The variables are expressed in millions.

systems have more passenger rail services compared to freight rail services, where the passenger share of rail services is 92%, 86% and 84% in Japan, Ireland, and Pakistan, respectively. In contrast, several railway systems display more freight rail services compared to passenger rail services, where the freight share of rail services is 96%, 93% and 92% in Latvia, Kazakhstan, and Azerbaijan, respectively.

5.5 Results

5.5.1 Input Distance Function Estimation

For the parametric approach, the input distance function has been estimated by the COLS method. Prior to estimation, all variables have been normalised by their sample mean. That is, each input and output variable has been divided by its mean. This approach has been implemented by Coelli and Perelman (2000), Kumbhakar et al. (2007), Last and Wetzel (2010) and Bougna and Crozet (2016). In this case, the coefficients of the first-order parameters can be interpreted as distance elasticities estimated at the sample means. It should be noted that the number of employees (x_3) has been selected for linear homogeneity, and this feature is to identify the distance function in input-orientation form, as discussed in Section 4.3.2. Thus, the input distance function that is estimated is written as follows:

$$
\begin{aligned}
-\ln(x_{3it}) = {} & \alpha_0 + \alpha_1 \ln(y_{1it}) + \alpha_2 \ln(y_{2it}) + \frac{1}{2}\alpha_{11} \ln(y_{1it})^2 + \alpha_{12} \ln(y_{1it}) \ln(y_{2it}) \\
& + \frac{1}{2}\alpha_{11} \ln(y_{2it})^2 + \beta_1 \ln\left(\frac{x_{1it}}{x_{3it}}\right) + \beta_2 \ln\left(\frac{x_{2it}}{x_{3it}}\right) + \frac{1}{2}\beta_{11} \ln\left(\frac{x_{1it}}{x_{3it}}\right)^2 \\
& + \beta_{12} \ln\left(\frac{x_{1it}}{x_{3it}}\right) \ln\left(\frac{x_{2it}}{x_{3it}}\right) + \frac{1}{2}\beta_{22} \ln\left(\frac{x_{2it}}{x_{3it}}\right)^2 + \theta_t T + \frac{1}{2}\theta_{tt} T^2 \\
& + \delta_{11} \ln(y_{1it}) \ln\left(\frac{x_{1it}}{x_{3it}}\right) + \delta_{12} \ln(y_{1it}) \ln\left(\frac{x_{2it}}{x_{3it}}\right) \\
& + \delta_{21} \ln(y_{2it}) \ln\left(\frac{x_{1it}}{x_{3it}}\right) + \delta_{22} \ln(y_{2it}) \ln\left(\frac{x_{2it}}{x_{3it}}\right) + \varphi_{1t} \ln(y_{1it}) T \\
& + \varphi_{2t} \ln(y_{2it}) T + \phi_{1t} \ln\left(\frac{x_{1it}}{x_{3it}}\right) T + \phi_{2t} \ln\left(\frac{x_{2it}}{x_{3it}}\right) T + \varepsilon_{it}
\end{aligned}
\tag{5.6}
$$

where y_1 is passenger output measured as pass-km; y_2 is freight output measured as tonne-km; x_1 is the total number of rolling stocks; x_2 is the network-line length in km; x_3 is the number of employees; t is time trend; α, β, θ,δ,φ and ϕ are model parameters to be estimated; and $i = 1, 2, \ldots, I$ is the number of railway systems. Table 5.2 reports the OLS estimation of the input-distance function by using the Stata software. To test the goodness-of-fit, the estimated model has a high value of adjusted R^2, which is around 91%. This means that the estimated model can be considered a reasonable fit for the observed data. The coefficients of the first-order have the expected signs and are statistically significant at a 99.9% confidence level. That is, the estimated input-distance function satisfies the monotonicity condition at the sample means (i.e., non-decreasing in x and non-increasing in y). For the concavity condition, the second-order input parameters, the coefficient of the cross-product interaction (β_{12}, β_{13} and β_{23}) have negative signs as expected. This proves that the concavity condition of the input distance function is satisfied at the sample means.

Table 5.2: The input distance function estimated by OLS

Variable	Parameter	OLS	Variable	Parameter	OLS
$\ln(y_1)$	α_1	-0.358***	T	θ_t	0.014
		(0.041)			(0.015)
$\ln(y_2)$	α_2	-0.417***	T^2	θ_{tt}	0.000
		(0.053)			(0.002)
$\ln(y_1)^2$	α_{11}	0.043*	$\ln(y_1)\ln(x_1)$	δ_{11}	0.357***
		(0.017)			(0.064)
$\ln(y_1)\ln(y_2)$	α_{12}	-0.031	$\ln(y_1)\ln(x_2)$	δ_{12}	-0.247***
		(0.022)			(0.055)
$\ln(y_2)^2$	α_{22}	-0.118**	$\ln(y_1)\ln(x_3)$	δ_{13}	-0.110
		(0.009)			
$\ln(x_1)$	β_1	0.294**	$\ln(y_2)\ln(x_1)$	δ_{21}	-0.134
		(0.115)			(0.085)
$\ln(x_2)$	β_2	0.308***	$\ln(y_2)\ln(x_2)$	δ_{22}	-0.016
		(0.105)			(0.088)
$\ln(x_3)$	β_3	0.398	$\ln(y_2)\ln(x_3)$	δ_{23}	0.150
$\ln(x_1)^2$	β_{11}	1.598***	$\ln(y_1)t$	φ_{1t}	0.010**
		(0.255)			(0.003)
$\ln(x_1)\ln(x_2)$	β_{12}	-0.794***	$\ln(y_2)t$	φ_{2t}	-0.007
		(0.197)			(0.004)
$\ln(x_2)^2$	β_{22}	0.957***	$\ln(x_1)t$	ϕ_{1t}	0.030*
		(0.219)			(0.012)
$\ln(x_3)^2$	β_{33}	0.967	$\ln(x_2)t$	ϕ_{2t}	-0.037***
					(0.01)
$\ln(x_1)\ln(x_3)$	β_{13}	-0.804	$\ln(x_3)t$	ϕ_{3t}	0.007
$\ln(x_2)\ln(x_3)$	β_{23}	-0.163	Constant	α_0	-0.420***
					(0.072)
N		434			
adj. R^2		0.91			
F		209.397			
LL		-167.567			

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.
Standard errors in parentheses.
Underlined coefficients are obtained by homogeneity restrictions.

According to Färe and Primont (1995), the scale elasticity of the input distance function is given by

$$EE = -1 \Bigg/ \left(\sum_{m=1}^{M} \frac{\partial \ln D}{\partial \ln y_m} \right) \tag{5.7}$$

As the variables are normalised by the means, the scale elasticity is calculated as the sum of the first-order output elasticities in absolute values. The sum of the coefficients of the first-order outputs in absolute value is 0.775, which is less than one. This finding complies with other empirical works that find increasing returns to scale in the railway systems (e.g., Coelli and Perelman (1996, 1999, 2000) and Kumbhakar et al. (2007)). Following Coelli and Perelman (1999), an approximate scale elasticity suggests 1.290 (1/0.775). This indicates that the sample size exhibits increasing returns to scale at the means, which means that if the inputs increase by 1%, the outputs will increase by 1.290.

The estimated input elasticities for the total number of rolling stocks (β_1) and for network-line length (β_2) are found to be equal to 0.294 and 0.308, respectively. The number of employees has an input elasticity of 0.398, which is calculated via the homogeneity restrictions. These elasticities can be interpreted as shadow shares. In this regard, the total number of rolling stock accounts for 29.4%, the network-line length accounts for 30.8%, and the number of employees accounts for 39.8% of passenger and freight rail services at the sample mean. The first-order time-trend parameter (θ_t) is 0.014 but not statistically significant. The positive sign indicates progressive technical change at the sample means.

5.5.2 Efficiency Levels

The measurements of the technical efficiency are obtained by COLS as a parametric approach and DEA as a non-parametric approach. Figure 5.1 visualises the average technical efficiency levels over the sample period measured by each estimation method. It appears that DEA under VRS has higher efficiency levels (0.6650 as an average value) compared to the two methods. The average values of technical efficiency estimated by DEA under VRS and CRS have the same pattern of change over the sample period. The COLS estimation has lower average efficiency scores over the sample period compared to the DEA estimation. This variation is expected because the DEA estimation identifies the best practice of DMUs of transforming inputs into outputs, and these DMUs score 1 as technical efficiency. As a result, DEA estimates higher average efficiency scores compared to other methods.

There is no clear evidence of the sudden jump in the efficiency scores in 2010 and 2011, around a 15% jump. However, the potential reason can be related to the procedure of

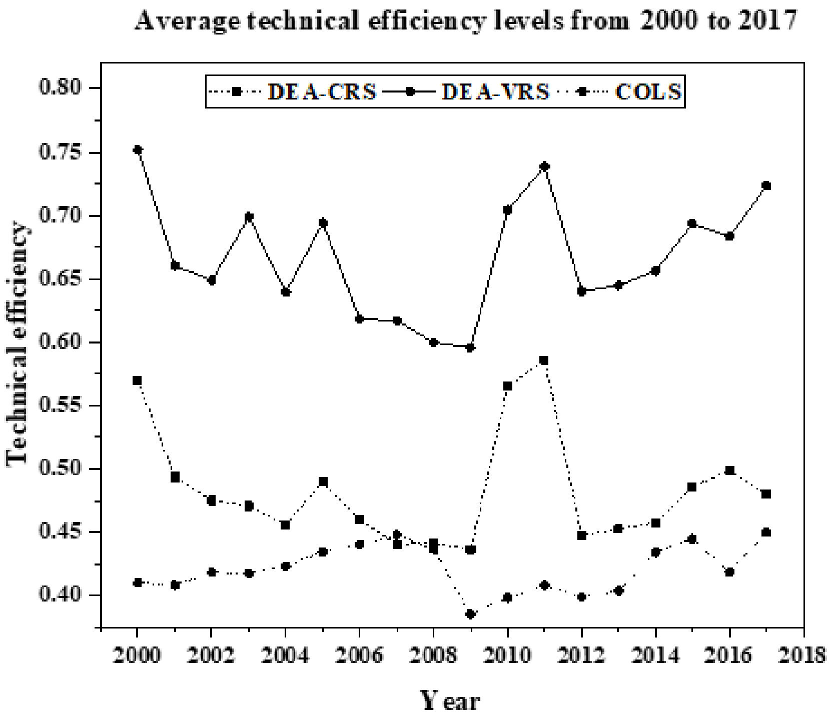


FIGURE 5.1: Average technical efficiency levels over the sample period.

TABLE 5.3: Spearman rank-order correlation coefficient.

	COLS	DEA^{VRS}	DEA^{CRS}
COLS	1		
TE^{VRS}	0.6863*	1	
TE^{CRS}	0.5994*	0.7361*	1

* indicates significance at the 1% level.

the DEA estimation as it is highly sensitive to the sample. As Japan and Kazakhstan are identified as the frontier for this sample, the available observations for Japan are only from 2000 to 2009. This means that the data for Japan is missing from 2010 onwards. On the other hand, the data for Kazakhstan is missing in 2010 and 2011. This indicates that new railway systems are identified as the frontier to calculate technical efficiency.

In order to assess the similarity in ranking railway systems among the estimation methods, the Spearman rank-order correlation coefficient between the efficiency scores is reported in Table 5.3. If the coefficient has a value close to one, it means that the efficiency rankings of the railway systems estimated by the three approaches are similar. The results indicate a positive and strong correlation between the efficiency scores estimated by the parametric and non-parametric approaches. In other words, the relative positions of the railway industries regarding efficiency scores are consistent in general.

TABLE 5.4: Average technical efficiency levels estimated by COLS and DEA.

Country	**DEA^{CRS}**	**DEA^{VRS}**	**COLS**	**Average**
Algeria	0.1113	0.3339	0.249	0.2314
Austria	0.4349	0.5256	0.4127	0.4577
Azerbaijan	0.3482	0.7457	0.3446	0.4795
Belgium	0.3871	0.5666	0.4431	0.4656
Bulgaria	0.1692	0.3738	0.2807	0.2746
Czech Rep.	0.2217	0.2736	0.2641	0.2531
Finland	0.6798	0.7554	0.4027	0.6126
France	0.5227	0.6681	0.3896	0.5268
Germany	0.4235	0.5535	0.3004	0.4258
Greece	0.1456	0.5401	0.3934	0.3597
Iran	0.9597	0.9632	0.4533	0.7921
Ireland	0.2066	0.9973	0.4371	0.5470
Italy	0.3837	0.4223	0.3089	0.3716
Japan	1	1	0.5022	0.8341
Kazakhstan	1	1	0.5719	0.8573
Latvia	0.9159	0.9930	0.3928	0.7672
Morocco	0.5525	0.8235	0.8705	0.7488
Pakistan	0.2875	0.3806	0.2653	0.3111
Poland	0.291	0.4079	0.2484	0.3158
Portugal	0.3459	0.6451	0.4683	0.4864
Saudi Arabia	0.4959	1	0.4379	0.6446
Slovakia	0.3477	0.5078	0.3780	0.4112
Slovenia	0.3801	0.9888	0.6934	0.6874
South Korea	0.8338	0.9336	0.6669	0.8114
Spain	0.5427	0.5678	0.3899	0.5001
Switzerland	0.6176	0.8061	0.5121	0.6453
Turkey	0.2827	0.3205	0.2974	0.3002
Turkmenistan	0.3755	0.6042	0.4922	0.4906
Uzbekistan	0.5022	0.6402	0.2963	0.4796
Average	0.4824	0.6650	0.4207	0.5203

The measurements of the technical efficiency for each railway system are tabulated in Table 5.4. These measurements are varied between the estimation methods. To assess the variation in the technical efficiency between railway systems, the Kruskal-Wallis test was conducted. This test has been performed on the three approaches separately. The results indicate rejection of the null hypothesis at the 1% level. The measurements of the technical efficiency between railway systems are not at the same level. For the COLS estimation, the top five railway systems with higher average efficiency levels are Morocco, Slovenia, South Korea, Kazakhstan, and Switzerland. In contrast, Poland, Algeria, the Czech Republic, Pakistan, and Bulgaria have lower average efficiency levels.

For the DEA estimation, the railway systems of Japan and Kazakhstan have the highest average efficiency scores under VRS and CRS. This means that the average score of the two countries is situated on the frontier for both cases. The technical efficiency levels for the remaining railway systems differ between VRS and CRS. For the VRS case, Saudi Arabia has a value of 1 for technical efficiency, which means that the average score is on the frontier. South Korea, Iran, Slovenia, Latvia, and Ireland obtained a value of above 0.90, which means that their average efficiency score is no more than 10 percentage points from the production frontier. Conversely, five railway systems (Czech Republic, Turkey, Algeria, Bulgaria, and Pakistan) scored less than 0.4 for average efficiency levels.

For the CRS assumption, no railway system has scored a value of 1, except Japan and Kazakhstan. This indicates that 26 railway systems are not located on the frontier. Iran and Latvia obtained a score above 0.90, which means their score is less than 10 percentage points from the frontier. On the other hand, Algeria, Greece, Bulgaria, Ireland, and the Czech Republic have the lowest average efficiency levels.

5.6 Effect Estimation

In this section, this research aims to develop a model that is capable of explaining rail inefficiency in terms of railway reform. In the previous section, the technical efficiency of railway systems has been estimated by COLS and DEA. The estimated efficiency scores are bounded between (0, 1], which indicates that these scores are limited from the right. This can be explained by the nature of the efficiency scores, where the railway system obtains a value equal to one if it is located on the frontier. For this reason, this research performs the Tobit model for effect estimation. However, this research is more interested in the determinants of inefficiency levels. Thus, the estimated efficiency scores are transformed into inefficiency scores as follows:

$$IEM = \left(\frac{1}{TE}\right) - 1 \tag{5.8}$$

where IEM is the inefficiency score and TE is the technical efficiency score. It should be noted that the inefficiency score now becomes censored from the left, [0, 1).

To estimate the impact of railway reforms, two sets of variables are considered. The first group is the test variables, which represent rail policies. The test variables are defined at two levels (i.e., individual and combined levels). The first level represents rail policies in terms of competition form for rail services, horizontal and vertical dimensions, which are as follows:

- Organisational separation (D_{OS}), which represents that IMs and RUs are organised with a separation of subsidies and decision-making producers but within a holding company.

- Institutional separation (D_{IS}), which represents that IMs and RUs are managed and operated by separated entities.
- Regional separation (D_{RS}), which represents dividing the rail services on a regional basis.
- Market function separation (D_{MES}), which represents that passenger and freight rail services are separated and provided by different railway companies.
- Off-track competition (D_{OFTC}), which represents competitive tendering for passenger rail services.
- On-track competition (D_{ONTC}), which represents a free entry for freight rail services.

Practically, introducing a new form of rail policies has been combined with other forms. Thus, this research project proposes a new measure, which can capture the levels of railway reforms. These levels are defined as follows:

- Level-1 (D_{L1}): Any form of vertical separation only.
- Level-2 (D_{L2}): Any form of horizontal separation only.
- Level-3 (D_{L3}): Any form of competition only.
- Level-4 (D_{L4}): Any form of vertical separation combined with any form of horizontal separation.
- Level-5 (D_{L5}): Any form of vertical separation combined with any form of competition.
- Level-6 (D_{L6}): Any form of vertical separation combined with any form of horizontal separation and any form of competition.

The second group is control variables, which could separate and determine the impact of railway reforms more accurately. The first control variable is GDP per capita (z_1). The second control variable is population density (z_2), which is the total population of each country divided by its area. The third control variable is line density (z_3), which is calculated as the total network-line length divided by country area divided by 1000.

The last control variable is the traffic density of the railway system. This variable has been included in the estimation by the literature (e.g., Mizutani and Uranishi (2013), Mizutani et al. (2015), Bougna and Crozet (2016) and Smith et al. (2018)). These authors defined the traffic density with respect to train-km density. Due to the lack of data, this research project calculates the traffic density as Unit Traffic Density (UTD) and denoted as (z_4), which was proposed by Thompson and Bente (2014). UTD is calculated as the sum of pass-km and tonne-km and divided by the network-line length. This research also evaluates the impact of organisational and institutional separation in conjunction with traffic density, following Mizutani and Uranishi (2013) and Mizutani et al. (2015).

It has been discussed that the level of regulation has an impact on efficiency. This measure has been used by Bougna and Crozet (2016), Lerida-Navarro et al. (2019) and Smith et al. (2018). The authors used the IBM measure, which focused on European countries. This research project involves countries that are not covered by this measure, and there is no possible measure that can cover all railway systems. Thus, this measure is dropped from the estimation. Based on these variables, two Tobit models are specified as follows:

$$\begin{aligned} \ln(IEM) = {} & \alpha_0 + \alpha_1 \ln(z_1) + \alpha_2 \ln(z_2) + \alpha_3 \ln(z_3) + \alpha_4 \ln(z_4) + \beta_1 \ln(z_4) D_{OS} \\ & + \beta_2 \ln(z_4) D_{IS} + \beta_3 D_{OS} + \beta_4 D_{RS} + \beta_5 D_{OS} + \beta_6 D_{MFS} \\ & + \beta_7 D_{ONTC} + \beta_8 D_{OFTC} + \varepsilon \end{aligned} \tag{5.9}$$

$$\begin{aligned} \ln(IEM) = {} & \alpha_0 + \alpha_1 \ln(z_1) + \alpha_2 \ln(z_2) + \alpha_3 \ln(z_3) + \alpha_4 \ln(z_4) + \delta_1 D_{L1} + \delta_2 D_{L2} \\ & + \delta_3 D_{L3} + \delta_4 D_{L4} + \delta_5 D_{L5} + \delta_6 D_{L6} + \varepsilon \end{aligned} \tag{5.10}$$

The results of the effect estimation obtained by the Tobit model are presented in Table 5.5. It should be noted that the technical inefficiency scores that are obtained by COLS are used in this estimation. The reason for this selection is that the DEA approach identifies the best practice in transforming inputs into outputs among the railway systems. This best practice will then be assigned as a frontier and will have a score of one as technical efficiency. The efficiency score for the remaining railway systems will be calculated as the distance from the frontier. This means that best practice will have a score of one as technical efficiency for the whole sample period. For this research, Japan and Kazakhstan are identified as best practices in the DEA approach and have a score of one for the sample period. Having said that, a constant efficiency score does not account for variations in efficiency levels due to changes in railway organisation. Therefore, the COLS estimation is used to determine the impact of railway reform on inefficiency levels.

For the control variables, the population density (z_2) variable has been dropped from the two models because it is not statistically significant, but the remaining control variables are. The coefficient of GDP per capita (α_1) is statistically significant at a 99.9% confidence level and has a negative sign. In other words, railway systems with higher GDP have lower inefficiency levels. This is expected because countries with high-income levels could lead to more utilisation of their railway systems.

The coefficient of line density (α_3) is statistically significant at a 99.9% confidence level and has a positive sign. Precisely, railway systems with high levels of line density have higher inefficiency levels. This finding can be explained by the complexity level of the railway network. Because this variable is derived by network-line density divided by country area, higher line density indicates a more complicated network (e.g., junctions and conflicts), which can contribute to congestion consequences.

The coefficient of traffic density (α_4) is statistically significant at a 99.9% confidence level and has a negative sign. This finding is similar to Bougna and Crozet (2016), Cantos et al. (2010) and Cowie and Riddington (1996). This finding suggests that railway systems with higher traffic density are likely to have lower levels of inefficiency. In other words, higher utilisation of railway networks could maximise the use of inputs. However, congestion effects might be caused by more traffic density. This effect can be captured more precisely if the traffic density is measured as train-km divided by network length rather than unit traffic unit density.

As well as the impact of the control variables, the results of the test variables are reported in Table 6.7. For the vertical dimension, the coefficients of institutional separation (β_3) and organisational separation (β_4) are statistically significant at a 99.9% confidence level and have a negative sign. This result is in line with Cantos et al. (2010)[1], Cantos et al. (2012) and Friebel et al. (2010)[2]. In other words, railway systems with organisational and institutional separation tend to reduce their inefficiency levels compared to vertical integration. When the impact of vertical separation is examined with the level of traffic density, both vertical separation forms appears to increase the technical inefficiency levels. The coefficients of institutional separation (β_2) and organisational separation (β_1) in conjunction with traffic density are statistically significant at a 99.9% confidence level and have a positive sign. These results indicate that railway systems that are vertically separated tend to increase the levels of technical inefficiency. These results are in line with Mizutani and Uranishi (2013)[3].

For the horizontal dimension, the coefficient of regional separation (β_5) is statistically significant at a 99.9% confidence level and has a positive sign. This finding indicates that the regional division of rail services tends to increase the inefficiency levels compared to the horizontal integration. The coefficient of market function separation (β_6) is not statistically significant. Although the coefficient has a negative sign, it is not statistically different from zero. This result is in opposition to Cantos et al.

[1] It should be noted that Cantos et al. (2010) did not distinguish between the organisational and institutional separation. They added both forms as one form of vertical separation.

[2] The finding of Friebel et al. (2010) is conditioned in the case of the vertical separation if it is appropriately phased, as discussed above.

[3] The aim of Mizutani and Uranishi (2013) is to assess the impact of vertical separation on cost reduction.

Table 5.5: Determinants of technical inefficiency levels.

Term	Parameter	COLS	
		Model (1)	Model (2)
$\ln(z_1)$	α_1	-0.242***	-0.198***
		(0.018)	(0.020)
$\ln(z_3)$	α_3	0.164***	0.122***
		(0.018)	(0.019)
$\ln(z_4)$	α_4	-0.494***	-0.389***
		(0.087)	(0.028)
$\ln(z_1)D_{OS}$	β_1	0.349***	
		(0.082)	
$\ln(z_1)D_{IS}$	β_2	0.348***	
		(0.078)	
D_{IS}	β_3	-0.383***	
		(0.094)	
D_{OS}	β_4	-0.537***	
		(0.136)	
D_{RS}	β_5	0.365***	
		(0.069)	
D_{MFS}	β_6	-0.085	
		(0.073)	
D_{ONTC}	β_7	0.207***	
		(0.049)	
D_{OFTC}	β_8	0.139*	
		(0.062)	
D_{L1}	δ_1		0.024
			(0.065)
D_{L2}	δ_2		0.280*
			(0.117)
D_{L3}	δ_3		0.138
			(0.117)
D_{L4}	δ_4		0.457**
			(0.163)
D_{L5}	δ_5		0.168**
			(0.058)
D_{L6}	δ_6		0.323***
			(0.071)
cons	α_0	2.749***	2.343***
		(0.157)	(0.176)
N		434	434
Pseud R^2		0.520	0.410
LL		-168.030	-206.472

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.
Standard errors in parentheses.

(2010)[4]. In other words, separating passenger rail services from freight rail services reduces the inefficiency levels. This is expected because freight rail services consume lower sources (Thompson and Bente, 2014).

For the competition forms, the coefficients of on-track competition (β_7) and off-track competition (β_8) are statistically significant at a 99.9% confidence level and have a positive sign. For competitive tendering, this result is in opposition to Friebel et al. (2010), Cantos et al. (2012) and Bougna and Crozet (2016) but is in line with Cantos et al. (2010)[5]. For the freight market deregulation, the results are in opposition to Cantos et al. (2010), Friebel et al. (2010) and Cantos et al. (2012). The difference between these research results and other findings is the sample size. Previous studies have mainly targeted the European railway systems, while this research includes systems from different regions. Thus, for this sample size, competitive tendering for passenger services and free entry for freight services increase the inefficiency levels. It has been discussed that railway systems have been restructured vertically and horizontally to introduce competition. However, the results show that although those railway systems with vertical separation scored lower inefficiency levels, competitive tendering and freight market deregulation do not appear to have favoured more efficient behaviour. For the impact of competitive tendering, Alexandersson and Hultén (2005) highlighted the factors that could be caused by the franchise system on railway performance. First, a low number of train operating companies submit bids for passenger rail services. In fact, some contracts are directly awarded to TOCs without competition (e.g., German railway system). Second, some TOCs underestimate the future developments, or the bid winner could be inefficient (Preston, 1996), and it is difficult to make adjustments.

Besides the assessment of each rail policy individually, this research estimates the impact of combining rail policies. According to these results, combining rail policies appears to increase the technical inefficiency levels. The coefficient of level-04 (δ_4) is statistically significant at a 99% confidence level and has a positive sign. In other words, restructuring the railway organisation with any form of vertical separation combined with any form of horizontal separation increases the inefficiency. The coefficient of level-05 (δ_5) is statistically significant at a 99% confidence level and has a positive sign. This finding means that reforming the railway industry with any form of vertical separation combined with any form of competition increases the inefficiency levels. The coefficient of level-06 (δ_6) is statistically significant at a 99.9% confidence level and has a positive sign. In other words, combining the three forms of rail policies increases the inefficiency levels. This result is in opposition to Cantos et al. (2012), where the combined rail policies increase railway efficiency.

[4]The result of Cantos et al. (2010) is not statistically significant

[5]It should be noted that the result of Cantos et al. (2010) is not statistically significant.

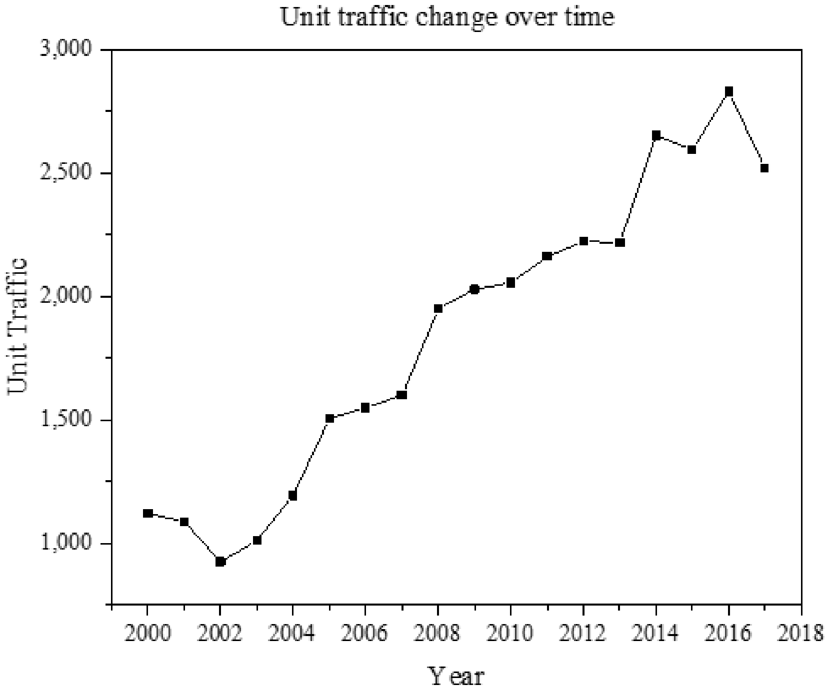


FIGURE 5.2: Traffic unit growth in the Saudi railway systems.

5.7 Discussion

This section discusses the efficiency results of the railway system in Saudi Arabia. In addition, it discusses the possible rail policies that can be applied to the Saudi railway system. There is traffic unit growth in the unit traffic over the sample period, as shown in Figure 5.2. This is mainly for freight rail services, but the passenger rail services are quite constant with some decline. Also, the data shows that there is no balance between the rail services, where freight output is greater than passenger output. This could be explained by limited network availability and fuel prices. Moreover, the rail network is limited to some part of the kingdom (i.e., Eastern and northern regions). Although the Saudi government has a strategic plan to expand the rail network by constructing several projects for passenger and freight sectors (e.g., the land bridge which connects the east with the west) as discussed in Section 3.2.2, there is no specific dates for these projects to be completed. Compared to the sample, the traffic unit has a quite steady pattern over time, and none of the passenger or freight outputs has an influence like the Saudi railway system.

For the efficiency estimation, the railway industry in Saudi Arabia scored a value of 0.4379, 1 and 0.4959 as an average efficiency estimated by COLS, DEA under VRS and CRS, respectively. The difference in the DEA efficiency estimation between VRS and CRS indicates scale inefficiency with a score of 0.4959.

To assess the rail policies, the Saudi government had organised its railway system in a form of vertical integration under SRO without any forms of horizontal separation or competition introduction. A new railway operator has been introduced, SAR, in 2006. SAR started in operation in 2011 to transport several types of minerals in the north region, and the passenger services were introduced in 2017. This indicates that the railway industry is organised in a form of vertical integration with a regional division of rail services. However, in 2019, there has been a merger in which SRO assets have been transferred to SAR. Thus, the current railway organisation can be described as vertical integration under the SAR operation.

The results show that rail policies have various impacts on technical inefficiency. The organisational and institutional separation improve the efficiency levels. Thus, both forms of the vertical separation policy are recommended for the Saudi railway industry. As the impact of vertical separation was assessed in conjunction with traffic density, Table 5.6 presents the average values of traffic density measured as traffic units in millions per network-line length in km for the railway systems. According to this table, the Saudi railway system has an average traffic density of 1.452, which is substantially below the average of the sample size (4.405). This suggests that separating IMs from RUs could increase technical efficiency. For the horizontal dimension, the regional division of rail services has a negative impact on inefficiency levels. This policy could not fit the Saudi railway industry. In fact, merging SRO into SAR can be described as a better reform in which the technical inefficiency can be reduced[6].

On the other hand, the results of the technical assessment do not support the market function separation, but the literature shows the separation between passenger and freight rail services increases the technical efficiency. This policy might be recommended for the Saudi railway system based on the findings of the literature. However, the results show that the horizontal separation combined with the vertical separation reduces the efficiency levels. The policies of competitive tendering for passenger services and free entry for freight services increase the levels of inefficiency. Hence, the policies of introducing competition are not recommended for the Saudi railway industry. Moreover, passenger and freight rail services are recommended to be provided by a single company without any form of competition. This option could improve the performance efficiency of the Saudi railway system. As the Saudi government plans to construct a land-bridge rail project, it is recommended to operate this project without any form of competition alongside the existing railway system. Thus, it is not recommended to reform the Saudi railway industry by introducing vertical separation and market function separation in one package.

[6] It should be noted that this research does not include SAR in the assessment due to the lack of data.

Table 5.6: Average values of traffic density for the railway systems.

Country	No obser.	Traffic density	Ratio
Algeria	18	0.666	0.151
Austria	14	4.923	1.118
Azerbaijan	16	3.928	0.892
Belgium	10	4.737	1.075
Bulgaria	18	1.361	0.309
Czech Rep.	18	2.139	0.485
Finland	18	2.300	0.522
France	14	3.975	0.902
Germany	18	4.602	1.045
Greece	7	0.951	0.216
Iran	16	4.471	1.015
Ireland	11	1.073	0.244
Italy	15	3.736	0.848
Japan	16	13.603	3.088
Kazakhstan	15	14.575	3.309
Latvia	12	7.186	1.631
Morocco	18	4.441	1.008
Pakistan	18	3.304	0.750
Poland	15	2.893	0.657
Portugal	9	2.077	0.472
Saudi Arabia	18	1.452	0.330
Slovakia	18	3.033	0.688
Slovenia	18	3.370	0.765
South Korea	18	10.851	2.463
Spain	18	2.244	0.509
Switzerland	18	7.818	1.775
Turkey	17	1.608	0.365
Turkmenistan	5	3.995	0.907
Uzbekistan	8	5.323	1.208
Average		4.405	1.000

5.8 Conclusion

This research aimed to evaluate the railway reform via technical assessment. Through the technical assessment, this research project has estimated the technical efficiency and productivity. This estimation relied on parametric and non-parametric approaches. Based on the efficiency estimation, it appeared that the non-parametric approach, DEA, produces higher average efficiency levels compared to the parametric approach. Despite this difference, the efficiency rankings of railway systems seemed acceptable based on the results of Spearman's rank correlation.

The efficiency estimation is then followed by determinants of inefficiency levels. These determinants are clustered between control and test variables. The test variables represent the forms of rail policies. The results show that the policy of vertical separation, whether organisational or institutional separation, is more beneficial in terms of reducing inefficiency levels compared to vertical integration. However, this effect is changed in conjunction with the traffic density levels. The vertical separation reduces the inefficiency at low density levels but increases the inefficiency at high density levels, where vertical integration is more beneficial. On the other hand, the other forms of rail policies, such as competition introduction and horizontal separation, increase the inefficiency levels. Finally, combining the three forms of rail policies, competition introduction, vertical and horizontal separation, increases the technical inefficiency levels.

Chapter 6

Cost Assessment

6.1 Introduction

In this chapter, this research evaluates railway reforms via conducting a cost assessment. This assessment focuses on the input prices and outputs of the railway systems to estimate their efficiency. Through this estimation, it is assumed that the primary objective of a railway system is to minimise the cost. In this regard, this research employs the concept of cost efficiency, where it is required to estimate the minimum cost for different railway systems. The cost efficiency can be estimated by parametric and non-parametric approaches, as discussed in the Methodology Chapter. The two approaches are used in order to check the robustness of the estimation and also to draw a comparison between the methods. The cost efficiency is then followed by an estimation of the effect of railway reform on cost efficiency. This could help to identify which railway organisation can achieve cost minimisation.

6.2 Literature Review

To review the empirical literature, there are numerous studies aimed to assess the impacts of railway reform on cost minimisation. These studies differ based on estimation methods and empirical data. Cost efficiency has been examined in terms of the application of railway reform. The first reform to discuss is competition introduction. As there are two forms of competition, the discussion starts with the impact of competitive tendering on cost efficiency. Asmild et al. (2009) showed that competitive tendering has a significant negative impact on material costs. In other words, competitive tendering increases material costs. This could be explained as passenger services having more attention from the political side to be more attractive. van de Velde et al. (2012) and Mizutani et al. (2015) concluded that the effect of competitive tendering on cost reduction is not statistically significant. This means that competitive tendering does not affect cost minimisation. Smith et al. (2018) found that competitive tendering has a significant impact on cost reduction. There is a cost reduction of around 7% due to competition in the market for the passenger sector, but this finding was concluded when the authors considered the role of economic regulation in the analysis.

Besides the impact of competitive tendering, there is an effect of on-track competition on cost efficiency. Asmild et al. (2009) found that market deregulation for the freight sector has a significant positive effect on staff costs. In other words, freight market deregulation reduces staff costs. This might be explained by the fact that the freight sector has more flexibility in adjusting staff expenditures than the passenger sector. van de Velde et al. (2012) and Mizutani et al. (2015) showed that the impact of freight market deregulation on railway costs is not statistically significant. This indicates that introducing competition in the freight market does not have any impact on cost reduction. Smith et al. (2018) found similar results to the previous literature. The impact of freight market deregulation on cost reduction is not statistically significant, which means that opening the market to new entrants does not lead to cost reduction.

Similarly to the discussion of competition introduction, structural reform has an effect on cost efficiency. The first form to discuss is vertical reform. Jensen and Stelling (2007) found that vertical reforms have an impact on cost performance. The vertically separated organisation increases the railway costs, but there is a cost reduction when vertical reforms are introduced in conjunction with the introduction of competition. Asmild et al. (2009) found that vertical reforms have varied impacts on cost efficiency. Accounting separation has a significant downward impact on staff and material costs, while there are no significant impacts caused by the full vertical separation. The authors concluded that accounting separation is more appropriate to improve efficiency compared to complete vertical separation.

The impact of vertical reforms on cost efficiency, however, has been assessed in terms of the level of train density. van de Velde et al. (2012) and Mizutani et al. (2015) found that the impact of vertical reforms on cost reduction varies depending on the organisational forms. The first form to consider is a vertically separated organisation; the impact of this organisation on cost reduction is not significantly different from zero when compared to vertical integration, but this impact is observed differently when the level of train density is taken into account. The vertically separated organisation reduces the total cost at lower levels of train density, but there is an increase in railway costs as train density increases. The second form is the vertical separation within a holding company; the effect of this organisation on cost reduction is statistically significant, which means that railway systems that are organised in the form of vertical separation within a holding company tend to reduce costs by 5% compared to vertical integration. When the impact on cost in terms of train density is considered, the effect of vertical separation within a holding company does not vary with train density. To conclude, the authors preferred the vertically separated organisation within a holding company to achieve cost minimisation because the impact of this organisation does not change with respect to changes in the level of train density. Mizutani and Uranishi (2013) concluded that vertical reforms have different impacts on cost reduction in terms of train density. Firstly, vertical separation, in general, tends to reduce the railway costs, but this impact is varied when the level of train density is considered in the assessment. At lower levels of train density, railways that are organised in a vertically separated form reduce the total cost, while this organisational form tends to increase costs as train density increases. In contrast, railways that are organised in the form of vertically integrated systems have the opposite impact on cost reduction in terms of train density levels.

As well as examining vertical reforms with respect to train density, the power of independent regulatory bodies and economic regulation have been examined in terms of cost efficiency. Smith et al. (2018), for example, found that strong economic regulation has a positive impact on cost reduction. Regarding vertical reforms, the impact of institutional and organisational separation on cost reduction is similar to van de Velde et al. (2012) and Mizutani et al. (2015). However, the main finding for Smith et al. (2018) is that strong economic regulation is capable of reducing costs for a vertically separated organisation at higher levels of train density than previously envisaged.

The second form of structural reform is horizontal reform. van de Velde et al. (2012) showed that horizontal reforms have a positive impact on cost reduction. Moreover, the horizontal separation between the passenger and freight sectors reduces the costs of railways. Mizutani and Uranishi (2013) found that horizontal reforms have a statistically significant effect on cost reduction. Moreover, the horizontal separation between passenger and freight rail services tends to reduce the railway costs regardless of the level of train density. Mizutani et al. (2015) showed that horizontal reforms have an impact on cost reduction. The horizontal separation between passenger and freight services reduces the total cost of railway systems by 24% compared to horizontal integration. Smith et al. (2018) found similar results for the impact of horizontal separation on cost reduction compared to the previous literature.

6.3 Methodology

It has been discussed in the technical assessment that railway industries are classified as a multiple-output market. One approach to dealing with this market is to assume cost minimisation behaviour. Thus, the cost function is a tool for calculating the cost of a firm by including a set of output quantities and input prices under an assumption of cost minimisation behaviour based on the concept of cost efficiency. This research employs parametric and non-parametric approaches to estimate the cost efficiency.

6.3.1 Parametric Approach

In the parametric approach, the cost efficiency can be estimated via several methods (e.g., COLS, SUR, and SFA). This research selects the SFA as a parametric approach for the following reasons. The assumption underlying the COLS method for fitting a cost function is that all deviations from the frontier result from inefficiency. This is a major drawback of the deterministic methods because they fail to account for the potential impact of data noise, such as that caused by measurement error or model miss-specification. On the other hand, the SUR estimation presumes that all railway systems have successfully reached the frontier. In fact, the SUR method does not directly estimate the cost efficiency, and Cowie (2002), for example, estimated the cost efficiency by applying the COLS approach to the residuals that are obtained by SUR estimation.

To illustrate the estimation of the cost efficiency via the SFA, recall the translog cost function that is stated in Equation (4.26), which is written as follows:

$$\begin{aligned}\ln(C_i^A) = \alpha_0 &+ \sum_{m=1}^{M} \alpha_m \ln(y_{mi}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln(y_{mi}) \ln(y_{ni}) \\ &+ \sum_{j=1}^{J} \beta_j \ln(w_{ji}) + \frac{1}{2}\sum_{j=1}^{J}\sum_{l=1}^{J} \beta_{jl} \ln(w_{ji}) \ln(w_{li}) \\ &+ \sum_{m=1}^{M}\sum_{j=1}^{J} \delta_{mj} \ln(y_{mi}) \ln(w_{ji}) + \varepsilon_i\end{aligned} \tag{6.1}$$

where C^A is the total cost; y is a vector of outputs; w is a vector of input prices; and α, β and δ are model parameters to be estimated. Following the concept of the SFA, the error term (ε) will be replaced with the statistical noise (v) and inefficiency (u) terms. Thus, the cost function can be rewritten as follows:

$$\begin{aligned}\ln(C_i^A) = \alpha_0 &+ \sum_{m=1}^{M} \alpha_m \ln(y_{mi}) + \frac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{M} \alpha_{mn} \ln(y_{mi}) \ln(y_{ni}) \\ &+ \sum_{j=1}^{J} \beta_j \ln(w_{ji}) + \frac{1}{2}\sum_{j=1}^{J}\sum_{l=1}^{J} \beta_{jl} \ln(w_{ji}) \ln(w_{li}) + \theta_t T + \frac{1}{2}\theta_{tt} T^2 \\ &+ \sum_{m=1}^{M}\sum_{j=1}^{J} \delta_{mj} \ln(y_{mit}) \ln(w_{jit}) + \sum_{m=1}^{M} \varphi_{mt} \ln(y_{mit}) T \\ &+ \phi_{jt} \sum_{j=1}^{J} \ln(w_{jit}) T + v_{it} + u_{it}\end{aligned} \tag{6.2}$$

It should be noted that the time-trend variable has been introduced to the cost model. To estimate the SFA with a panel data, there are several stochastic frontier models that can be applied to estimate the cost efficiency with a panel data, as listed in Table 4.1. It is noteworthy that the half-normal model proposed by Aigner et al. (1977) can be apply to estimate the cost efficiency. This approach has been applied by Farsi et al. (2005) and Bougna and Crozet (2016)[1]. This model neglects the panel effect and estimates the efficiency as a pooled model. The remaining of the stochastic frontier models will be estimated, and the results are discussed below.

6.3.2 Non-parametric Approach

In the parametric approach, the production technology set was specified and represented by an explicit production function with input prices. However, it is possible to measure the cost efficiency with unspecified functional form by using DEA. Recall Equation (4.21) from Section 4.3.3 that represents the cost minimisation problem. This equation is written as follows:

$$C(x,y) = \min_{x} \{wx : x \in L(y)\} \tag{6.3}$$

The linear programming problem that is stated in Equation (4.51) can be solved to estimate the minimum cost. Similar to the technical efficiency estimation, this problem is modified by adding a network constrain, as discussed in Section 5.3.1. Thus, the minimum cost can be driven through solving the DEA linear programming problem as follows:

$$\begin{aligned} C_0^* &= \min \sum_{i=1}^{N} x_{i0} w_{i0} \\ s.t. \sum_{j=1}^{N} \lambda_j x^j &\leq x_0; \\ \sum_{j=1}^{N} \lambda_j y^j &\geq y_0; \\ \sum_{j=i}^{N} \lambda_i LL^j &\leqslant \theta LL_0; \\ \sum_{j=1}^{N} \lambda_j &= 1; \\ \lambda_j \geq 0(j &= 1, 2, \ldots, N) \end{aligned} \tag{6.4}$$

[1]The authors employed the pooled half-normal model to estimate the technical efficiency.

To estimate the minimum cost with constant returns to scale, $\sum_{i=1}^{N} \lambda_i = 1$ is removed from the DEA problem. Once the minimum cost obtained from solving the linear problem, the cost efficiency of the selected railway systems can then be measured as follows:

$$CE_0 = \frac{C_0^*}{C_0} \tag{6.5}$$

6.3.3 Returns to Scale and Density

Commonly for rail cost studies, the firm size can be determined through measurements of the level of outputs (Caves et al., 1985). For the railway industry, the network size has a strong influence on the firm size. Thus, the cost per unit of output could be obtained differently among firms, depending on the nature of the railway network. Thus, the cost model can be written by including the network effect (N) as follows:

$$C^A = C^*(Y, N, W, T) \tag{6.6}$$

Returns to Density (RTD) can be defined as the change in outputs with respect to the change in inputs. However, because Equation 6.6 contains the network effect to calculate the cost, this requires defining whether the change in inputs leads to a change in outputs only or a change in both output and network. This leads to distinguishing between RTD and RTS. RTD can be described as the possibility of the change in outputs due to the change in inputs, with network and input prices held fixed. Thus, RTD can be measured as the inverse of the sum of the elasticities of the total cost with respect to the outputs:

$$RTD = 1 \Bigg/ \sum \frac{\partial \ln(C)}{\partial \ln(y)} \tag{6.7}$$

The RTD can be interpreted as greater than 1 means increasing returns to density, equal to 1 means constant returns to density and less than 1 means decreasing returns to density. RTS, on the other hand, can be obtained as the inverse of the sum of the elasticities of the total cost with respect to the outputs and network:

$$RTS = 1 \Bigg/ \left(\sum \frac{\partial \ln(C)}{\partial \ln(y)} + \sum \frac{\partial \ln(C)}{\partial \ln(N)} \right) \tag{6.8}$$

The RTS can be interpreted as greater than 1 means increasing returns to scale, equal to 1 means constant returns to scale and less than 1 means decreasing returns to scale.

6.4 Cost Model and Data

This research estimates the cost efficiency of railway systems. There are two classifications of railway costs to build a cost model, as highlighted by Mizutani (2004). The first classification is variable costs, which can be described as the costs that increase or decrease depending on the level of output produced. Generally, these costs represent two inputs, which are labour and energy consumed by rolling stock. Caves et al. (1985) used these cost variables in their cost model. Although this classification represents the change in response to the level of rail service outputs, the shortcoming is omitting some cost variables that could be varied with rail service outputs. This means that including only labour and energy costs could lead to an underestimation of the rail costs.

The second classification is total costs, which includes the costs of equipment, materials, and supplies besides labour and energy costs. This classification encloses all types of costs in the rail industry, but it is difficult to separate the cost of fixed components from equipment, materials, and supplies, particularly maintenance costs for rail infrastructure. A solution proposed by Savage (1997) is to allocate maintenance costs for rolling stock. However, this solution requires a level of detail in data that is sometimes difficult to obtain.

As this research involves railway systems with different organisational forms, this raises a concern about how the costs of each industry will be included in the estimation. Based on Mizutani and Uranishi (2013), van de Velde et al. (2012) and Mizutani et al. (2015), the total costs are the sum of energy, labour, capital, and material costs. For the vertically integrated organisation, the observed total cost will be used. On the other hand, the total costs of the vertically separated organisation will be used as the sum of the total costs of the train operating company and infrastructure manager. Mizutani and Uranishi (2013) proposed an approach to involve the total costs of the infrastructure manager in the capital costs. This approach seems reasonable to reflect the total costs of each industry without any underestimation. However, due to data availability, this research cannot calculate the total costs as mentioned in earlier studies. Thus, this research relies on the first classification of railway costs (i.e., only examines variable operating costs). This gives an indication of some cost categories that are missing, which could be classified as capital costs (e.g., financial expenses, TACs, depreciation), as highlighted by Mizutani and Uranishi (2013). These costs account for 30-37% of the total costs[2] of van de Velde et al. (2012), Mizutani and Uranishi (2013), Mizutani et al. (2015) and Smith et al. (2018).

This research builds a multiple-output cost model which includes passenger and freight outputs, network-line length, labour, and energy and material prices, and technology

[2] The percentage was obtained based on the estimated coefficient of the capital cost in the cost model.

index. Passenger carried (y_1) and freight carried (y_2) measured as pass-km and tonne-km, respectively, represent the output measure of the rail system. These two variables are commonly used in the literature as the final output measure of the rail services for the same reason that is mentioned in Section 5.4.

Network-line length (N) represents the network variable. Due to a lack of data, this variable is included as a network-line length in km rather than a network-track length. Although network-line length does not represent the rail infrastructure like the network-track, data availability is the reason for selecting this measure.

Energy and material prices (w_1) are included to represent the energy and material costs per rolling stock. This variable is calculated as the energy and material costs divided by the number of rolling stock. Labour price (w_2) is included to represent the labour costs per employee. This variable is calculated by dividing the labour costs by the total number of employees, as used by Mizutani and Uranishi (2013), van de Velde et al. (2012) and Mizutani et al. (2015).

Technology index (T) is included to represent the technology, based on previous studies such as Mizutani and Uranishi (2013), van de Velde et al. (2012) and Mizutani et al. (2015). Mizutani and Uranishi (2013) define the potential use of this variable to represent technological progress or time-trend. The technological progress could be described by the percentage of Automatic Train Stop (ATS), Automatic Train Control (ATC) and electrified line length (Mizutani and Uranishi, 2013). Due to the data availability, it is possible to use the percentage of electrified lines as a technology index, following the previous studies. However, after testing the cost model with the percentage of electrified lines and time-trend, the results more closely comply with the theory when the time-trend is included. Therefore, the technology index is represented by the time-trend. Finally, the cost model is written as follows:

$$
\begin{aligned}
\ln\left(\frac{VC_{it}}{w_{2it}}\right) = {} & \alpha_0 + \alpha_1 \ln(y_{1it}) + \alpha_2 \ln(y_{2it}) + \frac{1}{2}\alpha_{11}\ln(y_{1it})^2 \\
& + \alpha_{12}\ln(y_{1it})\ln(y_{2it}) + \frac{1}{2}\alpha_{22}\ln(y_{2it})^2 + \beta_1 \ln\left(\frac{w_{1it}}{w_{2it}}\right) \\
& + \frac{1}{2}\beta_{11}\ln\left(\frac{w_{1it}}{w_{2it}}\right)^2 + \vartheta_N \ln(N_{it}) + \frac{1}{2}\vartheta_{NN}\ln(N_{it})^2 + \theta_t T_{it} \\
& + \frac{1}{2}\theta_{tt}T_{it}^2 + \delta_{11}\ln(y_{1it})\ln\left(\frac{w_{1it}}{w_{2it}}\right) + \delta_{21}\ln(y_{2it})\ln\left(\frac{w_{1it}}{w_{2it}}\right) \\
& + \xi_{1N}\ln(y_{1it})\ln(N_{it}) + \xi_{2N}\ln(y_{2it})\ln(N_{it}) \\
& + \psi_{1N}\ln\left(\frac{w_{1it}}{w_{2it}}\right)\ln(N_{it}) + \varphi_{1t}\ln(y_{1it})T + \varphi_{2t}\ln(y_{2it})T \\
& + \phi_{1t}\ln\left(\frac{w_{1it}}{w_{2it}}\right)T_{it} + \zeta_{Nt}\ln(N_{it})T_{it} + v_{it} + u_{it}
\end{aligned}
\tag{6.9}
$$

where VC is variable costs; y_1 is passenger carried measured as pass-km; y_2 is freight carried measured as tonne-km; w_1 is price of energy and material consumption; w_2 is price of labour; N is network-line length; t is time-trend; $\alpha, \beta, \vartheta, \theta, \delta, \xi, \varphi, \psi, \phi, \zeta$ are model parameters to be estimated.

To conduct the cost assessment, a sufficient amount of data is required to estimate the cost efficiency. The data sources are the International Union of Railways (UIC), the World Bank, and annual reports by each railway company. These sources are commonly used in literature for conducting railway studies, especially UIC. As the research project involves countries from Europe, North Africa, East Asia, and the Middle East, it is difficult to obtain the data in case of missing data from online sources.

Table 6.1 contains the average values of all variables used for the cost models. According to this table, there is a large variation between the railway systems. For example, France, Germany, and Italy can be seen as large railway systems in terms of variable costs compared to Algeria, Greece, Ireland, and Portugal. In addition, the railway systems in France and Germany have higher rail outputs compared to those in Algeria, Ireland, and Saudi Arabia.

6.5 Results

6.5.1 Cost Model Estimation

This research applied the SFA and DEA to estimate the cost efficiency of railway systems. For the SFA estimation, the cost model has been estimated by several stochastic frontier models by using the Stata written commands by Kumbhakar et al. (2015). The Schmidt and Sickles (1984) and Pitt and Lee (1981) models are rejected for the same reasons that are discussed in Section 5.3.2. On the other hand, the pooled half-normal model is rejected because the LR test has a value of 4.852, which is lower than the critical value of 5.412. This means that the null hypothesis is not rejected. In other words, the OLS estimation is sufficient. The Kumbhakar (1990) and Battese and Coelli (1995) models do not converge to obtain the optimal solution for maximising the log-likelihood function. The only model that was able to produce an output and satisfies the LR test is the Battese and Coelli (1992) model. This means that proposed parameters (i.e., η_1 and η_2) to measure time-varying efficiency are statistically different from zero.

Prior to the SFA estimation, all variables have been normalised by their sample mean before the estimation. This approach has been applied by Coelli and Perelman (2000), Kumbhakar et al. (2007), Last and Wetzel (2010), Bougna and Crozet (2016). In this case, the coefficients of the first-order parameters can be interpreted as elasticities

Table 6.1: Average values of cost model variables at country level.

Country	**Obser**	$VC^{a,b}$	y_1^a	y_2^a	N	w_1^b	w_2^b
Algeria	10	416	1,122	1,332	3,935	8,658	27,264
Austria	12	2,640	9,464	16,593	5,321	24,282	46,575
Belgium	9	3,420	9,098	7,502	3,497	97,142	51,994
Bulgaria	18	529	2,219	3,468	4,147	35,780	7,534
Czech Rep.	15	2,460	6,833	13,902	9,469	36,612	19,028
Finland	18	680	3,749	9,797	5,888	32,612	36,349
France	9	15,100	80,856	36,749	29,575	193,593	38,425
Germany	18	33,800	75,854	81,578	34,273	176,257	45,809
Greece	7	477	1,761	610	2,491	42,552	37,620
Ireland	5	347	1,621	408	1,919	70,647	36,095
Italy	15	8,100	43,821	17,192	16,377	87,777	45,124
Morocco	13	419	3,422	5,150	1,982	36,994	20,521
Pakistan	18	948	21,715	4,031	7,791	39,190	2,107
Poland	15	4,560	16,644	40,819	19,814	30,633	18,035
Portugal	9	452	3,575	2,318	2,837	69,103	18,463
Saudi Arabia	18	607	340	1,506	1,249	204,812	71,124
Slovakia	17	1,100	2,485	8,560	3,636	42,576	16,234
Slovenia	18	527	755	3,352	1,219	62,555	33,829
South Korea	17	3,530	27,109	10,078	3,506	109,698	62,145
Spain	13	2,650	20,714	9,743	13,401	98,599	33,006
Switzerland	17	3,330	15,383	9,029	3,108	89,341	67,375
Turkey	17	1,890	5,134	9,618	9,190	51,849	30,402
Total	308						

[a] The variables are expressed in millions.
[b] The variables are expressed in PPP USD.

estimated at the sample means. The labour price (w_2) has been selected for the price homogeneity.

Table 6.2 reports the SFA estimation of the cost model. First, the monotonicity and concavity conditions are diagnosed. For the outputs, the coefficients of the first-order output (α_1 and α_2) are statistically significant at a 99.9% confidence level and have the expected sign. At the means, it appears that the outputs are non-negative, which means that the monotonicity conditions on the outputs are satisfied. The coefficients of first-order input price (β_1 and β_2) are statistically significant at a 99.9% confidence level and have the expected sign[3]. These coefficients are non-negative, and this indicates that the monotonicity conditions in the input prices are satisfied at the means. On the other hand, the coefficient of the second-order input prices, cross-product term, (β_{12}) has the expected sign. That is, the concavity condition in the input price is satisfied at the sample means. Thus, the estimated model satisfies the monotonicity and concavity conditions at the means.

[3]It should be noted that the coefficient of labour price is assumed to be statistically significant because it has been calculated via the homogeneity restrictions.

Table 6.2: Cost model estimated by SFA.

Term	Parameter	SFA	Term	Parameter	SFA
$\ln(y_1)$	α_1	0.271***	$\ln(y_1)\ln(w_1)$	δ_{11}	-0.097***
		(0.073)			(0.019)
$\ln(y_2)$	α_2	0.197***	$\ln(y_1)\ln(w_2)$	δ_{12}	$\underline{0.097}$
		(0.046)			
$\ln(y_1)^2$	α_{11}	0.057	$\ln(y_2)\ln(w_1)$	δ_{21}	0.097***
		(0.052)			(0.021)
$\ln(y_1)\ln(y_2)$	α_{12}	-0.252***	$\ln(y_2)\ln(w_2)$	δ_{22}	$\underline{-0.097}$
		(0.035)			
$\ln(y_2)^2$	α_{22}	0.199***	$\ln(y_1)\ln(N)$	ξ_{1N}	0.269***
		(0.027)			(0.060)
$\ln(w_1)$	β_1	0.440***	$\ln(y_2)\ln(N)$	ξ_{2N}	0.108
		(0.025)			(0.057)
$\ln(w_2)$	β_2	$\underline{0.560}$	$\ln(w_1)\ln(N)$	ψ_{1N}	-0.042
					(0.035)
$\ln(w_1)^2$	β_{11}	0.180***	$\ln(w_2)\ln(N)$	ψ_{2N}	$\underline{0.042}$
		(0.025)			
$\ln(w_1)\ln(w_2)$	β_{12}	$\underline{-0.180}$	$\ln(y_1)T$	φ_{1t}	0.006*
					(0.002)
$\ln(w_2)^2$	β_{22}	$\underline{0.180}$	$\ln(y_2)T$	φ_{2t}	0.001
					(0.003)
$\ln(N)$	ϑ_N	0.319**	$\ln(w_1)T$	ϕ_{1t}	-0.006*
		(0.095)			(0.002)
$\ln(N)^2$	ϑ_{NN}	-0.773***	$\ln(w_2)T$	ϕ_{2t}	$\underline{0.006}$
		(0.169)			
T	θ_t	-0.022**	$\ln(N)T$	ζ_{Nt}	0.003
		(0.006)			(0.003)
T^2	θ_{tt}	0.000	Constant	α_0	-0.677***
		(0.000)			(0.073)
N		308			
σ		0.133***			
		(0.013)			
γ		0.942***			
		(0.006)			
η_1		0.719***			
		(0.124)			
η_2		-0.010**			
		(0.004)			
LL		246.321			
LR		612.644			

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.
Standard errors in parentheses.
Underlined coefficients are obtained by homogeneity restrictions.

As the cost model is expressed in the natural logs, the coefficient of the model parameters can be interpreted as elasticities. The elasticities of the passenger-km (α_1) and tonne-km (α_2) outputs are found to be equal to 0.271 and 0.197, respectively. These parameter values suggest that a one percent increase in passenger-km and tonne-km will lead to an increase in the costs by 0.271% and 0.197%, respectively. The difference between the two elasticities is 0.074 (0.271 - 0.197 = 0.074). This percentage indicates that the freight output has a lower marginal effect on the cost by 7.4% than the passenger output.

The elasticities of energy and the material prices (β_1) and the labour prices (β_2) are found to be equal to 0.440 and 0.560, respectively. The estimation of the cost model indicates that energy and material costs account for 43.8% and labour costs account for 56.2% of the variable costs at the sample mean. These estimations might be inaccurate because they are calculated as an average. For example, the labour price is calculated by dividing the labour costs by the average annual number of employees. Similarly, the energy and material price is calculated by dividing the energy and material cost by the total number of rolling stock. Based on these calculations, the input prices might not represent the actual measurements.

The elasticity of the network-line length (ϑ_N), which represents the scale of the network, is statistically significant at a 99% confidence level. This indicates that railway systems with larger networks tend to increase costs, which implies that a one percent increase in the network size will lead to a 0.319% increase in costs. This effect is expected because larger networks might have more complexity in terms of the number of stations, level of crossings, etc. for operation. Lastly, the first-order time-trend parameter (θ_t) is -0.020 and statistically significant at a 99% confidence level. The negative sign indicates a decrease in cost over time.

6.5.2 Efficiency Levels

The measurements of the cost efficiency are estimated by DEA and SFA. These measurements over the sample period are shown in Figure 6.1. According to this figure, it appears that DEA under the VRS assumption has higher cost efficiency levels (0.7222 as an average value) compared to other methods. The average values of cost efficiency obtained by DEA under VRS and CRS have a relatively similar pattern of change over the sample period. The SFA estimation has lower average efficiency scores compared to the DEA estimation. This variation is expected because DEA measures the efficiency compared to the best practice of decision-making units, which is identified as the frontier, and these units obtain 1 as cost efficiency. Thus, DEA estimates higher average efficiency scores compared to other methods.

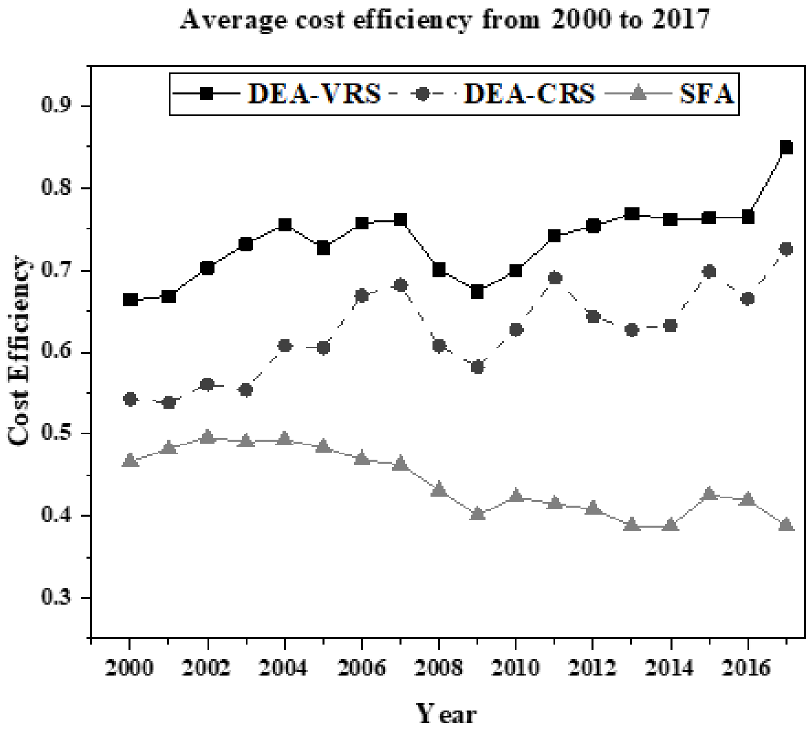


FIGURE 6.1: Average cost efficiency levels over the sample period

TABLE 6.3: Spearman rank-order correlation coefficient.

	DEA^{CRS}	DEA^{VRS}	SFA
SFA	1		
TE^{VRS}	0.8339*	1	
TE^{CRS}	0.5316*	0.5589*	1

* indicates significance at the 1% level.

To check the ranking similarity between the railway systems, the Spearman rank-order correlation coefficient between the efficiency scores obtained by DEA and SFA is reported in Table 6.3. If the coefficient has a value close to one, it indicates that railway systems have similar efficiency rankings obtained by the DEA and SFA estimations. The results show a positive and strong correlation between the efficiency scores estimated by the parametric and non-parametric approaches. In other words, the relative positions of the railway industries regarding efficiency scores are consistent in general.

The cost efficiency measurements of each railway system are tabulated in Table 6.4. A Kruskal-Wallis test has been conducted to assess the variation in the cost efficiency between railway systems. The test has been performed on each estimation approach. The results conclude rejection of the null hypothesis at the 1% level. In other words, the measurements of the cost efficiency between the railway systems are not at the same level. For the DEA estimation under VRS, five railway systems are located on

TABLE 6.4: Average cost efficiency levels per country estimated by DEA and SFA

Country	DEA^{VRS}	DEA^{CRS}	SFA
Algeria	0.2449	0.1607	0.2764
Austria	0.7871	0.5338	0.4090
Belgium	0.4772	0.4686	0.3578
Bulgaria	0.3742	0.3124	0.2698
Czech Rep.	0.4383	0.3684	0.2886
Finland	1	0.9963	0.9336
France	1	0.9117	0.4710
Germany	1	0.6978	0.2683
Greece	0.4925	0.2406	0.5748
Ireland	0.8506	0.3226	0.8343
Italy	0.7597	0.5593	0.4848
Morocco	0.9254	0.8818	0.5259
Pakistan	0.5730	0.5341	0.2966
Poland	0.7937	0.4746	0.2741
Portugal	0.7904	0.6425	0.7850
Saudi Arabia	1	0.7585	0.5005
Slovakia	0.5229	0.5121	0.3298
Slovenia	0.7133	0.6502	0.3145
South Korea	1	0.9949	0.4739
Spain	0.8702	0.8466	0.7950
Switzerland	0.7558	0.7368	0.3877
Turkey	0.5199	0.4794	0.4306

the cost frontier (i.e., Finland, France, Germany, Saudi Arabia, and South Korea). These railway systems can be described as cost efficient. The five most cost inefficient railway systems are Algeria, Bulgaria, the Czech Republic, Belgium, and Greece, where the efficiency score is below 0.5. For the DEA estimation under CRS, no railway systems are located on the cost frontier. Finland, France, and South Korea have railway systems that are located below the frontier by 0.1. The most inefficient are Algeria, Greece, Bulgaria, Ireland, and the Czech Republic, where the cost efficiency level is below 0.4. For the SFA estimation, Finland, Spain, Portugal, Ireland, and France are the top five railway systems that scored higher cost efficiency levels. In contrast, Bulgaria, Algeria, Slovenia, the Czech Republic, and Slovakia have the lowest efficiency levels.

6.5.3 RTS and RTD

This section discusses the estimation of RTS and RTD. There are two options to estimate the RTS and RTD. The first option is to define the railway outputs at the intermediate level by using train-km, while the second option is to define the railway outputs at the final level by using passenger-km and tonne-km. The first option seems reasonably accurate because the final measures are subjected to external effects (e.g.,

TABLE 6.5: Average values of RTS and RTD for each railway system.

Country	RTD	RTS
Austria	0.77	1.25
Belgium	3.81	1.16
Bulgaria	-10.48	0.84
Czech Rep.	5.14	0.91
Finland	1.09	0.98
France	-1.91	0.74
Germany	-2.42	0.77
Greece	1.24	1.07
Ireland	2.42	1.08
Italy	9.05	0.88
Morocco	1.79	1.05
Pakistan	-0.96	0.67
Poland	3.06	0.81
Portugal	-4.64	0.89
Saudi Arabia	-1.66	1.10
Slovakia	30.88	0.96
Slovenia	1.75	1.25
South Korea	2.64	1.16
Spain	-3.29	0.82
Switzerland	1.88	1.31
Turkey	2.34	0.81

the price of rail services and the competition of other transport modes). However, by following the first option, the sample size will be reduced from 308 observations to 279 observations for 21 railway systems.

To estimate the RTS and RTD, train-km has been introduced to the cost model. In this regard, the cost model is rewritten as follows:

$$\begin{aligned}\ln\left(\frac{VC_{it}}{w_{2it}}\right) &= \alpha_0 + \alpha_1 \ln(Q_{it}) + \frac{1}{2}\alpha_{11}\ln(Q_{it})^2 + \beta_1 \ln\left(\frac{w_{1it}}{w_{2it}}\right) \\ &+ \frac{1}{2}\beta_{11}\ln\left(\frac{w_{1it}}{w_{2it}}\right)^2 + \gamma_N \ln(N_{it}) + \frac{1}{2}\gamma_{NN}\ln(N_{it})^2 + \theta_T T_{it} \\ &+ \frac{1}{2}\theta_{TT}T_{it}^2 + \varphi_{11}\ln(Q_{it})\ln\left(\frac{w_{1it}}{w_{2it}}\right) + \lambda_{1N}\ln(Q_{it})\ln(N_{it}) \\ &+ \psi_{1T}\ln(Q_{it})T + \psi_{2T}\ln(y_{2it})T + \zeta_{1N}\ln\left(\frac{w_{1it}}{w_{2it}}\right)\ln(N_{it}) \\ &+ \phi_{1T}\ln\left(\frac{w_{1it}}{w_{2it}}\right)T_{it} + \delta_{NT}\ln(N_{it})T_{it} + \varepsilon_{it}\end{aligned} \tag{6.10}$$

where VC is variable costs; Q is train-km; w_1 is price of energy and material consumption, w_2 is price of labour; N is network-line length; t is time-trend; $\alpha, \beta, \gamma, \theta, \varphi, \lambda, \psi, \zeta, \phi, \delta$ are model parameters to be estimated. Table 6.5 reports the estimation of RTS and RTD for each railway system.

Following Preston (2001), the optimal size and density of a railway system can be determined by the inverse of network-line length and traffic density, respectively. First, the RTS is regressed against the inverse of network-line length, as shown in Equation (6.11). The network length is included because RTS is computed with respect to the network length. As this research aims to identify which optimal organisation can fit the Saudi railway system, this research attempts to determine the optimal size and organisation for each organisation (i.e., vertical integration, organisational and institutional separation). Having said that, the dummy variables are introduced in Equation (6.11). The optimal size then can be determined by assuming the RTS equals one (RTS = 1) in Equation (6.11).

$$RTS = \alpha_0 + \beta_1 N^{-1} + \beta_2 D_{IS} + \beta_3 D_{OS} + \varepsilon \tag{6.11}$$

where N is network-line length in km; D_{IS} is a dummy variable for institutional separation; D_{OS} is a dummy variable for organisational separation; and α and β are model parameters to be estimated.

Second, the RTD is regressed against the inverse of traffic density, as shown in Equation (6.12). The density is included because RTD is computed with respects to the traffic density. Again, the organisation-specific dummy variables are included in this equation to determine the optimal density for each organisation. The optimal density can be measured by assuming the RTS equals one (RTD = 1).

$$RTD = \alpha_0 + \beta_1 V^{-1} + \beta_2 D_{IS} + \beta_3 D_{OS} + \varepsilon \tag{6.12}$$

where V is traffic density; D_{IS} and D_{OS} are dummy variables for each organisation; and α and β are model parameters to be estimated. Table 6.6 reports the estimation results of Equation (6.11) and (6.12). According to these results, the adjusted R^2 of the RTD model has a relatively small value, which indicates that the RTD is poorly explained by this model. In fact, the F-test of this model is 0.059, which indicates that there is no relation between RTD and any independent variable. On the other hand, the adjusted R^2 of the RTS model has a value of 0.412, which indicates the RTS explains over 41.2% of the variation in the data. The results suggest that the optimal size of a railway system is 3,455 km, which is larger by 541 km than the findings of Preston (2001). This optimal size is for a vertically integrated organisation. To estimate the optimal size for other organisational forms, the coefficient of the institutional separation dummy variable is statistically significant at a 99% confidence level and has a negative value. This indicates that the optimal size for a vertically separated railway system is lower than for vertically integrated systems. The results suggest that the optimal size for a railway system that is separated organisationally is 2,180 km.

TABLE 6.6: Estimation of RTD and RTS regression models.

Term	Parameter	RTD	RTS
V^{-1}	β_1	-9.305	
		(30.187)	
N^{-1}	β_1		449.120***
			(37.758)
D_{IS}	β_2	0.593	-0.076***
		(4.352)	(0.022)
D_{OS}	β_3	0.415	0.014
		(4.944)	(0.022)
Constant	α_0	2.590	0.870***
		(5.005)	(0.020)
N		279	279
adj. R^2		0.001	0.412
F		0.059	64.164
LL		-1336.170	141.362

6.6 Effect estimation

In this section, this research aims to create a model that can explain cost efficiency in relation to different railway organisation forms. The procedure of this estimation is quite similar to the technical assessment that is explained in Section 5.6. In this case, the inefficiency determinante models are written as follows:

$$\begin{aligned}\ln(IEM) = {} & \alpha_0 + \alpha_1 \ln(z_1) + \alpha_2 \ln(z_2) + \alpha_3 \ln(z_3) + \alpha_4 \ln(z_4) + \beta_1 \ln(z_2) D_{OS} \\ & + \beta_2 \ln(z_4) D_{IS} + \beta_3 D_{OS} + \beta_4 D_{RS} + \beta_5 D_{OS} + \beta_6 D_{MFS} \\ & + \beta_7 D_{ONTC} + \beta_8 D_{OFTC} + \varepsilon\end{aligned} \tag{6.13}$$

$$\begin{aligned}\ln(IEM) = {} & \alpha_0 + \alpha_1 \ln(z_1) + \alpha_2 \ln(z_2) + \alpha_3 \ln(z_3) + \alpha_4 \ln(z_4) + \delta_1 D_{L1} + \delta_2 D_{L3} \\ & + \delta_3 D_{L4} + \delta_4 D_{L5} + \delta_5 D_{L6} + \varepsilon\end{aligned} \tag{6.14}$$

where IEM is the inefficiency cost score; z_1 is the GDP per capita; z_2 is the population density; z_3 is the network-line; z_4 is traffic density measured as traffic unit per network-line; D_{OS} is a dummy variable for organisational separation; D_{IS} is a dummy variable for institutional separation; D_{RS} is a dummy variable for regional separation; D_{MFS} is a dummy variable for market function separation; D_{OFTC} is a dummy variable for off-track competition; D_{ONTC} is a dummy variable for on-track competition; $D_{(LV.1)}$ is a dummy variable for any form of vertical separation only; $D_{(LV.3)}$ is a dummy variable for any form of competition only; $D_{(LV.4)}$ is any form of vertical separation combined with any form of horizontal separation; $D_{(LV.5)}$ is any form of vertical separation combined with any form of competition; $D_{(LV.6)}$ is any form of vertical separation combined with any form of horizontal separation and any form of competition; and α and β are model parameters to be estimated.

The estimation Model (6.13) and Model (6.14) are obtained by the Tobit model by using the Stata software and reported in Table 6.7. It should be noted that the cost inefficiency scores that are obtained by the SFA are used in this estimation. This is because of the shortcoming of the DEA estimation that is discussed in Section 5.6.

For the control variables, the population density (z_2) variable has been dropped from the two models because it is not statistically significant, but the remaining control variables are. The coefficient of GDP per capita (α_1) is statistically significant at a 99.9% confidence level and has a negative sign. This indicates that railway systems with higher GDP per capita are more cost efficient.

The coefficient of line density (α_3) is statistically significant at a 99.9% confidence level and has a positive sign. This indicates that railway systems with higher line density have a higher cost inefficiency. This finding is expected because line density represents the complexity of the rail infrastructure (e.g., crossing levels, stations, etc.), which requires more control compared to railway systems with lower line density.

The coefficient of traffic density (α_4) is statistically significant at a 99.9% confidence level and has a negative sign, although this coefficient is not statistically significant in the second model. In other words, railway systems with dense traffic have lower cost inefficiency levels. This finding is expected because more traffic means more utilisation of input resources. However, it is expected for the dense railway system to be more congested, which might increase the cost inefficiency levels in turn. The congestion effect could be captured by defining the traffic density as train per network-line in km.

Besides the control variables, the results of the test variables are reported in Table 6.7. The discussion of these results starts with the impact of rail policies individually. For the vertical dimension, the coefficient of institutional separation (β_3) is not statistically significant. This finding is in line with van de Velde et al. (2012) and Mizutani et al. (2015), where the impact is not statistically significant. In other words, railway systems that are organised vertically separated have lower cost inefficiency levels compared to the vertically integrated systems, but this difference is not statistically significant. On the other hand, the coefficient of organisational separation (β_4) is statistically significant at a 99.9% confidence level and has a negative sign. This finding is in line with Mizutani and Uranishi (2013)[4], van de Velde et al. (2012) and Mizutani et al. (2015) but in opposite to Jensen and Stelling (2007) and Asmild et al. (2009).

[4]It should be noted that Mizutani and Uranishi (2013) did not specify the degree of the vertical separation.

Table 6.7: Determinants of cost inefficiency levels.

Term	**Parameter**	**SFA**	
		Model (1)	**Model (2)**
$\ln(z_1)$	α_1	-0.285***	-0.235***
		(0.018)	(0.023)
$\ln(z_3)$	α_3	0.248***	0.166***
		(0.020)	(0.024)
$\ln(z_4)$	α_4	-0.240***	-0.071*
		(0.043)	(0.036)
$\ln(z_1)D_{OS}$	β_1	-0.045	
		(0.086)	
$\ln(z_1)D_{IS}$	β_2	0.498***	
		(0.096)	
D_{IS}	β_3	-0.155	
		(0.097)	
D_{OS}	β_4	-0.519***	
		(0.146)	
D_{RS}	β_5	0.635***	
		(0.067)	
D_{MFS}	β_6	0.139	
		(0.077)	
D_{ONTC}	β_7	0.220***	
		(0.051)	
D_{OFTC}	β_8	-0.374***	
		(0.062)	
D_{L1}	δ_1		-0.011
			(0.079)
D_{L3}	δ_2		0.285*
			(0.116)
D_{L4}	δ_3		0.396*
			(0.160)
D_{L5}	δ_4		0.119
			(0.065)
D_{L6}	δ_5		0.547***
			(0.077)
cons	α_0	2.527***	2.107***
		(0.161)	(0.200)
N		308	308
Pseud R^2		0.700	0.422
LL		-75.298	-145.354

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.

Standard errors in parentheses.

This research also evaluates the impact of the vertical separation forms with respect to traffic density. First, the coefficient of institutional separation with respect to traffic density (β_2) is statistically significant at a 99.9% confidence level and has a positive sign. This result indicates that vertical separation with high levels of traffic density increases the cost inefficiency levels. This finding is in line with Mizutani and Uranishi (2013), van de Velde et al. (2012) and Mizutani et al. (2015). It is worth noting that this research defines the traffic density as the sum of pass-km and tonne-km divided by line, whereas the other studies define the traffic density as train-km divided by network-track in km by 365. Regardless of how traffic density is defined, the conclusion is the same; vertical separation increases cost inefficiency at high levels of traffic density.

Second, the coefficient of organisational separation with respect to traffic density (β_1) is not statistically significant. This result means that the impact of vertical separation within a holding company does not vary with the level of train density. This finding is in line with van de Velde et al. (2012) and Mizutani et al. (2015). Again, regardless of how traffic density is measured, vertical separation within a holding company has no effect on the cost inefficiency levels associated with traffic density.

For the horizontal dimension, the coefficient of regional separation (β_5) is statistically significant at a 99.9% confidence level and has a positive sign. This means that railway systems with regional separation have higher levels of cost inefficiency compared to regionally integrated systems. On the other hand, the coefficient of market function separation (β_6) is not statistically significant, although it has a positive sign. In other words, railway systems with a separation between passenger and freight rail services have higher cost inefficiency levels compared with integrated services, but this difference is not statistically significant. This finding is opposite to Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018). This finding is unexpected because separating passenger from freight rail services could lead to cost minimisation. This can be explained by the fact that freight services tend to consume lower input resources compared to passenger services, which could lead to cost minimisation in turn. However, by comparing these results with the findings of others, the capital costs have been included in their cost models, and Mizutani and Uranishi (2013) explain the cost reduction that may result from the separation of market functions as the elimination of hidden costs (e.g., subsidies).Another possible explanation is that this research used the traffic unit outputs (i.e., passenger-km and tonne-km) to measure the cost efficiency. However, the literature found similar results in cost reduction when they used the traffic unit output in their cost models.

For the competition forms, the coefficient of on-track competition (β_7) is statistically significant at a 99.9% confidence level and has a positive sign. In other words, railway systems with free entry for freight services have higher levels of cost inefficiency compared to other systems. This finding is opposite to Asmild et al. (2009), van de Velde et al. (2012), Mizutani et al. (2015), and Smith et al. (2018). The difference between this research finding and the others is the sample size. This research involves railway systems from outside Europe (e.g., Algeria, Morocco, Pakistan, and Saudi Arabia), where the freight market is not deregulated for new entrants.

The coefficient of off-track competition (β_8) is statistically significant at a 99.9% confidence level and has a negative sign. This finding is in line with Smith et al. (2018)[5] and contrary to Asmild et al. (2009), van de Velde et al. (2012) and Mizutani et al. (2015)[6]. This indicates that railway systems with the franchise system for passenger rail services have lower cost inefficiency levels compared to other systems.

Besides the assessments of rail policies individually, this research examines the level of reforms with respect to the degree of combining rail policies, and Table 6.7 presents these results. Based on these results, combining rail policies appears to increase the cost inefficiency levels. The coefficient of level-04 (δ_3) is statistically significant at a 95% confidence level and has a positive sign. This indicates that railway systems that are reformed with a vertical separation combined with horizontal separation have higher cost inefficiency levels compared to not reformed systems. The coefficient of level-05 (δ_4) is not statistically significant, and it has a positive sign. This indicates that railway systems that have been reformed with a vertical separation combined with any form of competition have a higher cost inefficiency compared to not reformed systems, but this difference is not statistically significant. The coefficient of level-06 (δ_5) is statistically significant at a 99.9% confidence level and has a positive sign. This means that railway systems that have been reformed with the three forms have higher levels of cost inefficiency compared to non-reformed systems. In other words, completing the three forms of the railway reform appears to increase the cost inefficiency levels.

6.7 Discussion

This section discusses the results of the cost assessment for the Saudi railway system. The railway industry has faced a rapid increase in the traffic unit-km between 2000 and 2017, as discussed in Section 5.7. Figure 6.2 displays the changes in the outputs and costs of the Saudi railway system from 2000 to 2017. According to this figure, the changes in railway costs are not in line with the growth in the traffic unit. These variations in costs could be related to the changes in the Saudi railway system.

[5] Smith et al. (2018) included the power of regulation in their cost model.

[6] van de Velde et al. (2012) and Mizutani et al. (2015) found that the impact of competitive tendering is not statistically significant.

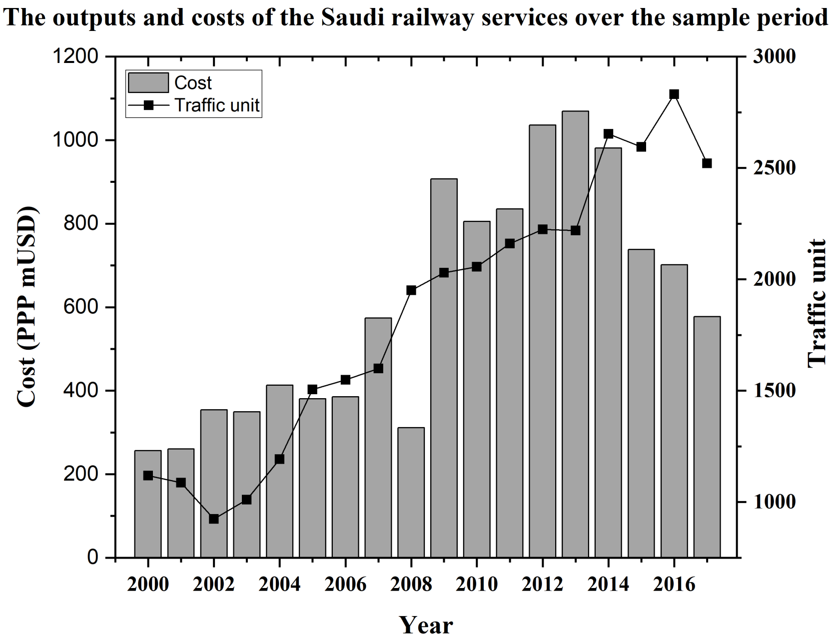


FIGURE 6.2: Average technical efficiency levels over the sample period

In this assessment, the cost efficiency of the Saudi railway system has been estimated via the DEA and SFA methods. First, the system scored a value of 1 and 0.7585 obtained by the DEA estimation under the VRS and CRS assumptions, respectively. These cost efficiency scores are seen as pure efficiency due to the nature of the DEA estimation. The difference between the DEA scores indicates a scale efficiency of 0.7585. Second, the Saudi railway system has a value of 0.5005 as cost efficiency estimated via the SFA method.

To discuss the policy recommendations for the Saudi railway system based on the cost assessment, the changes in the Saudi system have been discussed in detail in Section 3.2 and briefly mentioned in Section 5.7. For the vertical structure, the results show that the vertical separation within a holding company tends to reduce the cost inefficiency levels, particularly at low traffic density. Table 6.8 contains the average values of traffic unit density measured as million traffic unit-km per network-line length in km for the railway systems[7]. According to this table, the Saudi railway system has an average value of traffic unit density of 1.452, which is below the average traffic density of the sample (3.516). This suggests that the vertical separation between IMs and RUs is recommended for the Saudi railway system.

For the horizontal structure, the results show that the regional division of rail services increases the cost inefficiency levels. This would suggest that the merger between SRO and SAR is more valuable reform to increase the cost efficiency. On the other hand, the separation between passenger and freight rail services is not supported by the cost assessment because the results are not statistically significant, although the literature

[7]These average values are different from those in Table 5.6 due to the difference in the sample size.

Table 6.8: Average values of traffic unit density for the railway systems.

Country	No. obser.	Traffic density	Ratio
Algeria	10	0.629	0.179
Austria	12	4.923	1.400
Belgium	9	4.749	1.351
Bulgaria	18	1.361	0.387
Czech Rep.	15	2.190	0.623
Finland	18	2.300	0.654
France	9	3.978	1.131
Germany	18	4.602	1.309
Greece	7	0.951	0.270
Ireland	5	1.058	0.301
Italy	15	3.736	1.062
Morocco	13	4.317	1.228
Pakistan	18	3.304	0.940
Poland	15	2.893	0.823
Portugal	9	2.077	0.591
Saudi Arabia	18	1.452	0.413
Slovakia	17	3.036	0.863
Slovenia	18	3.370	0.958
South Korea	17	10.758	3.059
Spain	13	2.285	0.650
Switzerland	17	7.855	2.234
Turkey	17	1.608	0.457
Average		3.516	1

suggested the market function separation (e.g., Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018)). Having said that, market function separation is recommended for the Saudi railway system based on the findings of the literature.

For the competition forms, the assessment found that each competition form has a different impact on cost inefficiency. Moreover, the freight market deregulation appears to increase the cost inefficiency levels, while the competitive tendering for passenger rail services reduces these levels. Thus, the competitive tendering system is the only policy that can be recommended for the Saudi railway system to increase cost efficiency. This finding means that the horizontal separation between passengers and freight rail services should be applied alongside with the competitive tendering system. Moreover, TOCs can compete each other for passenger rail services via competitive tendering system, while one TOC will be responsible for freight rail services without any form of competition.

Lastly, the results found that combining rail policies at any level will increase the cost inefficiency levels. Although it has been recommended for the Saudi railway system to be reformed with vertical separation combined with market function separation and competitive tendering for passenger rail services, this research does not recommend these three policies to be introduced in one package.

6.8 Conclusion

This research conducted a cost assessment to evaluate the impact of railway reform on cost efficiency. To estimate the cost efficiency, this research employed DEA and SFA as non-parametric and parametric approaches, respectively. The estimation results indicate that, on average, the DEA produces higher cost efficiency scores than the SFA estimation. The reason for this is that the DEA has one or more railway systems that have cost efficiency scores of one because they are identified as the frontier for the other railway systems. Despite this difference, the efficiency rankings of railway systems seemed acceptable based on the results of Spearman's rank correlation.

After the cost efficiency estimation, this research evaluated the determinants of cost inefficiency levels. The results indicated that the vertical separation between IMs and RUs tends to reduce the cost inefficiency levels. However, when the vertical separation is assessed with respect to traffic density, the results suggest that the vertical separation with higher traffic density tends to increase the costs. This could be explained given that dense railway systems need more alignment between IMs and RUs. At the horizontal structure, the regional separation of rail services appeared to increase the cost inefficiency levels, while the results of this research do not support the separation between passenger and freight rail services, as the results are not statistically significant. For the competition forms, the results showed that the on-track and off-track competition have different impacts on the cost inefficiency. Freight market deregulation tends to increase the cost inefficiency levels while competitive tendering reduces these levels. Finally, the results indicated that combining rail policies at any level increases the cost inefficiency levels.

Chapter 7

Welfare Assessment

7.1 Introduction

This chapter discusses the application of the welfare assessment. Whereas the previous two assessments sought to examine the impact of railway reform on performance from the producer perspective (i.e., railway operators), this research seeks to select the optimum organisation for society as a whole by employing the concept of welfare economics. Through this assessment, this research involves the forms of railway reform to capture its impact on welfare and its components (i.e., producer and consumer surplus).

7.2 Literature Review

Several studies have been published in order to examine the effects of railway reform on welfare. These studies used a range of estimation frameworks and empirical data. Ellig (2002), for example, examined the impact of the deregulation of the US railroads. The author found that the deregulation contributed to an improvement in productivity and economic welfare. The deregulation of the US railroads contributed to a reduction in rail fares. In fact, the deregulation was aimed at reducing the cost of the US railroads, which ultimately led to a reduction in rail fares. The author also highlighted that the majority of railroad productivity gains were the result of excess capital and labour being eliminated.

Boardman et al. (2009) aimed to examine the impact of railway reform on the Canadian railway system (CN). The authors found that there is a financial improvement in the CN performance gained from the privatisation. They indicated no change in consumer demand since prices and volumes were constant in comparison to the counterfactual measure. In addition, they found that the impact of privatisation on labour was insignificant.

Preston and Robins (2013) investigated the financial implications of privatising the British railway system from 1997 onwards. They found that the railway reform has the advantage of lowering the rail fare and increasing train services compared to the counterfactual, which in turn increases demand and raises the consumer surplus and revenue. In contrast, there is an increase in the total cost, which leads to losses in welfare.

Preston (2008, 2018) and Preston and Bickel (2020) assessed the franchising phases of the British passenger rail services. They found that there are losses in welfare due to the increase in the total costs of the railway industry, but these may be attributed to infrastructure cost increases and franchising *per se* may be welfare positive.

7.3 Methodology

This research develops a framework based on welfare economics, which is built on the work of others (e.g., Boardman et al. (2009), Preston and Robins (2013) and Preston and Bickel (2020)). This framework aims to calculate the changes in welfare with respect to the changes in railway systems due to railway reforms. Recall Equation 4.55 with slight changes. Based on the concept of producer surplus that stated in Section 4.7, the producer surplus can be calculated as the difference between revenue and costs of rail services. With regards to externalities, this research assumes no material externalities due to the lack of comparative data at an international level. Thus, the welfare (W) for a railway system is calculated as follows:

$$W_t = SC_t + R_t - C_t \tag{7.1}$$

where CS is the consumer surplus; R is the revenue generated by rail services; C is the cost of the rail services; and the subscript t refers to time.

To calculate the consumer surplus, a demand model has been developed. Based on economic theory, the demand model is specified as a function of income and price. In this framework, the demand model consists of control and test variables. The control variables are listed and described as follows:

- Revenue per traffic unit km ($RTUKM$).
  This variable represents the revenue per traffic unit-km. The price is calculated by dividing the total revenue generated from passenger and freight rail services each year by the sum of passenger-km and tonne-km. In addition, this price is

expressed in PPP USD. This variable is assumed to cover the cost of passenger and freight travelled by rail services. Thus, this variable is used as a proxy for the price of passenger and freight rail services. This assumption might not be accurate in representing the price that aims to cover the cost of passenger and freight travelled by rail services for two reasons. First, this assumption does not differentiate between the different forms of tickets sold to the passengers (e.g., first- and second-class tickets, long and short journeys). Second, this assumption implies that the cost of rail transport is the same for both passengers and freight, where freight services tend to consume fewer resources than passenger services. Having said that, this assumption has been made due to data availability.

- GDP per capita (GDP).
  This variable represents the GDP per capita for each country. This variable is measured in current international dollars converted by the purchasing power parity (PPP) conversion factor. This variable is included in the demand model to represent the purchasing power of the population.

- Passenger rail services share ($PKMS$).
  This variable describes the share of passenger rail services in the total rail services.

- Network-line length in km (N).
  This variable defines the length of railway systems. This variable shows how the demand changes with respect to the length of the railway network. This variable is also used as a proxy for dummy variables for each railway system because when railway-specific dummy variables were added to the demand model, there was a high correlation between these variables and the network-line length.

The test variables, on the other hand, are dummy variables representing rail policies to capture the impact of railway reform on demand. These policies are listed as follows:

- Organisational separation (D_{OS}) represents that the Infrastructure Manager (IM) and the Railway Undertaking (RU) are organised with a separation of subsidies and decision-making producers but within a holding company.

- Institutional separation (D_{IS}) represents that the IM and RU are managed and operated by separate entities.

- Market function separation (D_{MFS}) represents that passenger and freight rail services are separated and provided by different railway companies.

- Regional separation (D_{RS}) represents dividing the rail services on a regional basis.

- Off-track competition (D_{OFTC}), which represents competitive tendering for passenger rail services.
- On-track competition (D_{ONTC}), which represents free entry for freight rail services.

Thus, the demand model is written as follows

$$\begin{aligned}\ln(TUKM_{it}) = {} & \alpha + \beta_1 RTUKM_{it} + \beta_2 GDP_{it} + \beta_3 PKMS_{it} + \beta_4 N_{it} \\ & + \sum \gamma_n D_n + \varepsilon_{it}\end{aligned} \tag{7.2}$$

where $TUKM$ is traffic unit km measured as the sum of pass-km and tonne-km; $RTUKM$ is revenue per $TUKM$ measured as revenue divided by $TUKM$; GDP is Gross Domestic Product per capita; $PKMS$ is the passenger shares of total rail services calculated as pass-km divided by $TUKM$; N is the network-line length in km; D represents the dummy variables of rail policies; subscript i refers to rail systems; subscript t refers to time-trend; and α, β and γ are model parameters to be estimated. The estimated demand model will then be used to forecast the traffic unit km, which requires the model variables to be forecasted.

Given the negative exponential demand function specification in Equation (7.2), once the demand model has been estimated, the consumer surplus can be calculated by direct integration as follows

$$CS_{it} = \int_{RTUKM}^{max} TUKM_{it} \, d\, RTUKM = -\frac{1}{\beta_1} TUKM_{it} \tag{7.3}$$

where β_1 is the coefficient of revenue per $TUKM$. The negative sign is included because the expected sign of the price (or revenue per $TUKM$) parameter value is negative. This is because when the price decreases (increases), the demand is expected to increase (decrease). Thus, the consumer surplus cannot be negative, and this is achieved by transposing the upper and lower limits of integration.

On the other hand, the producer surplus is calculated as revenue minus cost. The revenue calculations depend on the demand models. Once the $TUKM$ has been obtained from different demand models, the revenue then can be calculated as

$$R = TUKM \times RTUKM \tag{7.4}$$

To calculate the cost, this research develops a cost model that consists of control and test variables. The control variables are passenger and freight outputs, labour and energy nd material prices and network-line length. The passenger and freight outputs are measured as pass-km and tonne-km, respectively. The labour and energy and material prices are the input prices. The test variables are similar to the variables that are included in the demand model.

$$
\begin{aligned}
\ln(C_{it}) = {} & \alpha_0 + \alpha_1 \ln(PKM_{it}) + \alpha_2 \ln(TKM_{it}) + \frac{1}{2}\alpha_{11} \ln(PKM_{it})^2 \\
& + \alpha_{12} \ln(PKM_{it}) \ln(TKM_{it}) + \frac{1}{2}\alpha_{22} \ln(TKM_it)^2 + \beta_1 \ln(w_{Lit}) \\
& + \beta_2 ln(w_{Mit}) + \frac{1}{2}\beta_{11} \ln(w_{Lit})^2 + \beta_{12} \ln(w_{Lit}) \ln(w_{Mit}) \\
& + \frac{1}{2}\beta_{22} ln(w_{Mit})^2 + \vartheta_N \ln(N_{it}) + \frac{1}{2}\vartheta_{NN} \ln(N_{it})^2 + \theta_T T + \frac{1}{2}\theta_{TT} T^2 \\
& + \delta_{11} \ln(PKM_{it}) \ln(w_{Lit}) + \delta_{12} \ln(PKM_{it}) \ln(w_{Mit}) \\
& + \delta_{21} \ln(TKM_{it}) \ln(w_{Lit}) + \delta_{22} \ln(TKM_{it}) \ln(w_{Mit}) \\
& + \xi_{1N} \ln(PKM_{it}) \ln(N_{it}) + \xi_{2N} \ln(TKM_{it}) \ln(N_{it}) \\
& + \psi_{1N} \ln(w_{Lit}) \ln(N_{it}) + \psi_{2N} \ln(w_{Mit}) \ln(N_{it}) + \varphi_{1T} \ln(PKM_{it}) T \\
& + \varphi_{2T} \ln(TKM_{it}) T + \phi_{1T} \ln(w_{Lit}) T + \phi_{2T} \ln(w_{Mit}) T + \zeta_{NT} \ln(N_{it}) T \\
& + \sum_n \gamma_n D_n + \varepsilon_{it}
\end{aligned}
\tag{7.5}
$$

where C is costs; PKM is passenger output measured as pass-km; y_2 is ffreight output measured as tonne-km; w_L is labour price; w_M is material and energy price; N is network-line length in km; t is time-trend; D is dummy variables for rail policies as stated above; and $\alpha, \beta, \vartheta, \theta, \delta, \xi, \psi, \varphi, \phi, \zeta, \gamma$ are model parameters to be estimated. By estimating a series of cost models, one without dummy variables and the other with them, the effect of rail policies on railway costs will be captured. It should be noted that the cost model developed in the cost assessment can be used in the welfare assessment, but a new cost model has been developed due to the availability of data. This is because of the fact that the sample size for the welfare assessment is smaller than the sample size for the cost assessment, and the sample data will also be used for the demand model, which can deliver a comprehensive estimation of welfare.

To measure the change in welfare due to railway reform, this research attempts to forecast welfare and its components in two scenarios. In the first scenario, this research attempts to forecast the measurements of welfare and its components with respect to the change in railway systems due to railway reform. To determine the form of railway reform, only rail policies that have a significant impact on demand and cost models will be selected. Then, the welfare, consumer surplus, revenue and cost will be forecasted with the effect of railway reform. In contrast, the second scenario is forecasting the welfare and its components without introducing any form of railway reform, which will be named the counterfactual scenario. This means that the demand and cost models will be estimated without dummy variables of rail policies. Then, the welfare, consumer surplus, revenue and cost will be forecasted. After calculating the two scenarios, this research determines the difference between the welfare and its components as the change in welfare due to railway reform.

7.4 Data

In this research, the data sample covers 22 railway systems from Asia, Europe and North Africa from 2001 to 2017. This sample has been selected based on the criteria that are mentioned in Section 3.3. The data is collected from the International Union of Railways (UIC) and completed from the annual reports of each railway company and statistical reports of each country. The data sample consists of rail service outputs, revenue generated, costs, rail infrastructure, and input prices. PKM and TKM are rail service outputs measured as pass-km and tonne-km, respectively. The revenue (R) represents the revenue that is generated from rail services. The cost (C) represents the operating costs of rail services. The input prices are the labour prices (w_L) and material and energy prices (w_M). The labour price is calculated as the total labour cost divided by the number of employees. The energy and material price is obtained as the total energy and material cost divided by the total number of rolling stocks, which includes locomotives, coaches, and wagons. The network-line length (N) in km represents the rail infrastructure.

7.5 Results

7.5.1 Demand Model Estimation

This section discusses the results of the welfare assessment. The calculations of the welfare and its components for the factual and counterfactual scenarios are built on the estimation of the demand and cost model. In these two models, the dummy variables for rail policies are introduced to determine the effect of railway reform on welfare. Thus, it is essential to determine the demand and cost models that will represent the factual and counterfactual scenarios.

To achieve this, a series of demand models have been estimated at different levels by introducing the dummy variables of rail policies gradually. This procedure could help to determine which rail policies have a statistically significant impact on demand. To implement this procedure, this research has developed 13 models based on introducing the dummy variables gradually at two levels (i.e., individual and combined levels), which are specified as follows:

- **Model (1)**: no dummy variables are included.

TABLE 7.1: Average values of demand and cost model variables.

Country	PKM^a	TKM^a	$R^{a,b}$	$C^{a,b}$	N	w_1^b	w_2^b
Algeria	1,137	1,169	134.63	443.64	3,975	30,218.22	8,283.33
Austria	9,713	16,629	3,801.41	2,656.41	5,268	9,053.14	21,233.79
Belgium	9,269	7,480	1,869.83	3,544.88	3,500	4,079.25	1,940.82
Bulgaria	2,396	4,318	500.51	603.64	4,192	4,569.56	30,419.99
Czech Rep.	6,802	13,665	2,210.08	2,489.60	9,477	9,505.18	37,880.99
Finland	3,799	9,774	774.44	706.39	5,893	38,127.61	34,901.53
France	80,856	36,749	4,663.86	5,099.36	9,575	8,425.39	93,592.73
Germany	76,255	83,759	8,662.65	36,375.40	33,940	46,204.25	192,665.65
Greece	1,779	581	123.92	463.38	2,480	36,301.87	42,065.02
Ireland	1,621	408	188.46	347.31	1,919	36,094.60	70,647.10
Italy	44,429	17,859	6,825.24	7,996.93	16,345	43,873.73	77,231.65
Morocco	3,506	5,237	694.17	423.6	1,984	20,728.35	37,762.52
Pakistan	21,872	4,006	899.48	965.95	7,791	2,152.98	40,291.37
Poland	16,426	39,877	3,972.85	4,617.83	19,618	18,456.96	31,741.52
Portugal	3,551	2,362	370.45	461.49	2,844	19,396.33	75,190.97
Saudi Arab.	344	1,544	174.83	628.08	1,266	70,767.43	214,349.61
Slovakia	2,439	8,223	888.92	1,095.79	3,633	16,986.49	44,455.76
Slovenia	761	3,447	332.66	546.98	1,220	35,235.11	67,191.67
S. Korea	27,047	10,045	2,870.99	3,680.46	3,530	65,241.00	12,471.71
Spain	20,493	9,868	2,302.37	2,606.41	3,325	2,773.60	95,684.43
Switzerland	15,731	8,988	2,851.80	3,439.89	3,124	9,423.78	93,674.38
Turkey	5,058	9,765	640	1,955.63	9,259	1,873.39	54,059.15
Average	16,637	14,398	4,112.50	4,354.16	8,620	36,200.26	79,172.15

[a] The variables are expressed in millions.
[b] The variables are expressed in PPP USD.

- **Model (2)**: the dummy variables represent the vertical separation, whether organisational or institutional separation.
- **Model (3)**: the dummy variable represents the market function separation.
- **Model (4)**: the dummy variable represents the regional separation.
- **Model (5)**: the dummy variable represents the off-track competition.
- **Model (6)**: the dummy variable represents the on-track competition.
- **Model (7)**: the dummy variable represents the vertical separation without specifying the degree of separation.
- **Model (8)**: the dummy variable represents the vertical separation combined with horizontal separation.
- **Model (9)**: the dummy variable represents the vertical separation combined with the competition introduction.

TABLE 7.2: Estimation of demand models

Variable	**Parameter**	**Model (1)**	**Model (2)**
$RTUKM$	β_1	-3.030*	-3.917**
		-1.218	-1.192
GDP	β_2	1.36e-05***	9.41e-06**
		-2.80E-06	-2.86E-06
$\%PKM$	β_3	1.023***	1.266***
		-0.201	-0.202
N	β_4	1.09e-04***	1.02E-04***
		-5.95E-06	-5.96E-06
D_{ONTC}	γ_1		0.464***
			-0.104
Constant	α_0	8.206***	8.123***
		-0.128	-0.125
N		263	263
Adj. R^2		0.692	0.714
LL		-269.609	-259.746

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.
Standard errors in parentheses.

- **Model (10)**: the dummy variable represents the vertical separation combined with horizontal separation and competition introduction.
- **Model (11)**: the dummy variables represent all individual rail policies.
- **Model (12)**: the dummy variables represent all individual rail policies, where the degree of vertical separation is not specified.
- **Model (13)**: the dummy variables represent the level of railway reforms.

These demand models have been estimated by the OLS method. The goal of this estimation is to obtain a demand model that has all statistically significant variables, including rail policy variables.

The estimation of the demand model (1) is reported in Table 7.2. This model has an adjusted R-squared of 0.692, which indicates that the demand model explains over 69.2% of the variation in the data. The coefficient of revenue per $TUKM$ (β_1) is statistically significant at a 95% confidence level and has a negative sign. To determine the price elasticity of the Saudi railway system, the mean of $RTUKM$ is multiplied by the coefficient (β_1). The price elasticity at the mean is found to be -0.297. This is relatively low but may reflect a high degree of captivity, particularly for bulk freight, and a non-profit maximising pricing policy.

The coefficient of GDP per capita (β_2) is statistically significant at a 99.9% confidence level and has a positive sign. This impact is expected because countries with higher income levels could lead to more utilisation of rail services. The coefficient of the passenger share of rail services (β_3) is statistically significant at a 99.9% confidence level and has a positive sign. This indicates the greater the passenger share of rail services the more demand for rail services in terms of traffic units, all other things being equal. This parameter value is not easy to explain. Around the sample mean, the result suggests that a railway system with 100% passengers will carry 3.55 times more $TUKM$ (exp(1.266)=3.55) than one with 100% freight. This might suggest that passenger-km are easier to carry than tonnes, with an equivalence of 3.54 passengers to one tonne. The coefficient of network-length (β_4) is statistically significant at a 99.9% confidence level and has a positive sign. The network size has an impact on the demand. This can be explained by the fact that better connectivity between substantial locations (e.g., business and industrial areas) could lead to more transport of large volumes of passengers and freight.

To determine which rail policy has a significant impact on demand, there is only one model that has statistically significant parameter values for all variables. This model is reported in Table 7.2 as Model (2). In this model, the impact of on-track competition on demand is statistically significant at a 99.9% confidence level and has a positive sign. This indicates that railway systems with free entry for freight services have more outputs of rail services, which is equivalent to 59% (Exp(0.464) = 1.59).

7.5.2 Cost Model Estimation

Besides the estimation of the demand model, the cost model has been estimated with the same procedure. First, the cost model has been estimated without any dummy variables, and the results are reported as Model (1) in Table 7.3. The adjusted R^2 for this model has a value of 0.949, which indicates that the estimated model seems to be a proper fit for the data.

To capture the impact of rail policies on rail costs, a series of cost models have been estimated by introducing the dummy variable gradually. The estimation process concludes that the dummy variable of the institutional separation (D_{IS}) is the only variable that is statistically significant. The estimated cost model is reported in Table 7.3 as Model (2). This model has a value of 0.963 for the adjusted R^2. This indicates that the estimated model seems to be a proper fit for the data. In other words, separating the infrastructure manager from train operations can lead to cost reduction. This reduction has been found to be 14% (Exp(-0.147)=0.863). This finding is in line with the results of the technical and cost assessment, where the vertical separation has a positive impact on the efficiency of railway performance, as discussed in Section 5.6 and 6.6.

TABLE 7.3: Estimation of cost models.

Para.	**Model (1)**	**Model (2)**	**Para.**	**Model (1)**	**Model (2)**
α_1	1.076*	0.219	δ_{12}	-0.171***	-0.160***
	(0.520)	(0.464)		(0.049)	(0.043)
α_2	0.572	0.663	δ_{21}	-0.176	-0.140
	(0.432)	(0.371)			
α_{11}	0.016	0.050	δ_{22}	0.176***	0.140***
	(0.082)	(0.072)		(0.043)	(0.037)
α_{12}	-0.066	0.026	ξ_{1N}	-0.063	-0.105
	(0.093)	(0.081)		(0.079)	(0.076)
α_{22}	0.430***	0.355***	ξ_{2N}	-0.389**	-0.409**
	(0.064)	(0.056)		(0.148)	(0.127)
$\beta 1$	0.613	2.214	ψ_{1N}	0.020	-0.220
$\beta 2$	0.387	-1.214*	ψ_{2N}	-0.020	0.220*
	(0.615)	(0.556)		(0.098)	(0.088)
$\beta 11$	0.479	0.316	φ_{1T}	0.017***	0.013**
				(0.005)	(0.004)
$\beta 12$	-0.479	-0.316	φ_{2T}	-0.008	-0.016*
				0.017***	(0.006)
$\beta 22$	0.479***	0.316***	ϕ_{1T}	-0.003	-0.002
	(0.053)	(0.050)			
ϑ_N	-0.657	3.873**	ϕ_{2T}	0.003	0.002
	(1.194)	(1.165)		(0.006)	(0.005)
ϑ_{NN}	0.601*	0.126	ζ_{NT}	-0.006	0.004
	(0.251)	(0.235)		(0.010)	(0.008)
θ_T	-0.048	-0.050	γ_{IS}		-0.147**
	(0.054)	(0.046)			(0.055)
θ_{TT}	0.001	0.002	α_0	1.595	-13.813***
	(0.002)	(0.002)		(3.167)	(3.263)
δ_{11}	-0.171	0.160			
N	263	263			
Adj. R2	0.949	0.963			
LL	-50.477	-9.237			

*, **, *** indicates significance at the 0.05, 0.01, and 0.001 level, respectively.
Standard errors in parentheses.
Underlined coefficients are obtained by homogeneity restrictions.

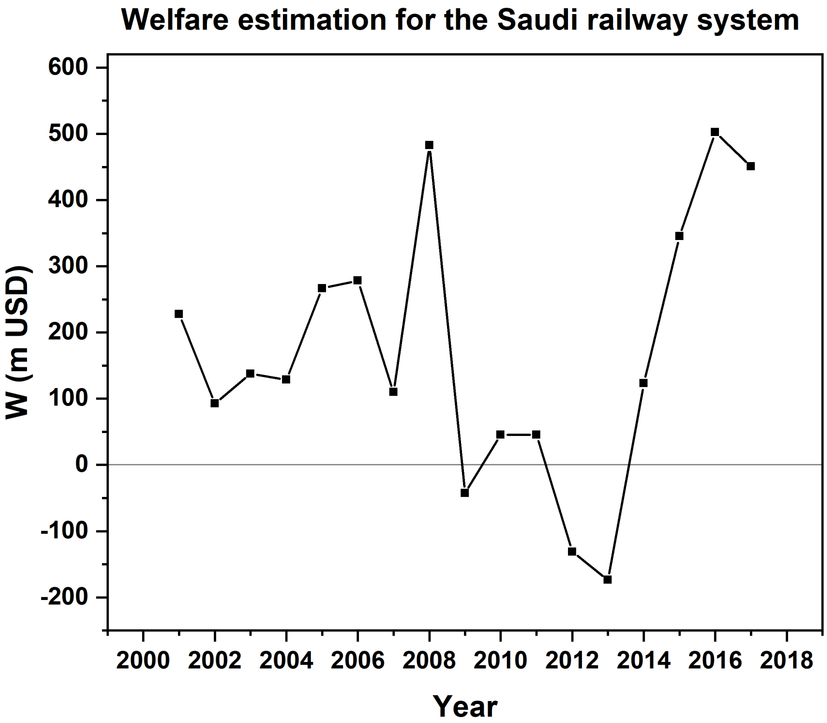


FIGURE 7.1: The welfare estimation of the Saudi railway system

7.5.3 Welfare Calculations

Before presenting the results of the forecasted welfare, Figure 7.1 displays the estimation of the welfare of the Saudi railway system between 2001 and 2017. This estimation represents the actual measure of the railway system. According to this figure, there is a large variation in welfare. This variation can be explained by the components of welfare, as shown in Figure 7.2. First, the consumer surplus has a quite steady increase over time. As the consumer surplus is calculated based on the traffic unit and price, which is stated in Equation (7.3), the Saudi railway system has a growth in the traffic unit over time, where the freight output is the main contributor to this increase.

Second, due to the increase in demand for rail services, there is an increase in revenue generated from both passenger and freight rail services. Lastly, there is a large variation in the cost of the Saudi rail services. Although there is an increase in the inputs (e.g., employees and rolling stocks) and outputs (e.g., passenger and freight services), the cost does not line with this increase. Figure 7.2 shows a fluctuation in the cost of rail services. There is a slight increase in the cost between 2001 and 2008, although the cost has dropped in some years, However, there is a sharp increase in the cost between 2009 and 2014, then followed by a continuous drop from 2015 to 2017.

To forecast the welfare and its components from 2018 onwards for the Saudi railway system, the variables of the demand and cost models have been forecasted. Table 7.4

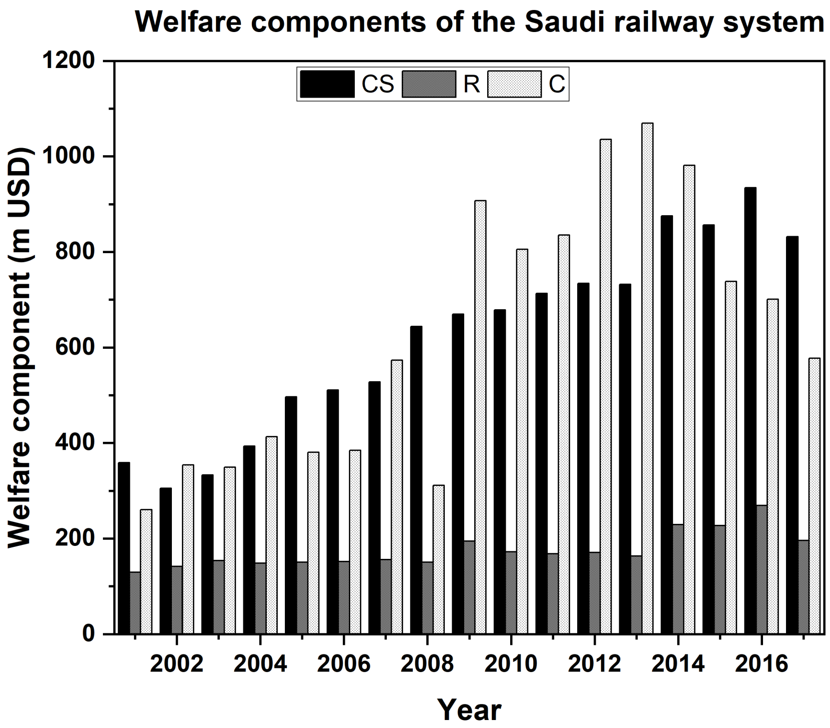


FIGURE 7.2: The estimation of welfare components for the Saudi railway system.

reports the forecasted variables of the demand model from 2018 to 2030. The forecasted measure of GDP per capita is calculated via a fitted trend-line based on moving average (5 years centred average) data rather than the actual data. Second, to obtain the forecasted measure of $RTUKM$, the passenger and freight outputs and their revenue have been forecasted via a fitted trend-line. Then, the forecasted revenue is divided by the $TUKM$, which is the sum of the forecasted passenger and freight outputs. Lastly, the forecasted passenger share of rail services is calculated as the percentage of the forecasted passenger to the forecasted $TUKM$. Appendix A provides the figures that show the actual and forecasted measures of demand variables. For the network-line length, it is assumed to be constant, although the Saudi government has a plan to expand its railway network, but there is no timeline for these projects.

To calculate the consumer surplus for each scenario, Demand Model (2) that is reported in Table 7.2 is applied to calculate the factual measure of the traffic unit. This demand model is written as follows:

$$\begin{aligned} \ln(TUKM_t) = {} & 8.123 - 3.917\, RTUKM_t^F + 9.41 \times 10^{-06}\, GDP_t^F + 1.266\, PKMS_t \\ & + 1.02 \times 10^{-04}\, N_t + 0.464\, D_{ONTC} + \varepsilon_t \end{aligned} \tag{7.6}$$

This research assumes that the freight rail market will be open to free entrants from 2023 onwards. This means that the policy variable (D_{ONTC}) in the demand model will

TABLE 7.4: The forecasted measurements of the Saudi railway system for the demand model.

Year	$RTUKM^F$ **(USD)**	GDP^F**(USD)**	$\%PKM$	N **(km)**
2017	0.078	20,803.75	0.177	1,710
2018	0.078	26,301.22	0.144	1,710
2019	0.077	27,269.78	0.142	1,710
2020	0.076	28,238.34	0.139	1,710
2021	0.075	29,206.90	0.137	1,710
2022	0.074	30,175.47	0.135	1,710
2023	0.073	31,144.03	0.133	1,710
2024	0.072	32,112.59	0.131	1,710
2025	0.072	33,081.15	0.13	1,710
2026	0.071	34,049.71	0.128	1,710
2027	0.071	35,018.28	0.127	1,710
2028	0.07	35,986.84	0.125	1,710
2029	0.069	36,955.40	0.124	1,710
2030	0.069	37,923.96	0.123	1,710

be set to 1 to estimate the factual scenario from 2023 onwards. On the other hand, this variable will be set to 0 to estimate the counterfactual measure of traffic units for the same period. Table (7.5) reports the factual ($TUKM^F$) and counterfactual ($TUKM^C$) of traffic units. The results show that there is an increase in the traffic unit due to the on-track competition policy. It should be noted that both measures of the traffic unit show an increase. This increase can be explained by the extrapolated trend-line of the moving average for the demand model variables.

After forecasting the traffic units, Equation (7.3) can be applied to calculate the consumer surplus for both scenarios. Table 7.5 contains the forecasted consumer surplus for both scenarios. The results indicate that both measures of the consumer surplus show a rapid increase based on the demand model forecasting. Again, this can be explained by the effect of the extrapolated trend-line of the moving average for the demand model variables.

To measure the producer surplus, first, the revenue generated for the two demand models can be calculated as the product of $TUKM$ and $RTUKM$, which is written as follows

$$TUKM_t^j \times RTUKM_t^F = R_t^j \tag{7.7}$$

where j refers to the forecasted measures of $TUKM$ as listed in Table 7.5; and $RTUKM_t^F$ is the forecasted $RTUKM$ as listed in Table 7.4. For the factual measure of revenue (R^F), the traffic unit that is forecasted via demand model (2) is multiplied by the forecasted measure of $RTUKM$. On the other hand, the counterfactual

Table 7.5: The forecasted measurements of $TUKM$ and consumer surplus of the Saudi railway system.

Year	$TUKM^{Fa}$	$TUKM^{Ca}$	$CS^{Fa,b}$ **(USD)**	$CS^{Ca,b}$ **(USD)**
2018	2,527		645.18	
2019	2,670		681.59	
2020	2,820		720.02	
2021	2,979		760.58	
2022	3,147		803.39	
2023	5,286	3,324	1,349.60	848.58
2024	5,584	3,511	1,425.45	896.27
2025	5,897	3,708	1,505.53	946.62
2026	6,228	3,916	1,590.06	999.77
2027	6,578	4,136	1,679.28	1,055.87
2028	6,947	4,368	1,773.48	1,115.10
2029	7,336	4,613	1,872.91	1,177.62
2030	7,747	4,871	1,977.88	1,243.62

[a] The variables are expressed in millions.
[b] The variables are expressed in PPP USD.

Table 7.6: The calculation of revenue of each forecasted scenario.

Year	R^{Fa} **(USD)**	R^{Ca} **(USD)**
2018	197.31	
2019	205.70	
2020	214.61	
2021	224.05	
2022	234.06	
2023	389.10	244.65
2024	406.93	255.86
2025	425.78	267.71
2026	445.70	280.24
2027	466.75	293.47
2028	488.98	307.45
2029	512.45	322.21
2030	537.22	337.79

[a] The variables are expressed in millions and in PPP USD.

measure of revenue (R^C) is calculated by multiplying the counterfactual measure of $TUKM$ by the forecasted $RTUKM$. The results of these two measures of revenue are reported in Table 7.6. The results show an increase in the revenue for both scenarios. Again, this increase is attributed to the growth in the forecasted $TUKM$ that is produced by the two demand models.

On the other hand, the cost models that are reported in Table 7.3 are used to calculate the factual and counterfactual measures of cost by using the forecasted variables of the

Table 7.7: The forecasted measurements of the Saudi railway system for the cost model.

Year	PKM^{Fa}	TKM^{Fa}	w_L^{Fb} (USD)	w_M^{Fb} (USD)	N (km)
2018	425	2,517	70,453.96	339,485.81	1,710
2019	434	2,625	70,419.18	353,384.96	1,710
2020	443	2,733	70,384.40	367,284.11	1,710
2021	452	2,841	70,349.62	381,183.26	1,710
2022	461	2,949	70,314.84	395,082.41	1,710
2023	470	3,057	70,280.06	408,981.56	1,710
2024	479	3,165	70,245.28	422,880.71	1,710
2025	488	3,273	70,210.50	436,779.86	1,710
2026	497	3,381	70,175.72	450,679.01	1,710
2027	506	3,489	70,140.94	464,578.16	1,710
2028	515	3,597	70,106.16	478,477.32	1,710
2029	524	3,705	70,071.38	492,376.47	1,710
2030	533	3,813	70,036.60	506,275.62	1,710

[a] The variables are expressed in millions.
[b] The variables are expressed in PPP USD.

cost model, which are reported in Table 7.7. First, rail services (pass-km and tonne-km) are forecasted via a fitted trend-line based on moving average (5 years centred average) data. Second, labour prices and energy and material prices are forecasted via a fitted trend-line. The figures that show the actual and forecasted measure of the cost model variables are provided in Appendix A. Lastly, the network-line length is assumed to be constant.

Cost model (2) that is reported in Table 7.3 is used to forecast the cost from 2018 to 2030. This cost model represents the two scenarios as follows. Again, this research assumes that the vertical separation policy will be introduced in 2023. This means that the dummy variable (D_{IS}) will be set to 1 from 2023 onwards. On the other hand, the same cost model will be used to calculate the counterfactual cost by switching off the dummy variable ($D_{IS} = 0$). The results of both measures of cost are shown in Table 7.8. The results indicate that there is a quite steady increase in cost in both scenarios, but the factual measure of cost is lower than the counterfactual cost. This cost reduction is related to the effect of the vertical separation between the infrastructure manager and train operations.

Lastly, the welfare for both scenarios can be calculated by using Equation (7.1). Table 7.9 reports the calculations of the factual welfare (W^F) and the counterfactual welfare (W^C). The results show there is a quite steady increase in both measures of welfare. In addition, the factual welfare is significantly higher than the counterfactual welfare. This change is related to the rail policies of freight market deregulation and vertical separation.

TABLE 7.8: The calculation of cost of each forecasted scenario

Year	C^{Fa} **(USD)**	C^{Ca} **(USD)**
2018	1,048.80	
2019	1,056.85	
2020	1,064.99	
2021	1,073.32	
2022	1,081.91	
2023	941.9	1,090.85
2024	949.98	1,100.21
2025	958.51	1,110.08
2026	967.53	1,120.53
2027	977.13	1,131.65
2028	987.36	1,143.49
2029	998.29	1,156.15
2030	1,009.98	1,169.70

[a] The variables are expressed in millions and in PPP USD.

TABLE 7.9: The results of welfare and its components for each scenario

Year	CS^{Fa}	R^{Fa}	C^{Fa}	W^{Fa}	CS^{Ca}	R^{Ca}	C^{Ca}	W^{Ca}
2018	645.18	197.31	1,048.80	-206.31				
2019	681.59	205.70	1,056.85	-169.56				
2020	720.02	214.61	1,064.99	-130.37				
2021	760.58	224.05	1,073.32	-88.69				
2022	803.39	234.06	1,081.91	-44.46				
2023	1,349.60	389.10	941.90	796.80	848.58	244.65	1,090.85	2.38
2024	1,425.45	406.93	949.98	882.40	896.27	255.86	1,100.21	51.92
2025	1,505.53	425.78	958.51	972.80	946.62	267.71	1,110.08	104.25
2026	1,590.06	445.70	967.53	1,068.22	999.77	280.24	1,120.53	159.47
2027	1,679.28	466.75	977.13	1,168.90	1,055.87	293.47	1,131.65	217.70
2028	1,773.48	488.98	987.36	1,275.10	1,115.10	307.45	1,143.49	279.06
2029	1,872.91	512.45	998.29	1,387.07	1,177.62	322.21	1,156.15	343.67
2030	1,977.88	537.22	1,009.98	1,505.12	1,243.62	337.79	1,169.70	411.71

[a] The variables are expressed in millions and in PPP USD.

7.6 Discussion

This section evaluates and discusses the effect of railway reform on the welfare analysis. Table 7.10 contains the results of the changes in welfare and its components between the two scenarios. These changes are presented as the difference between the factual and counterfactual scenarios and the percentage of change. The results show there is an increase in the welfare of the Saudi railway system as a result of the implementation of vertical separation and on-track competition policies. This is expected because the on-track competition has a positive impact by increasing the

TABLE 7.10: The change in Welfare and its components.

Year	ΔCS^a	ΔR^a	ΔC^a	ΔW^a
2023	501.02	144.45	-148.95	794.42
2024	529.18	151.07	-150.23	830.47
2025	558.91	158.06	-151.57	868.55
2026	590.29	165.46	-153.00	908.75
2027	623.41	173.27	-154.52	951.20
2028	658.38	181.53	-156.14	996.04
2029	695.29	190.24	-157.87	1,043.40
2030	734.26	199.44	-159.72	1,093.41

[a] The variables are expressed in millions and in PPP USD.

demand for rail services, particularly for freight. In addition, institutional separation has a positive impact by reducing the cost of rail services.

Although the results show a quite steady increase in welfare between the two scenarios, externalities are not included in the welfare calculations, which could affect the welfare calculations. This research assumes the externalities (e.g., congestion costs, environmental damage, and accidents) to be zero. This assumption has been made due to a lack of data. Because of this assumption, the results of the welfare analysis might not capture the full factors that might be used to calculate the welfare. In the two scenarios, the demand model suggests an increase in rail services due to the on-track competition, particularly for freight rail services. This increase in rail services could lead to an increase in the negative externalities, particularly the environmental damage and the congested rail network. However, there is some evidence showing that externalities are comparatively low for long-distance, inter-urban passenger and freight rail traffic. For example, Almujibah (2021, Table 1, Appendix A) found that the external costs are relatively small to the total costs, around 1.5%. This would suggest a relatively small change in the welfare calculations between the two scenarios. In addition, the results of increased rail external costs could be offset by less congestion and environmental damage to the road system. However, a multi-model, which incorporates the different modes of transport, is outside the scope of this research.

7.7 Conclusion

To conclude, this research attempted to conduct a welfare assessment to determine the optimal railway organisation for the Saudi railway system. This assessment assumed several forms of railway reform. Based on the literature, railway reforms have an impact on railway performance. These impacts vary between the forms of railway reform (e.g., vertical separation and competition forms).

Based on the results of the welfare assessment, two rail policies can be recommended. The on-track competition policy for freight rail services is recommended. This is because the results show that this policy has a positive impact on demand. This impact shows an increase in the traffic units, which is the sum of the passenger and freight outputs, which leads to an increase in revenue and consumer surplus in turn. The second policy to recommend is the vertical separation between IMs and RUs. The results show that the cost of rail services can be reduced by separating IMs from RUs. By combining these two recommended policies, the results show there is an increase in the welfare calculations. However, these calculations do not include the side effects of the increase in rail services, which could be the cost of environmental damage and traffic congestion on the railway system, but they may be offset by reductions on the road system.

Chapter 8

Policy Recommendation and Validation

8.1 Introduction

This chapter discusses and validates the policy recommendations for the Saudi railway industry. This research conducted technical, cost, and welfare assessments to assess the impact of railway reform on efficiency. In this chapter, this research discusses the findings of three assessments and outlines the similarities and differences between them. It should be noted that these findings have been obtained from practical work. Thus, this research conducts interviews with experts from the Saudi railway industry. Moreover, the findings of the research assessments will be compared with the results of the expert interviews for validation purposes.

8.2 Policy Recommendations

Through the technical, cost, and welfare assessments, this research project has drawn some recommendations for rail policies for the Saudi railway system. Table 8.1 summarises the findings of the three assessments regarding the impact of the rail policies. In this table, the positive impact means that the policy reduces the technical or cost inefficiency levels or has a significant impact on the demand and/or cost model in the welfare assessment. It should be noted that the significant impacts in the welfare assessment mean demand growth and/or cost reduction. In contrast, the negative impact means that the policy either increases the technical or cost inefficiency levels or has no significant impact on the demand and/or cost model in the welfare assessment.

At the vertical dimension, the findings of the three assessments indicate that the organisational separation policy has been found to reduce the technical and cost

TABLE 8.1: Summary of the impact of the rail policies.

Policy	TA	CA	WA
Organisational separation	Positive	Positive	Negative*
Institutional separation	Positive	Positive*	Positive
Regional separation	Negative	Negative	Negative*
Market function separation	Positive*	Negative*	Negative*
On-track competition	Negative	Negative	Positive
Off-track competition	Negative	Positive	Negative*
COMB	Negative	Negative	Negative*

Notes: TA = Technical Assessment. CA = Cost Assessment. WA = Welfare Assessment.
* indicates not statistically significant.
COMB means any level of combination between rail policies.

inefficiencies, while this policy has no significant effect on the demand and cost model in the welfare assessment. For the institutional separation policy, the three assessments concluded that this policy increases the efficiency levels. It should be noted that the cost assessment found that institutional separation reduces the cost inefficiency levels, but this is not statistically significant. When both vertical separation forms are assessed with respect to traffic density in the technical and cost assessment, the results show that both separation forms increase the technical and cost inefficiency levels. In other words, separating IMs from RUs for highly dense railway systems increases the technical and cost assessments. Therefore, both vertical separation forms can be recommended as the Saudi railway system has low density compared to the sample size.

At the horizontal dimension, this research found that the regional division of rail services increases the cost and technical inefficiency levels. On the other hand, this policy has no significant impact on the demand and cost models in the welfare assessment. Thus, the regional separation policy is not recommended for the Saudi railway system. With regards to the market function separation policy, the results of this research do not support this policy. The technical assessment concluded that this policy reduces the technical inefficiency levels, but this is not statistically significant. In contrast, the cost assessment found that this policy increases the cost inefficiency levels, but this is also not statistically significant. In the welfare assessment, this policy has no significant effect on the demand and cost models. As a result, the findings do not support the separation of passenger and freight rail services, although there is a mixture of findings in the literature. For example, Cantos et al. (2010) concluded that the market separation reduces technical efficiency, while, Cantos et al. (2012), Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018) found a positive impact of this separation on railway performance.

For the competition policies, the findings of three assessments found that the on-track competition policy increases the technical and cost inefficiency levels. In contrast, this policy has a significant impact on the demand model in the welfare assessment. It should be noted that the welfare assessment did not include the externalities in the welfare calculations that could be generated by the demand growth, as discussed in Section 7.6.

With regards to the off-track competition policy, the findings of the three assessments are observed differently. This policy was found to increase the technical inefficiency levels but reduce the cost inefficiency. This policy also has no significant impact on the demand and cost models in the welfare assessment. Thus, the results of this research do not recommend a competitive tendering policy for passenger rail. The main difference between the two assessments is the sample size. The sample size of the technical assessment consists of 434 observations for 29 railway systems from Europe, the Middle East, and Central and East Asia, whereas the sample size of the cost assessment is reduced to 308 observations for 22 railway systems. This indicates that there is a loss in the number of observations and railway systems due to data availability, particularly from Asia (e.g., Azerbaijan, Iran, Japan, Kazakhstan, Turkmenistan, and Uzbekistan). As a result, the findings of the technical assessment capture the effect of franchise, which leads to increasing technical inefficiency levels, whereas the findings of the cost assessment comply more with the findings of Smith et al. (2018)[1], which capture the role of economic regulations. As the sample size of the cost assessment spans until 2017 and includes European railway systems, which are the only systems with competitive tendering, the findings of this assessment suggest that this competition form with strong economic regulations reduces the cost inefficiency levels.

For policy combination, this research has examined different levels of combining rail policies, which are listed in Section 5.6. The findings of the three assessments indicate that any level of combination between rail policies has negative impacts on the technical and cost inefficiency levels and does not have significant impacts on the demand and cost models in the welfare assessment. It should be noted that, in the welfare assessment, two rail policies have been included to calculate the welfare and its components. However, there is a weakness in the welfare calculations because the externalities that could be generated from the increase in the demand for rail services are not included. Despite this weakness, Almujibah (2021) found that the external costs of the Saudi railway system are relatively small compared to the total costs. Therefore, it is difficult to recommend for the Saudi railway industry to combine rail policies.

[1] The sample size of Smith et al. (2018) covers a period between 2002 and 2010, but this research extends the sample period to 2017, which could capture the effect of the latest EU railway packages on passenger market competition.

8.3 Expert Interviews

In this section, the results of the expert interviews are presented. The participant questionnaire consists of two sections. The first section consists of open questions designed to gather information about the participants (for example, years of experience, areas of expertise, and occupations). The second section consists of questions that are built on a Likert scale. The Likert scale was proposed by Likert (1932) as a technique to measure the attitudes of the participants. In this research, the questions are designed to obtain the opinions of the Saudi rail experts to measure the levels of policy recommendations, which are listed as follows:

1. Institutional separation policy;
2. Organisational separation policy;
3. Regional separation policy;
4. Market function separation policy;
5. On-track competition policy for freight rail services;
6. Off-track competition policy; and
7. Open-access competition policy for passenger rail services.

The interviews were conducted via email by sending the participant information sheet, consent form, and participant questionnaire between March 15 and April 15, 2022, and the participant questionnaire is provided in Appendix B. The results of the expert interviews are discussed as follows. First, the discussion starts by presenting the structure and background of the participants, and Table 8.2 reports the results of this section. The sample of this survey consists of seven participants from the Saudi railway industry. Six out of seven participants are from the railway companies, and only one participant is from the PTA. The occupations of the participants are one project manager, four engineers, and two technicians. The participants have years of experience ranging from three to eight years, with two participants having three years of experience. Although this survey aimed to target experts who have at least five years of experience, all participants are included due to low response. The participants have expertise in project management, track maintenance, and facility management.

Second, Figure 8.1 shows the levels of policy recommendations by participants. According to this figure, one out of seven participants disagreed with recommending the institutional separation policy, while three participants strongly agreed with this policy. The remaining three participants are neutral about this recommendation. This

Table 8.2: Details of participant background.

ID	Location	Occupation	Years of experience	Area of expertise
P01	PTA	Acting Project Manager	5	Project management
P02	SAR	Electrical Engineering	3	Facilities management
P03	SAR	Equipment Maintenance Senior Engineer	5	Track and equipment maintenance
P04	SAR	Equipment Specialist	8	Equipment maintenance
P05	SAR	Signals and communications technician	5	Signals and communications
P06	SAR	Facility management (FM) engineer	5	Facility maintenance
P07	SAR	Electrical Engineer	3	Facility maintenance

indicates that Saudi rail experts recommend reforming the railway industry through a vertical separation between the infrastructure manager and train operations.

Besides the institutional separation, the results of the expert interviews regarding the organisational separation policy are as follows. Three out of seven participants agreed to recommend this policy. In fact, two participants strongly agreed to recommend this policy. In contrast, four out of seven participants disagreed with recommending this policy for the Saudi railway industry. This number is split between three and one for strongly disagreed and disagreed, respectively. This means that the rail experts do not recommend vertical separation within a holding company. As a result, the institutional separation policy is more recommended by the Saudi rail experts. This recommendation is in line with Cantos et al. (2010)[2], Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018)[3]

With regards to the horizontal dimension, Figure 8.1 shows the results of the expert interviews for the regional separation policy. The results indicate an equivalent number of participants (two to two) who strongly disagreed and strongly agreed with the regional separation of rail services, while four participants are neutral about recommending this policy. Besides the regional separation, the results of recommending the market function separation policy are shown in Figure 8.1. Only one participant strongly disagreed with recommending this policy. In contrast, six out

[2]The authors did not specify the degree of vertical separation.

[3]Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018) assessed the impact of vertical separation in conjunction with external factors (e.g., train density and economic regulation).

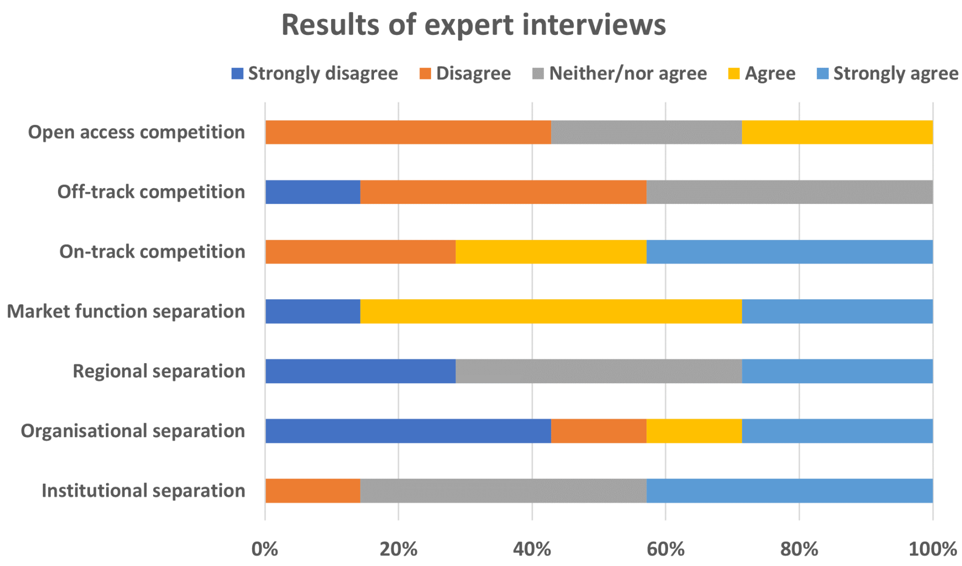


FIGURE 8.1: Results of expert interviews for each rail policy.

of seven participants are in a position to recommend the market function separation. This number is divided between four and two who agreed and strongly agreed with this policy, respectively. Based on these results, it appears that Saudi rail experts recommend separating passenger from freight rail services. This recommendation is in line with Cantos et al. (2012), Mizutani and Uranishi (2013), van de Velde et al. (2012), Mizutani et al. (2015) and Smith et al. (2018).

For the competition form policies, the results of the expert interviews regarding these policies are shown in Figure 8.1. For the on-track competition policy, two out of seven participants disagreed with recommending this policy. In contrast, this policy has been recommended by five participants. This number consists of two and three of the participants who agreed and strongly agreed, respectively. According to these results, the Saudi rail experts recommend market deregulation for free entry for freight rail services, which is in line with Mulder et al. (2005), Asmild et al. (2009), Cantos et al. (2010) and Cantos et al. (2012).

With regards to the off-track competition policy, four participants (one strongly disagreed and three disagreed) out of seven disagreed with recommending this policy, while the remaining three participants are neutral about this recommendation. It appears that none of the participants has recommended this policy. Thus, the franchise system for passenger rail services is not recommended for the Saudi railway industry based on the expert interviews.

The final policy is the open-access competition for passenger rail services. Two out of seven participants agreed to recommend this policy. In contrast, three participants disagreed with recommending this policy. The remaining two participants are neutral about this recommendation. Based on these results, the Saudi rail experts do not recommend market deregulation for free entry for passenger rail services.

Finally, the participants have been asked to recommend rail policies in the form of a combination. These choices are set based on the levels, which are as follows:

1. No rail policies are recommended (L1).
2. Vertical separation only (L2).
3. Horizontal separation only (L3).
4. Competition introduction only (L4).
5. Vertical separation combined with horizontal separation only (L5).
6. Vertical separation combined with competition introduction only (L6).
7. Vertical separation combined with horizontal separation and competition introduction (L7).

Figure 8.2 displays the number of participants for recommending the levels of reforms. According to this figure, completing the three forms of railway reforms is most recommended by the Saudi rail experts (four out of seven). This choice is expected based on the recommendations of individual policies above. In other words, this recommendation might represent institutional separation, market function separation, and on-track competition policies. This result is in line with Cantos et al. (2012). The second recommendation is to reform the railway system with vertical separation combined with competition introduction only. In other words, two out of seven participants recommend reforming the railway system in the form of institutional separation combined with on-track competition. This recommendation is in line with Jensen and Stelling (2007). Lastly, only one participant recommends reforming the railway system by introducing competition for rail services only.

8.4 Discussion and Conclusion

This section provides a comparison between the outcomes of the technical, cost, and welfare assessments and the expert interviews. Overall, there is some consistency between the findings of this research and the expert interviews. For the vertical structure, both this research and expert interviews are in line with recommending the institutional separation policy. For the organisational separation policy, although this research recommends this policy based on the findings of the three assessments, this policy does not fit the Saudi railway system based on the expert interviews. Thus, the

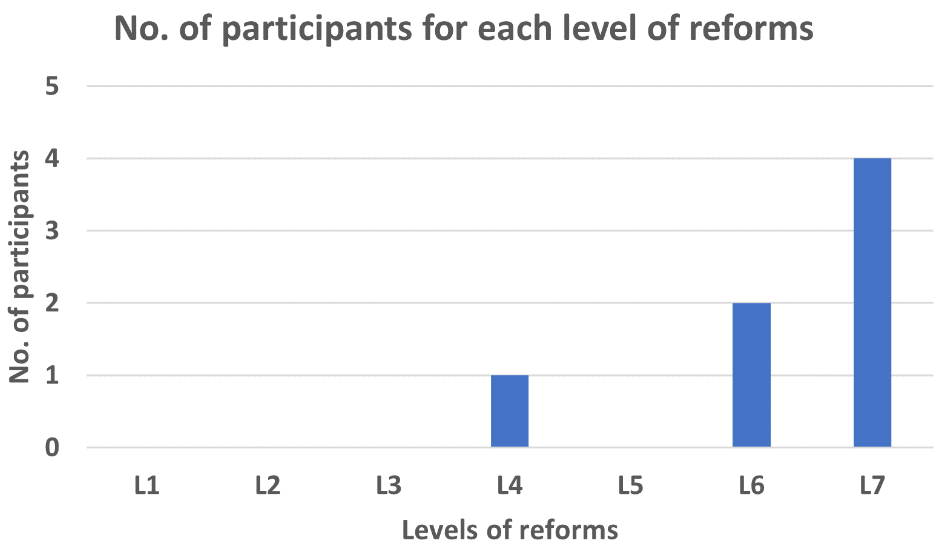


FIGURE 8.2: Results of expert interviews for levels of reforms.

vertical separation between the infrastructure manager and train operations is the preferred policy for the vertical structure.

For the horizontal structure, it is not recommended to separate rail services on a regional basis based on the findings of the three assessments. On the other hand, this policy has an equivalent percentage of recommending this policy based on the expert interviews. This split might be explained by the merger between SAR and SRO. As the expert interviews involve participants from the Saudi railway companies, it appears that participants have different positions on this merger. With regards to the market function separation policy, the three assessments have a mix of findings based on the three assessments, but the impact is not statistically significant. In contrast, the majority of the participants recommend separating passenger from freight rail services. Thus, this research recommends the market function separation policy only for the horizontal structure.

For the competition forms, although this research considers the on-track and off-track competition policies in the three assessments, the expert interviews added the open-access competition for passenger rail services. First, the three assessments have a mix of findings regarding the on-track competition policy. The market deregulation for free entry for freight services has a positive impact based on the welfare assessment alone. In contrast, the majority of rail experts seem to recommend this policy for the Saudi railway industry. Second, the results of the three assessments indicate a mix of outcomes. The competitive tendering policy for passenger services has a positive impact as found by the cost assessment only. In contrast, the majority of rail experts do not recommend this policy for the Saudi railway industry. Last, this research did not include the open-access competition policy in the three assessments. However, based on the expert interviews, it appears that this policy is not recommended by the rail experts. Thus, this research does not recommend the open-access competition policy for passenger services based on the results of expert interviews. The only policy that might be recommended is the on-track competition policy.

For the combined policies, the three assessments and the expert interviews have different outcomes. The assessments of this research concluded that any level of combination between rail policies has a negative impact on railway efficiency. In contrast, the results of the expert interviews indicate that combining the three policies is the more recommended form by the majority of the rail experts. This indicates that the findings of the research assessments are in opposition to the results of the expert interviews, which makes it difficult to draw some recommendations. However, the welfare assessment indicates that the freight market deregulation will result in an increase in the demand for rail services, while the vertical separation will reduce costs. Furthermore, deregulating the freight rail market will lead to the formation of specialist rail freight companies. This requires the Saudi railway system to be separated vertically and horizontally (market function separation). If freight market deregulation is the preferred option, it is not recommended for the Saudi railway system to introduce the three policies in a one-package, based on the results of the technical and cost assessments.

The vertical separation policy does not necessarily mean that the TOC is selected via a competitive tendering system. There are several examples (e.g., Turkey and Slovakia) where there is no competitive tendering combined with vertical separation. This option means that the TOC could be one entity owned by the government for providing rail services. If the Saudi authority has selected to adopt the competitive tendering system, there is one experience with the HHSR project where the government requested an alliance between one Saudi company and foreign companies to enter the competitive tendering.

With regard to yardstick competition, this research did not evaluate this form via the three assessments. However, if this competition form is considered by the Saudi authority for the railway system, the regional division is required, but this research found that the regional separation of rail services reduces the efficiency of the railway system. This indicates that the yardstick competition is not feasible.

Chapter 9

Conclusion and Future Work

9.1 Overview

The railway industry was reformed widely to create a competitive environment to improve performance efficiency. Additionally, the introduction of competition was primarily made possible by a combination of horizontal and vertical railway system reorganisation and railway market privatisation. For the structural reform, the railway organisation was restructured in the form of vertical separation, separating IMs from RUs, and horizontal separation, separating passenger from freight rail services. The vertical separation aims to provide fair access to TOCs to operate on the rail infrastructure through a competition process. The horizontal separation, on the other hand, aims to separate the railway services in terms of profitability. Furthermore, passenger rail services, especially at local and regional levels, are often unprofitable and hence subsidised to provide these services at affordable prices, whereas freight rail services are often profitable and unsubsidised.

However, several countries still operate their railway systems as integrated state-owned companies. In addition, there are some countries, such as Saudi Arabia, that aim to reform the railway industry for development purposes. Although the Saudi government has started to reform its railway industry by following the international trends in railway reform, the railway organisation might be considered underdeveloped compared to other railway systems. Therefore, the essential aim of this research is to deliver an assessment of reforms to optimise railway organisation in the Saudi context.

The effect of the railway reform on performance has been investigated in the literature. The effects of introducing the competition are varied between the competition forms. Competitive tendering for passenger rail services increases technical efficiency and has no effect on cost reduction, except that Smith et al. (2018) discovered that competitive tendering reduces costs, but they considered the role of economic regulation in the analysis. Freight market deregulation, on the other hand, increases technical efficiency but has no statistically significant impact on cost reduction.

Besides the effect of introducing competition, the influence of structural reform on the technical and financial performance of railway systems is varied. The horizontal separation between passenger and freight rail services is more beneficial in terms of cost reduction and technical performance increase. On the other hand, some studies (e.g., Cantos et al. (2010) and Friebel et al. (2010)) show that vertical reform increases technical efficiency, whereas other studies (e.g., Cantos et al. (2012)) find that vertical separation by itself (without introducing competition) has no significant impact on railway performance. For the financial performance, there are various impacts of vertical separation and integration on cost reduction. At high levels of train density, vertical integration is more beneficial in terms of cost reduction compared to vertical separation, while there are opposite impacts at low levels of train density.

9.2 Main Findings

The findings of this research are discussed in two parts. This part is the findings that can be related to the whole sample size, which will be based on the technical and cost assessments. The Saudi railway system will be the main topic of discussion in the second part, which involves the findings of the technical, cost and welfare assessments and the expert interviews.

The results of the technical and cost assessments are summarised as follows. First, the vertical separation between IMs and RUs is the most recommended structural form based on the results of the two assessments, particularly the institutional separation. However, this structural form has a negative impact at high levels of traffic unit density. In other words, the vertical integration between infrastructure managers and train operations is more beneficial for railways with high levels of traffic density. This implies that vertical integration is seen to have the ability to provide close coordination between IMs and RUs, which would simplify the complex technical operations of the rail sector at high levels of traffic density.

Second, based on the results of the two assessments, regional rail service separation and on-track competition for freight rail services are the least recommended rail policies. The results indicate that the regional separation of rail services has a negative impact on railway performance. Similarly, the results suggest that deregulation of the freight market for new entrants has a negative influence on railway performance. Third, the two assessments concluded that the impact of the market function separation is not statistically significant. Although the literature has found a positive impact of separating passengers from freight rail services on railway performance, particularly in minimising railway costs, the results of this research do not support this separation.

Fourth, the two assessments have different results regarding the impact of off-track competition. The competitive tendering system increases the technical inefficiency levels, while this competition form reduces the cost inefficiency levels. The difference between the two findings could be related to the sample size of each assessment. The technical assessment captures the effect of the franchise system, whereas the cost assessment captures the effect of the role of economic regulations. Moreover, the cost assessment captures the role of input prices, where competitive tendering encourages the optimal prices between inputs.

Last, this research has examined railway systems that have different organisational forms in terms of combining rail policies. The two assessments found that combining rail policy at any level had a negative influence on railway performance. This finding suggests that the railway system can be reformed by introducing rail policies gradually rather than in a one-package approach. This option might assist the governments in maintaining the process of reforming the railway industry in several ways. First, sequential reform can enable governments to select the appropriate rail policies that fit their plans. Second, governments may be able to facilitate the railway system to accommodate new regulations or parties. Finally, governments can monitor railway performance to make any necessary adjustments to improve the reform process.

The second part is to combine the findings of the technical and cost assessments with the welfare assessment and expert interviews to draw some recommendations for the Saudi railway system. The above recommendations can be applied to the Saudi railway system with slight adjustments that are driven and combined based on the welfare assessment and expert interviews. For the vertical structure, vertical separation is recommended for the Saudi railway system. The three assessments and expert interviews are in line with this recommendation, particularly the institutional separation. Furthermore, the technical and cost assessments indicated that the vertical separation has a positive impact at lower traffic density, which the Saudi railway system has.

For the horizontal structure, the three assessments found that the regional separation of rail services has negative impacts on the technical cost inefficiency levels and is not statistically significant in the demand and cost model in the welfare assessment. In contrast, the results of the expert interviews showed an even split between supporters and opponents of the recommendation for this document. This split can be explained by the merger between SRO and SAR, which shows the position of the participants regarding this merger. Having said that, the regional separation policy is not recommended for the Saudi railway system.

Besides the regional separation, the results of the three assessments indicated that the impact of the market function separation is not statistically significant. In contrast, the majority of Saudi railway experts recommend the separation between passenger and freight rail services, which is in line with the literature. Although the market function separation is not supported by the results of the three assessments, this research will recommend this policy based on the results of the expert interviews. Therefore, the separation between passenger and freight rail services is the only policy that is recommended at the horizontal level.

The results of assessing the impact of on-track competition vary between the three assessments and expert interviews. The technical and cost assessments indicated that this policy has a negative impact, which increases the technical and cost inefficiency levels. On the other hand, this policy has a statistically significant impact on the demand side of the welfare assessment. In addition, the Saudi rail experts recommend deregulating the freight rail market for new entrants. Having said that, there is a mixture in the findings of this research. However, this research can recommend freight market deregulation because of the results of the welfare assessment and expert interviews.

With regards to off-track competition, this research has different findings between the three assessments and expert interviews. In addition to the difference in the impact of competitive tendering between the technical and cost assessments as discussed above, this policy has no significant impact on the demand and cost models in the welfare assessment. The Saudi rail experts also do not recommend this policy for the Saudi railway system. As a result, competitive tendering for passenger rail services is not recommended for the Saudi railway system.

This research took into consideration the open-access competition for passenger rail services in conducting the expert interviews, in addition to the on-track and off-track competition forms. The results of these interviews indicated that the majority of Saudi rail experts are in a position of disagreement and neutral about recommending this policy. Therefore, freight market deregulation is the only policy that can be recommended for the Saudi rail industry.

Regarding combined rail policies, the welfare assessment has similar findings to the technical and cost assessments. Combining rail policies at any level has no statistically significant impact on the demand and cost models in the welfare assessment. However, the majority of Saudi rail experts recommend vertical separation combined with horizontal separation and competition introduction. These policies can be identified as institutional separation, market function separation, and freight market deregulation policies. To achieve freight market deregulation, it requires the railway system to be separated vertically and functionally. If these policies are recommended, it is not recommended to reform the Saudi railway system in one package as discussed above.

9.3 Limitations of the Study

This research has some limitations to be conducted. Data availability is the major limitation of this research, which can be summarised as follows:

- Each assessment of technical, cost and welfare has a different sample size. The technical assessment has the highest number of observations, and these observations dropped significantly for the cost and welfare assessments because the sources for the cost data are quite limited.
- Data availability played a major role in selecting the railway system. The sample size has been selected based on the available data from the UIC. There was an attempt to collect data from other sources, but the language and data use licences were the obstacles.
- The data is missing when a change in the railway system has occurred. For example, when the French railway system was reorganised from organisational to institutional separation, the data was missing. In addition, the issue is similar to that of the Turkish railway system when the system was reformed to be vertically separated. These changes could help to improve the findings of this research.

Another limitation is that several variables were missing or misspecified, and these can be summarised as follows:

- In the cost assessment, the cost efficiency has been estimated based on the variable costs, not the total costs. This decision has been made because the capital costs and infrastructure access charges were missing. This could improve the estimation of the cost efficiency, particularly for the railway systems that are organised in a vertical separation form.
- In the welfare assessment, the welfare has been calculated assuming no externalities. This could strengthen the calculations of welfare more accurate, particularly with the increase in demand due to the freight market deregulation.

Another limitation is the low response from the rail experts. Although this research aimed to involve rail experts who have at least five years of experience, this condition has been violated by including experts with only three years of experience. Additionally, there is no balance between the participants who are from the Saudi railway companies and the Saudi railway authorities represented by the Ministry of Transport and Public Transport Authority (PTA). The sample size contains only one participant from the PTA, while the remaining six participants are from the Saudi railway companies.

9.4 Directions for Future Research

The directions for future work can be planned to address the above limitations, which can be summarised as follows:

- Expanding the research to include more railway industries;
- Calculating the costs of railway systems more precisely;
- The balance sample size for the three assessments;
- Includes the externalities in the welfare calculations.
- Combining the different models in a consisted framework (e.g. technical assessment (COLS and DEA models), cost assessment (SFA and DEA models) Welfare assessment (negative exponential demand and translog cost model).
- Developing a welfare framework by incorporating different modes of transportation.

Appendix A

Figures for the Welfare Assessment

This appendix presents the figures of demand and cost model variables.

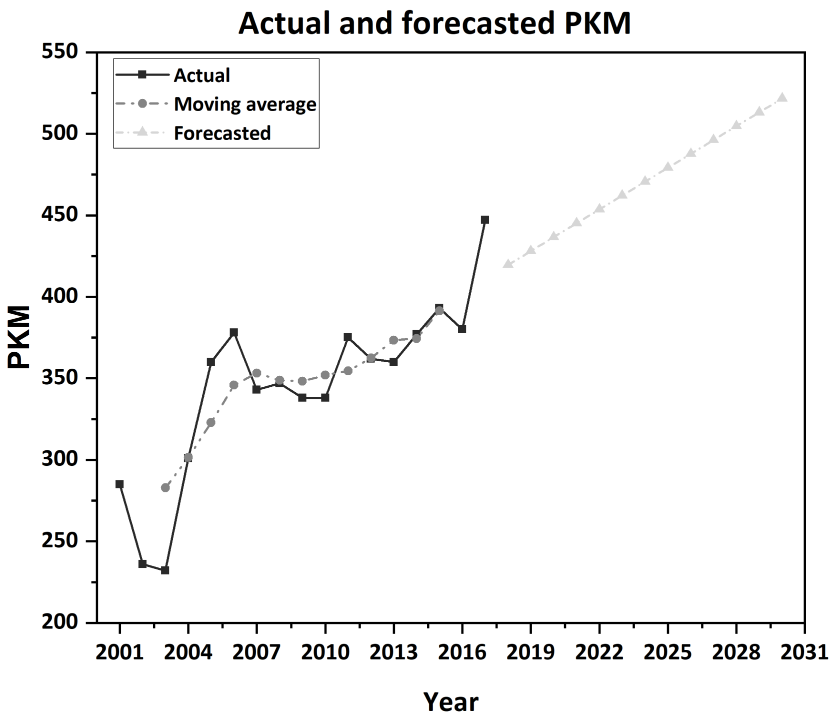


Figure A.1: Actual and predicted PKM based on the moving average trend-line.

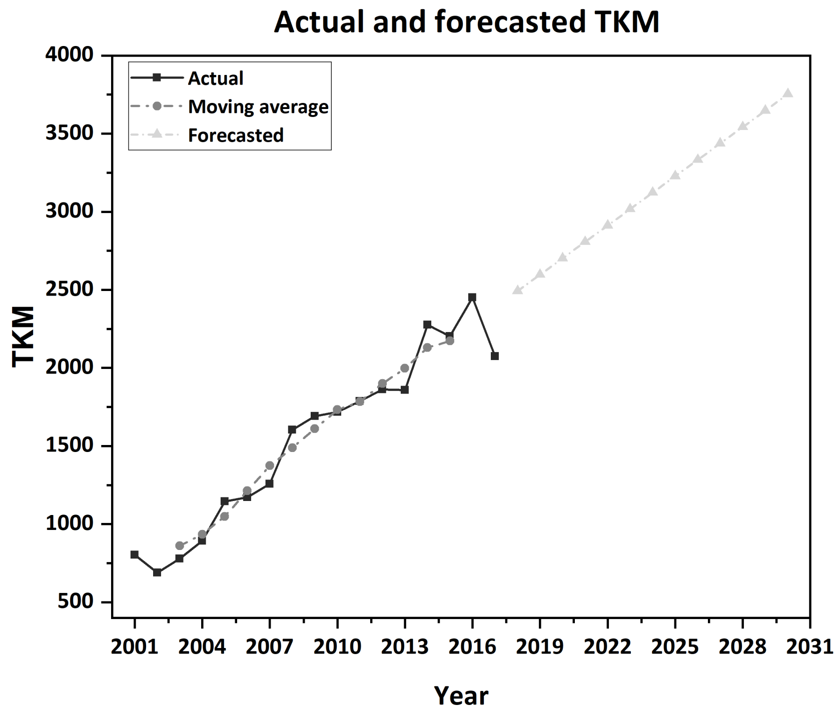


FIGURE A.2: Actual and predicted TKM based on the moving average trend-line.

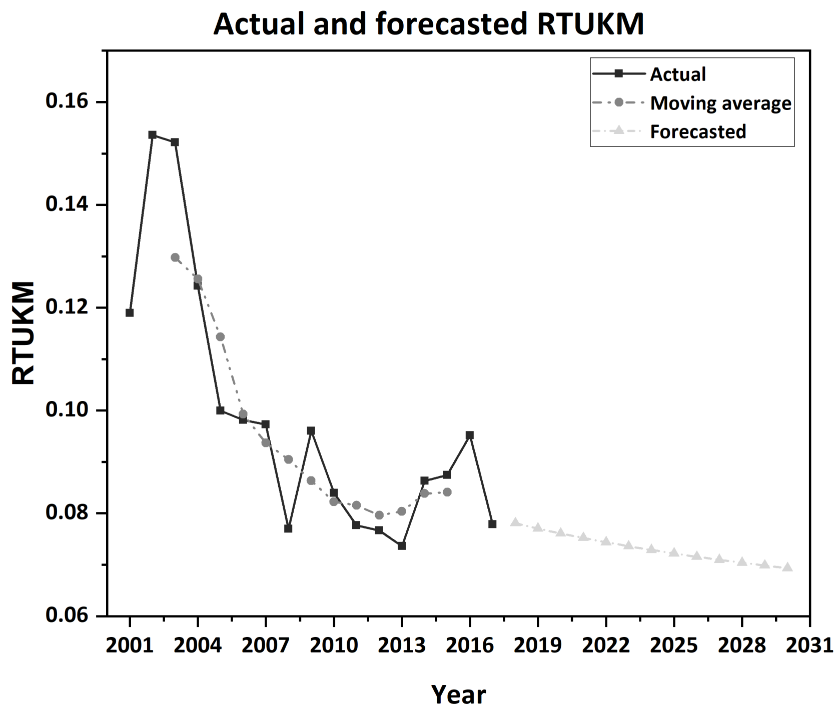


FIGURE A.3: Actual and predicted RTUKM based on the moving average trend-line.

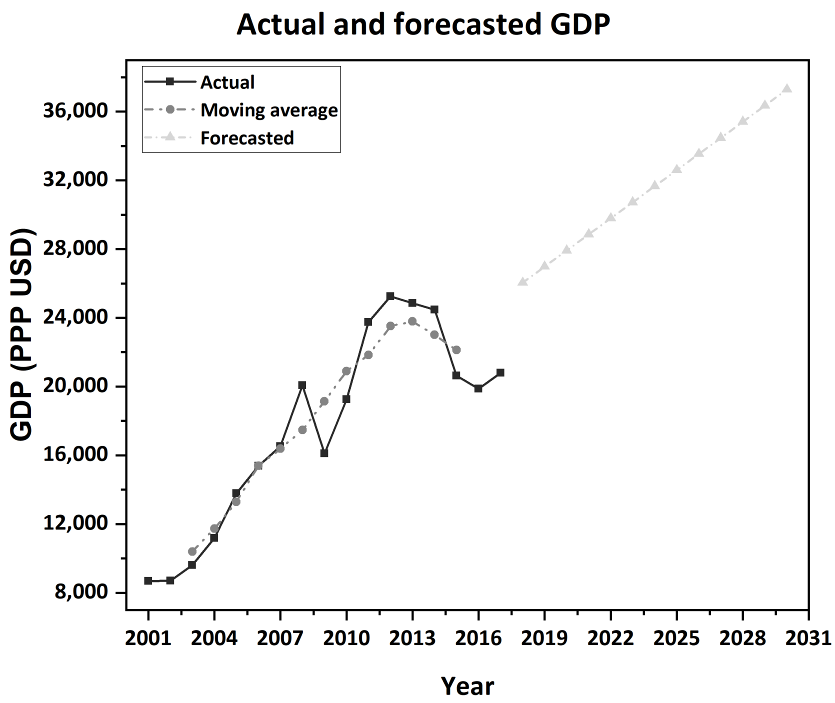


FIGURE A.4: Actual and predicted GDP per capita based on the moving average trend-line.

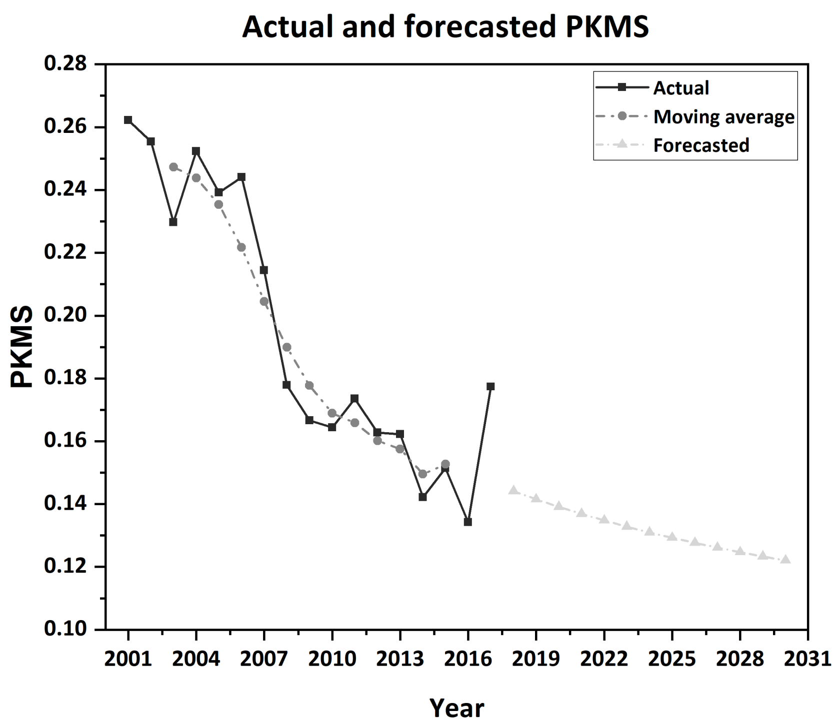


FIGURE A.5: Actual and predicted the passenger share of rail services based on the moving average trend-line.

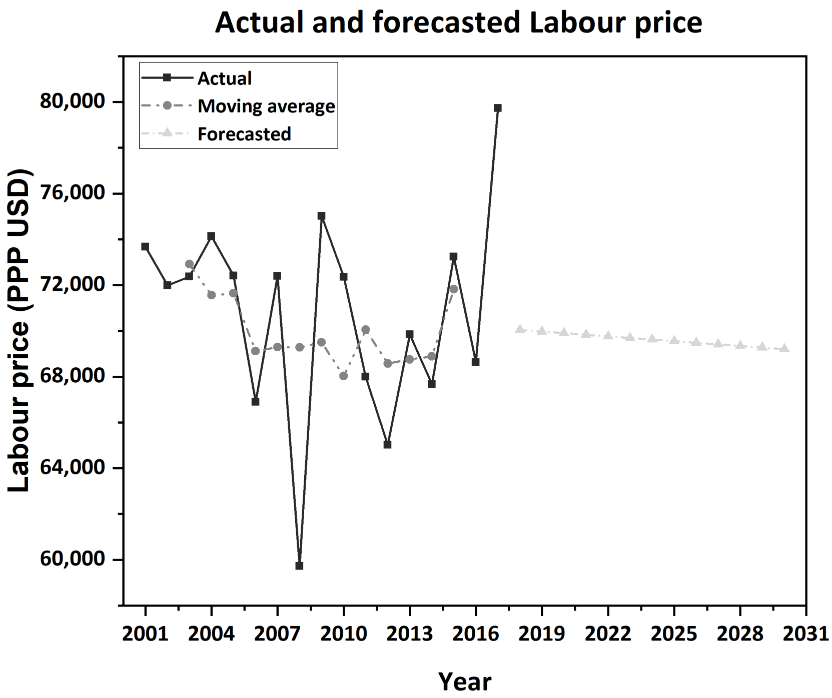


FIGURE A.6: Actual and predicted labour price based on the moving average trend-line.

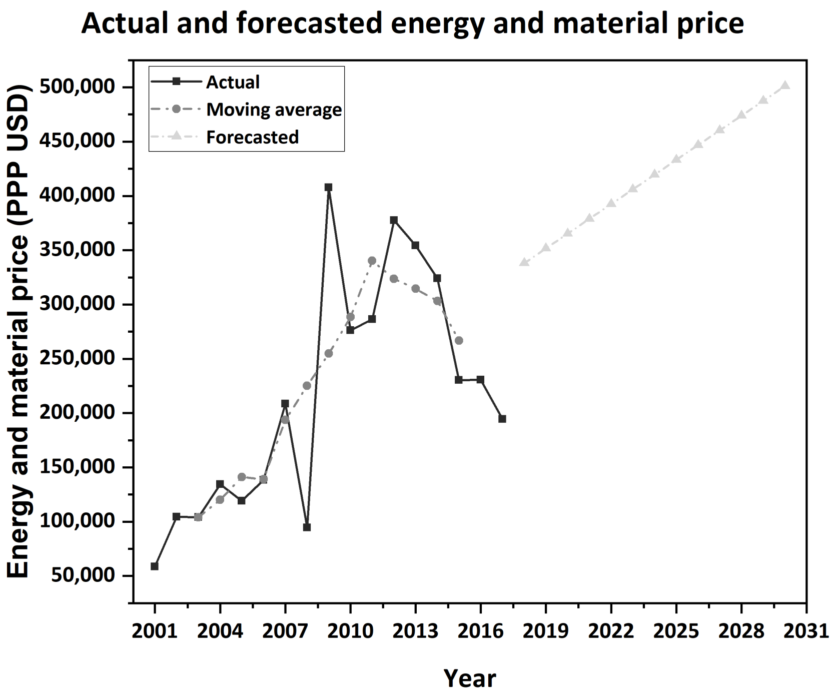


FIGURE A.7: Actual and predicted energy and material price based on the moving average trend-line.

Appendix B

Participant Questionnaire

This section presents the participant questionnaire.

Research title: The theoretical and practical aspects of railway reform: implications for Saudi Arabia.

عنوان البحث: الجوانب النظرية والعملي ??? ا السكك الحديدية —□ الآثار المترتبة على المملكة العربية السعودية.

Dear Sir/Madam,

سيدي العزيز / سيدتي

The purpose of this survey questionnaire is to obtain feedback from rail experts of the Saudi railway system. This feedback will be analysed to validate the recommendations of this research. In this regard, I need your kind cooperation to collect the required feedback to validate these recommendations.

الغرض من هذا الاستبيان هو الحصول على رأي خبراء السكك الحديدية في نظام السكك الحديدية السعودي. سيتم تحليل ??ظة هذه للتحقق من صحة توصيات هذا البحث في هذا الصدد، أحتاج إلى تعاونكم اللطيف لجمع ا ??ظات المطلوبة للتحقق من صحة هذه التوصيات.

I appreciate your time and effort in filling this short survey questionnaire. This will only take 20-25 minutes to complete. I can assure you that, your personal identity will be kept anonymous and confidential. Data collected from you will be used for the sake of academic research. Your participation in this study is entirely voluntary and if you feel uncomfortable at any step in filling this survey, you might leave without informing the researcher.

أقدر وقتك وجهدك في ملء الاستبيان الاستبيان القصير هذا. سيستغرق هذا 20-25 دقيقة فقط لإكماله. يمكنني أن أؤكد لك أن هويتك الشخصية ستبقى مجهولة وسرية. سيتم استخدام البيانات التي يتم جمعها منك من أجل البحث ?□كاديمي. مشاركتك في هذه الدراسة تطوعية تمامًا وإذا شعرت بعدم ا?رتياح في أي خطوة في ملء هذا □ستبيان فيمكنك المغادرة دون إبلاغ ?الباحث.

Section 1: Participant background

القسم ا?ول: خلفية المشارك

1. Where are you based?

…………………………

1. أين يقع مقر عملك؟

……………………………………

3. What is your occupation?

…………………………

3. ما هي مهنتك؟

……………………………………

2. How many years of professional railway experience do you have?

…………………………

2. كم عدد سنوات الخبرة المهنية في السكك الحديدية لديك؟

……………………………………

4. What is your area of expertise?

…………………………

4. ما هو مجال خبرتك؟

……………………………………

Section 2: Recommendations of rail policies for the Saudi railway system

القسم الثاني: توصيات سياسات السكك الحديدية لنظام السكك الحديدية السعودية

1. How likely is it that you would recommend the vertical separation between the infrastructure management and the train operations?

- 0 – Not at all likely.
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 – Extremely likely.

1. ما مدى احتمالية أن توصي بالفصل الرأسي بين البنية التحتية وتشغيل القطارات؟

- 0 - غير محتمل على ؟ ؟؟.
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 - محتمل للغاية.

2. How likely is it that you would recommend the vertical separation between the infrastructure management and the train operations within a holding company?

- 0 – Not at all likely
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 – Extremely likely

2. ما مدى احتمالية أن توصي بالفصل الرأسي بين البنية التحتية وتشغيل القطار داخل شركة قابضة؟

- 0 - غير محتمل على ؟ ؟؟.
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 - محتمل للغاية.

3. How likely is it that you would recommend the horizontal separation of rail services on a regional basis?

- 0 – Not at all likely
- 1
- 2
- 3
- 4
- 5
- 6
- 7

3. ما مدى احتمالية أن توصي بـ ؟!؟ أفقي لخدمات السكك الحديدية على أساس إقليمي؟

- 0 - غير محتمل على ؟ ؟؟.
- 1
- 2
- 3
- 4
- 5
- 6
- 7

- ○ 8
- ○ 9
- ○ 10 – Extremely likely

- ○ 8
- ○ 9
- ○ 10 - محتمل للغاية.

4. How likely is it that you would recommend the horizontal separation between passenger and freight rail services?

- ○ 0 – Not at all likely
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10 – Extremely likely

4. ما مدى احتمالية أن توصي بـ قي بين خدمات السكك الحديدية للركاب والشحن؟

- ○ 0 - غير محتمل على ؟ ؟؟.
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10 - محتمل للغاية.

5. How likely is it that you would recommend a competitive tendering system for passenger rail services?

- ○ 0 – Not at all likely
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10 – Extremely likely

5. ما مدى احتمالية أن توصي بنظام المناقصات التنافسية لخدمات السكك الحديدية للركاب؟

- ○ 0 - غير محتمل على ؟ ؟؟.
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10 - محتمل للغاية.

6. How likely is it that you would recommend market deregulation for new operators for freight rail services?

- ○ 0 – Not at all likely
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8

6. ما مدى احتمالية أن توصي بإلغاء تنظيم السوق لمشغلي الشحن الجدد لخدمات الشحن بالسكك الحديدية؟

- ○ 0 - غير محتمل على ؟ ؟؟.
- ○ 1
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7

- 9
- 10 – Extremely likely

- 8
- 9
- 10 - محتمل للغاية.

7. How likely is it that you would recommend market deregulation (open access competition) for new operators for passenger rail services?

- 0 – Not at all likely
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 – Extremely likely

7. ما مدى احتمالية أن توصي بإلغاء تنظيم السوق (منافسة الوصول المفتوح) للمشغلين الجدد لخدمات السكك الحديدية للركاب؟

- 0 - غير محتمل على ؟ ؟؟.
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10 - محتمل للغاية.

8. How likely is it that you would recommend a combination of rail policies?

- No rail policies are recommended.
- Vertical separation only.
- Horizontal separation only.
- Competition introduction only.
- Vertical separation combined with horizontal separation only.
- Vertical separation combined with competition introduction only.
- Vertical separation combined with horizontal separation and competition introduction.

8. ما مدى احتمالية أن توصي بمجموعة من سياسات السكك الحديدية؟

- يوصى بسياسات السكك الحديدية.
- الفصل الرأسي فقط.
- الفصل الأفقي فقط.
- تقدمة عن المنافسة فقط.
- الفصل الرأسي مع الفصل .
- الفصل الرأسي مع تقدمة المنافسة فقط.
- الفصل الرأسي مع الفصل الأفقي وإدخال المنافسة.

References

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Alexandersson, G. and Hultén, S. (2005). Strategic Pricing by Oligopolists in Public Tenders of Passenger Railway Services. 9th Thredbo Conference in Lisbon. Septemper 2005.

Alexandersson, G. and Hultén, S. (2007). Competitive tendering of regional and interregional rail services in Sweden. In *Competitive Tendering of Rail Services*, pages p. 165–187. OECD Publishing, Paris.

Almanidis, P. and Sickles, R. C. (2012). The Skewness Issue in Stochastic Frontiers Models: Fact or Fiction?". In Van Keilegom, I. and Wilson, P. W., editors, *Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift for Léopold Simar*, pages 201–227. Physica-Verlag HD, Heidelberg.

Almujibah, H. (2021). *High-speed rail forecasting demand model, using Riyadh-Dammam corridor, Saudi Arabia as a case study.* PhD thesis, University of Southampton.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1):3–61.

Asharq Al-Awsat (2021). The merger of railways is a step to privatise the Saudi transport sector. shorturl.at/luxy6. [Accessed 07 July 2022].

Asmild, M., Holvad, T., Hougaard, J. L., and Kronborg, D. (2009). Railway reforms: do they influence operating efficiency? *Transportation*, 36(5):617–638.

Banker, R. D., Charnes, A., and Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9):1078–1092.

Battese, G. E. and Coelli, T. J. (1988). Prediction of firm level technical inefficiencies with a generalized frontier production function. *Journal of Econometrics*, 38:387–399.

Battese, G. E. and Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3(1-2):153–169.

Battese, G. E. and Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20(2):325–332.

Benedetto, V., Smith, A. S., and Nash, C. A. (2017). Evaluating the roles and powers of rail regulatory bodies in Europe: A survey-based approach. *Transport Policy*, 59:116–123.

Berg, S. V. and Tschirhart, J. (1988). *Natural monopoly regulation: principles and practice*. Cambridge University Press, Cambridge.

Boardman, A. E., Laurin, C., Moore, M. A., and Vining, A. R. (2009). A Cost-Benefit Analysis of the Privatization of Canadian National Railway. *Canadian Public Policy*, 35(1):59–83.

Bognetti, G. and Obermann, G. (2008). Liberalization and privatization of public utilities: origins of the debate, current issues and challenges for the future. *Annals of Public and Cooperative Economics*, 79:461–485.

Bošković, B. and Bugarinović, M. (2015). Why and how to manage the process of liberalization of a regional railway market: South-Eastern European case study. *Transport Policy*, 41:50–59.

Bougna, E. and Crozet, Y. (2016). Towards a liberalised European rail transport: Analysing and modelling the impact of competition on productive efficiency. *Research in Transportation Economics*, 59:358–367.

Brage-Ardao, R. (2017). *Determinants of Metro Operating Costs: An Econometric Analysis*. PhD thesis, Imperial College London.

Brenck, A. and Peter, B. (2007). Experience with competitive tendering in Germany. In *Competitive Tendering of Rail Services*, pages 139–164. OECD Publishing, Paris.

Briginshaw, D. (2016). SAR to take responsibility for Saudi rail construction. https://www.railjournal.com/regions/middle-east/sar-to-take-responsibility-for-saudi-rail-construction/ [Accessed 17 October 2019]. [Accessed 17 October 2019].

Briginshaw, D. (2017a). SAR starts first passenger services from Riyadh. https://www.railjournal.com/regions/middle-east/sar-starts-first-passenger-services-from-riyadh/. [Accessed 17 October 2019].

Briginshaw, D. (2017b). Saudi Railways reports strong traffic growth. https://www.railjournal.com/regions/middle-east/saudi-railways-reports-strong-traffic-growth/. [Accessed 17 October 2019].

Briginshaw, D. (2019). Saudi Arabia to fold all rail operations into SAR. https://www.railjournal.com/regions/middle-east/saudi-merges-rail-operations/. [Accessed 17 October 2019].

Cantos, P. and Maudos, J. (2001). Regulation and efficiency: The case of European railways. *Transportation Research Part A: Policy and Practice*, 35(5):459–472.

Cantos, P., Pastor, J. M., and Serrano, L. (1999). Productivity, efficiency and technical change in the European railways: A non-parametric approach. *Transportation*, 26(4):337–357.

Cantos, P., Pastor, J. M., and Serrano, L. (2000). Efficiency measures and output specification: the case of European railways. *Journal of Transportation and Statistics*, 3:61–68.

Cantos, P., Pastor, J. M., and Serrano, L. (2002). Cost and revenue inefficiences in the European railways. *International Journal of Transport Economics*, 29:279–308.

Cantos, P., Pastor, J. M., and Serrano, L. (2010). Vertical and Horizontal Separation in the European Railway Sector and its Effects on Productivity. *Journal of Transport Economics and Policy*, 44(2):139–160.

Cantos, P., Pastor, J. M., and Serrano, L. (2012). Evaluating European railway deregulation using different approaches. *Transport Policy*, 24:67–72.

Casullo, L. (2016). The efficiency impact of open access competition in rail markets: The case of domestic passenger services in Europe. International Transport Forum Discussion Papers No. 2016/07, OECD Publishing, Paris.

Caves, D. W., Christensen, L. R., Tretheway, M. W., and Windle, R. J. (1985). Network effects and the measurement of returns to scale and density for U.S. railroads. *Analytical Studies in Transport Economics*, pages 97–120.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Christopoulos, D., Loizides, J., and Tsionas, E. G. (2001). Efficiency in European railways: not as inefficient as one might think. *Journal of Applied Economics*, 4(1):63–88.

Coelli, T. (1995). Estimators and hypothesis tests for a stochastic frontier function: A Monte Carlo analysis. *Journal of Productivity Analysis*, 6(3):247–268.

Coelli, T. and Perelman, S. (1996). Efficiency measurement, multiple-output technologie and distance functions: With application to European Railways. CREPP Working Papers 96/05, University of Liege.

Coelli, T. and Perelman, S. (1999). A comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operational Research*, 117(2):326–339.

Coelli, T. and Perelman, S. (2000). Technical efficiency of European railways: a distance function approach. *Applied Economics*, 32(15):1967–1976.

Cornwell, C., Schmidt, P., and Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics*, 46(1-2):185–200.

Cowie, J. (1999). The technical efficiency of public and private ownership in the rail industry: The case of Swiss private railways. *Journal of Transport Economics and Policy*, 33(3):241–252.

Cowie, J. (2002). The production economics of a vertically separated railway: the case of the British train operating companies. *Trasporti Europei*, (20):96–103.

Cowie, J. (2017). Contract Re-Award and Consumer Sovereignty, the case of British Rail Privatisation. Paper presented at Competition and Ownership in Land Based Public Transport (Thredbo 15).

Cowie, J. and Riddington, G. (1996). Measuring the efficiency of european railways. *Applied Economics*, 28(8):1027–1035.

Diewert, W. E. and Wales, T. J. (1987). Flexible functional forms and global curvature conditions. *Econometrica*, 55(1):43–68.

Drew, J. and Ludewig, J. (2011). *Reforming Europe's Railways - Learning from Experience*. Community of European Railway and Infrastructure, Brussels, 2nd edition.

Ellig, J. (2002). Railroad deregulation and consumer welfare. *Journal of Regulatory Economics*, 21(2):143–167.

Färe, R. and Primont, D. (1995). *Multi-Output Production and Duality: Theory and Applications, Kluwer Academic*. Kluwer Academic, Boston, Mass, USA.

Farsi, M., Filippini, M., and Greene, W. (2005). Efficiency Measurement in Network Industries: Application to the Swiss Railway Companies. *Journal of Regulatory Economics*, 28(1):69–90.

Finger, M. (2014). Governance of competition and performance in European railways: An analysis of five cases. *Utilities Policy*, 31:278–288.

Finger, M. and Messulam, P. (2015). Rail Economics and Regulation. In Finger, M. and Messulam, P., editors, *Rail Economics, Policy and Regulation in Europe*. Edward Elgar Publishing.

Fitzová, H. (2017). European railway reforms and efficiency: Review of evidence in the literature. *Review of Economic Perspectives*, 17(2):103–120.

Friebel, G., Ivaldi, M., and Vibes, C. (2010). Railway (De)regulation: A European efficiency comparison. *Economica*, 77(305):77–91.

Fröidh, O. and Nelldal, B. L. (2015). The impact of market opening on the supply of interregional train services. *Journal of Transport Geography*, 46:189–200.

Gómez-Ibáñez, J. A. and de Rus, G. (2006). *Competition in the railway industry: An international comparative analysis*. Edward Elgar Publishing.

Green, A. and Mayes, D. (1991). Technical Inefficiency in Manufacturing Industries. *Economic Journal*, 101(406):523–538.

Greene, W. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, 19(2-3):179–190.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303.

Growitsch, C. and Wetzel, H. (2009). Testing for economies of scope in European railways: an efficiency analysis. *Journal of Transport Economics and Policy*, 43(1):1–24.

Grushevska, K., Notteboom, T., and Shkliar, A. (2016). Institutional rail reform: The case of Ukrainian Railways. *Transport Policy*, 46:7–19.

Heatley, D. and Schwass, M. (2014). Rail Transport in New Zealand. In *Priorities and Pathways in Services Reform - Part II*, pages 151–184. World Scientific.

IBM and Kirchner, C. (2011). *Rail Liberalisation Index 2011. Market opening: comparison of the rail markets of the Member States of the European Union, Switzerland and Norway*. IBM Corporation, Berlin-Zurich-Brussels.

International Railway Journal (2011). Spanish consortium is preferred bidder for Saudi HS contract. https://www.railjournal.com/news/spanish-consortium-is-preferred-bidder-for-saudi-hs-contract/. [Accessed 17 October 2019].

Ishida, Y. (2011). Japan. In Drew, J. and Ludewig, J., editors, *Reforming Europe's Railways - Learning from Experience*, pages p. 23–31. Community of European Railway and Infrastructure, Brussels, 2nd edition.

Jensen, A. and Stelling, P. (2007). Economic impacts of Swedish railway deregulation: A longitudinal study. *Transportation Research Part E: Logistics and Transportation Review*, 43(5):516–534.

Johnson, D. and Nash, C. (2012). Competition and the provision of rail passenger services: A simulation exercise. *Journal of Rail Transport Planning and Management*, 2(1-2):14–22.

Kodde, D. A. and Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, pages 1243–1248.

Kumbhakar, S., Orea, L., Rodríguez-Álvarez, A., and Tsionas, E. (2007). Do we estimate an input or an output distance function? An application of the mixture approach to European railways. *Journal of Productivity Analysis*, 27(2):87–100.

Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46(1-2):201–211.

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.

Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2017). Stochastic Frontier Analysis: Foundations and Advances. Working Papers 2017-10, University of Miami, Department of Economics.

Kumbhakar, S. C., Wang, H. J., and Horncastle, A. P. (2015). *A practitioner's guide to stochastic frontier analysis using Stata*. Cambridge University Press.

Kurosaki, F. (2018). A study of vertical separation in Japanese passenger railways. *Case Studies on Transport Policy*, 6(3):391–399.

Laird, P. G. (2013). Government rail asset sales, and return to the public sector, in new zealand and tasmania. *Research in Transportation Business & Management*, 6:116–122.

Lan, L. W. and Lin, E. T. J. (2006). Performance Measurement for Railway Transport: Stochastic Distance Functions with Inefficiency and Ineffectiveness Effects. *Journal of Transport Economics and Policy*, 40(3):383–408.

Lang, M., Laperrouza, M., and Finger, M. (2010). The effects of increased competition in a vertically separated railway market. Working Papers 0025, Swiss Economics.

Last, A. K. and Wetzel, H. (2010). The efficiency of German public theaters: a stochastic frontier analysis approach. *Journal of Cultural Economics*, 34(2):89–110.

Lee, Y. H. and Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*, pages 237–255.

Leibenstein, H. (1966). Allocative efficiency vs." X-efficiency". *The American Economic Review*, 56(3):392–415.

Lerida-Navarro, C., Nombela, G., and Tranchez-Martin, J. M. (2019). European railways: Liberalization and productive efficiency. *Transport Policy*, 83:57–67.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*, 22:1–55.

Link, H. (2004). Rail infrastructure charging and on-track competition in Germany. *International Journal of Transport Management*, 2(1):17–27.

Link, H. and Merkert, R. (2011). Success factors and pitfalls of regional rail franchising in Germany. *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti*, 38:173–200.

Loizides, J. and Tsionas, E. (2002). Productivity growth in European railways: a new approach. *Transportation Research Part A: Policy and Practice*, 36(7):633–644.

Lovell, C., Travers, P., Richardson, S., and Wood, L. (1994). Resources and functionings: A new view of inequality in Australia. In Eichhorn, W., editor, *Models and measurement of welfare and inequality*, pages 787–807, Berlin. Springer-Verlag.

Lust, A. (2017). Broken rails: the privatisation of Estonian railways. *Post-Communist Economies*, 29(1):71–89.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, New York.

Mankiw, N. G. (2016). *Principles of economics.* Cengage Learning, Boston, 8th ed. edition.

Meeusen, W. and van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, 18(2):435–444.

Mehmood Alam, K. (2017). Efficiency Evaluation of Pakistan Railways Using Data Envelopment Analysis. *Civil and Environmental Research*, 9(4):1 – 6.

Merkert, R. (2012). An empirical study on the transaction sector within rail firms. *Transportmetrica*, 8(1):1–16.

Merkert, R., Smith, A. S., and Nash, C. A. (2012). The measurement of transaction costs evidence from european railways. *Journal of Transport Economics and Policy*, 46(3):349–365.

Mizutani, F. (2004). Privately owned railways' cost function, organization size and ownership. *Journal of Regulatory Economics*, 25(3):297–322.

Mizutani, F., Kozumi, H., and Matsushima, N. (2009). Does yardstick regulation really work? Empirical evidence from Japan's rail industry. *Journal of Regulatory Economics*, 36:308–323.

Mizutani, F. and Nakamura, K. (1997). Privatization of the Japan national railway: Overview of performance changes. *International Journal of Transport Economics*, 24(1):75–99.

Mizutani, F., Smith, A., Nash, C., and Uranishi, S. (2015). Comparing the Costs of Vertical Separation, Integration, and Intermediate Organisational Structures in European and East Asian Railways. *Journal of Transport Economics and Policy*, 49(3):496–515.

Mizutani, F. and Uranishi, S. (2013). Does vertical separation reduce cost? An empirical analysis of the rail industry in European and East Asian OECD Countries. *Journal of Regulatory Economics*, 43(1):31–59.

MoT (2017). Roadways. https://mot.gov.sa/en-us/TKingdom/Pages/Road.aspx. [Accessed: 6 January 2018].

Mulder, M., Lijesen, M., Driessen, G., and Van de Velde, D. (2005). Vertical separation and competition in the Dutch rail industry. A cost-benefit analysis. In *Third Conference on Railroad Industry Structure, Competition and Investment*, pages 1–32. Stockholm, 21 and 22 October 2005.

Nash, C. (2012). Towards optimal regulatory structures and practices for the rail sector. Issue Paper No. 3, Centre on Regulation in Europe.

Nash, C. and Preston, J. (1992). Barriers to entry in the railway industry. Working Paper 354, Institute of Transport Studies, University of Leeds, Leeds, UK.

Nash, C. and Smith, A. (2011). Britain. In Drew, J. and Ludewig, J., editors, *Reforming Europe's Railways - Learning from Experience*, pages p. 89–101. Community of European Railway and Infrastructure, Brussels, 2nd edition.

Nash, C. A., Smith, A. S. J., van de Velde, D., Mizutani, F., and Uranishi, S. (2014). Structural reforms in the railways: Incentive misalignment and cost implications. *Research in Transportation Economics*, 48:16–23.

Obermauer, A. (2001). National Railway Reform in Japan and the EU: Evaluation of Institutional Changes. *Japan Railway and Transport Review*, 29:24–31.

Oum, T. H., Waters, W. G., and Yu, C. (1999). A Survey of Productivity and Efficiency Measurement in Rail Transport. *Journal of Transport Economics and Policy*, 33(1):9–42.

Oum, T. H. and Yu, C. (1994). Economic efficiency of railways and implications for public policy: a comparative study of the OECD countries' railways. *Journal of Transport Economics and Policy*, pages 121–138.

Pels, E. and Rietveld, P. (2003). Rail cost functions and scale elasticities: A meta-analysis. Serie Research Memoranda 0003, VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics.

Pitt, M. M. and Lee, L. F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics*, 9(1):43–64.

Preston, J. (1996). The economics of British Rail privatization: An assessment. *Transport Reviews*, 16(1):1–21.

Preston, J. (2001). Franchising and refranchising of passenger rail services in Britain. *Transportation Research Record*, 1742(1):1–8.

Preston, J. (2002). The transaction cost economics of railways. *Trasporti Europei*, 20-21:6–15.

Preston, J. (2008). A review of passenger rail franchising in Britain: 1996/1997–2006/2007. *Research in Transportation Economics*, 22(1):71–77.

Preston, J. (2012). Competition and cooperation, organisation and market: how to deal with barriers to entry and market powers? Issue Paper No. 2, Centre on Regulation in Europe.

Preston, J. (2018). The wheels keep on turning: Is the end of rail franchising in Britain in sight? *Research in Transportation Economics*, 69:187–196.

Preston, J. and Bickel, C. (2020). And the beat goes on. The continued trials and tribulations of passenger rail franchising in Great Britain. *Research in Transportation Economics*, 83:100846.

Preston, J. and Robins, D. (2013). Evaluating the long term impacts of transport policy: The case of passenger rail privatisation. *Research in Transportation Economics*, 39(1):14–20.

Public Transport Authority (2017). *Passenger and freight train services in the Kingdom of Saudi Arabia – Expression of Interest (EoI).* Procurement and Contract at PTA, Riyadh.

Public Transport Authority (2018). *Annual report.* Procurement and Contract at PTA, Riyadh.

Quinet, E. (2011). France. In Drew, J. and Ludewig, J., editors, *Reforming Europe's Railways - Learning from Experience*, pages p. 79–87. Community of European Railway and Infrastructure, Brussels, 2nd edition.

Railway Sector (2018). *Saudi Public Transport Authority - Railway Sector: annual report - 2018.* Railway Sector at PTA, Riyadh.

Ramanathan, R. (2003). *An introduction to data envelopment analysis: a tool for performance measurement*. Sage Publications, New Delhi.

Ranghetti, D. (2018). Railway Reform in the ECE region. Report No. ECE/TRANS/261, United Nations Publication.

Ritter, C. and Simar, L. (1997). Pitfalls of normal-gamma stochastic frontier models. *Journal of Productivity Analysis*, 8(2):167–182.

Riyadh Chamber of Commerce (2005). *Saudi Arabia - Regional and global transport hub: a public-private partnership opportunity*. Riyadh Economic Forum.

Riyadh International Convention and Centre (2018). Saudi Transport and logistics infrastructure [pdf].

Savage, I. (1997). Scale economies in United States rail transit systems. *Transportation Research Part A: Policy and Practice*, 31(6):459–473.

Schmidt, P. and Sickles, R. C. (1984). Production Frontiers and Panel Data. *Journal of Business & Economic Statistics*, 2(4):367–374.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton University Press, Princeton, NJ, USA.

Sherman, H. (1992). Data Envelopment Analysis (DEA): Identifying New Opportunities to Improve Productivity. *Tijdschrift voor Economie*, 37(2).

Shleifer, A. (1985). A theory of yardstick competition. *The RAND journal of Economics*, pages 319–327.

Sickles, R. C. and Zelenyuk, V. (2019). *Measurement of Productivity and Efficiency: Theory and Practice*. Cambridge University Press, Cambridge.

Smith, A. S. (2012). The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector. *Transportation Research Part E: Logistics and Transportation Review*, 48(2):503–515.

Smith, A. S., Benedetto, V., and Nash, C. (2018). The impact of economic regulation on the efficiency of European railway systems. *Journal of Transport Economics and Policy*, 52(2):113–136.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13(1):57–66.

Thompson, L. and Bente, H. (2014). What is Rail Efficiency and How Can it Be Changed? International Transport Forum Discussion Papers No. 2014/23, OECD Publishing, Paris.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1):24–36.

van de Velde, D., Nash, C., Smith, A., Mizutani, F., Uranishi, S., Lijesen, M., and Zschoche, F. (2012). EVES-Rail - Economic effects of Vertical Separation in the railway sector. Report for CER-Community of European Railway and Infrastructure Companies, by inno-V (Amsterdam) in cooperation with University of Leeds–ITS, Kobe University, VU Amsterdam University and Civity management consultants, Amsterdam/Brussels.

Vare, R. (2011). Estonia. In Drew, J. and Ludewig, J., editors, *Reforming Europe's Railways - Learning from Experience*, pages p. 185–195. Community of European Railway and Infrastructure, Brussels, 2nd edition.

Winsten, C. (1957). Discussion on Mr. Farrell's paper. *Journal of the Royal Statistical Society*, 120(3):282–284.

World Bank (2017). Railway Reform: Toolkit for Improving Rail Sector Performance. World Bank Group, Washington, D.C.