Determinants of Metro Operating Costs: An Econometric Analysis

by Ruben Brage-Ardao

A thesis submitted to Imperial College London

for the degree of Doctor of Philosophy and the Diploma of Imperial College

to

Imperial College London

Department of Civil and Environmental Engineering

Centre for Transport Studies

August 2017

Abstract

Cities are indispensable for economic development, although urban agglomerations can fail to deliver their potential due to congestion. Metros are key to achieving the benefits of urban densities without the costs of congestion, by means of providing unparalleled fast and high-capacity transportation in densely populated areas. These urban rail systems require large one-off initial investments but also substantial operational costs, in order to provide the regular day-to-day rail service and maintain the metro's fixed and mobile assets in good condition, such as the rail tracks and the trains, respectively.

The aim of this thesis is to better understand what determines these operational costs. The analysis is based on a panel of more than 20 large and medium-sized urban rail systems across the world during the period 2005 to 2014. The research is based on the econometric estimation of cost functions for the aggregate operating costs but also for some of its constituents like train service, station service and rolling stock maintenance costs. Likewise, different outputs are considered to evaluate the cost efficiency of the metros, such as car kilometres, passenger journeys and train hours among others.

The analysis quantifies the effects of economies of scale and economies of density in metro operations, which can vary significantly between different

divisions of the metro operations leading to substantial policy implications. Moreover, the effect of wages, and how this is linked to labour productivity, is also examined. The dissertation also estimates the corresponding cost elasticities for other cost factors such as labour outsourcing, subsidies and network characteristics. Lastly, the research describes the effect of age on urban rail operating costs and also outlines the background time trends on how the metro operating costs could evolve in the future.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work

.....

Ruben Brage-Ardao

Declaration

I hereby declare and confirm that this thesis is entirely the result of my own original work and that I am the only author of this thesis. The contribution of my supervisors was only supervisory and editorial.

I further declare that all sources cited or quoted are indicated and acknowledged in the list of references in this thesis. Reedited segments of this dissertation have been presented in academic conferences, published and submitted for journal publication. This is clearly indicated at the beginning of the corresponding chapters of this thesis.

.....

Ruben Brage-Ardao

.....

Professor Daniel J. Graham

.....

RTSC Director Richard Anderson

Acknowledgements

I would like to thank Professor Dan Graham and Dr. Patricia Melo for giving me the opportunity to join Imperial College back in 2011. Professor Graham has not only supervised this thesis but he has also offered professional guidance and has been a role model to me over the last few years.

I would also like to express my gratitude to the RTSC Director, Richard Anderson, who has supervised this PhD and whose expertise in metros is probably unparalleled worldwide. I am also truly thankful to the RTSC management led by Richard Anderson, Mark Trompet, Alexander Barron and Ben Condry, for giving me the opportunity to work for the RTSC since 2012, providing funding and making my work compatible with my PhD. Thanks also to all the inspirational people working at the RTSC, including Shane Canavan, Judith Cohen, Oscar Criado, Sally Kao, Lindsey Morse, Deborah Mundy, Dan Piner, John Sing, Haobo Yang and PhD students Daniel Hörcher, Ramandeep Singh, Samira Barzin and the Centre for Transport Studies.

Lastly, I would also like to thank my family and friends, particularly Henrique, for his example, to Rosa, for her courage, and to Elena, for her relentless support and motivation.

Contents

Abstract **1**

Declaration **3**

Declaration **4**

Acknowledgments **5**

List of Figures **11**

List of Tables **14**

List of Acronyms **16**

Notations and Definition of Variables **18**

1 Introduction **20**

1.1 Research Background 20

1.2 Scope of the Research 23

1.3 Research Aims and Objectives 27

1.4 Contribution of this Research 32

1.5 Academic Publications and Conferences 34

2 Literature Review **36**

2.1 Concept of Cost Efficiency 36

2.2 Scope of the Review 37

2.3 Approaches to Assess Rail Cost Efficiency 39

2.3.1 Total factor productivity 39

2.3.2 Nonparametric: data envelopment analysis 43

2.3.3 Parametric: deterministic and stochastic 49

2.4 Literature Review Conclusions 55

3 Methodology for Cost Efficiency Estimation **63**

3.1 Nonparametric Frontier Methods 65

3.1.1 Data envelopment analysis 65

3.1.2 Stochastic data envelopment analysis 71

3.2 Parametric Frontier Methods 72

3.2.1 Deterministic econometric analysis 72

3.2.2 Stochastic frontier analysis 83

3.3 Methodological Challenges of the Research 88

3.3.1 Unobserved heterogeneity 88

3.3.2 Functional form 90

3.3.3 Omitted variable bias 92

3.3.4 Reverse causality 94

3.3.5 Measurement error 95

3.4 Methodology Review Conclusions 97

4 Data **99**

4.1 RTSC and CoMET-Nova Consortia 100

4.2 Balanced and Unbalanced Panel Data 105
4.3 CoMET-Nova Financial Comparability 113

5 Operating Costs in Urban Rail Systems 116

5.1 Introduction 116
5.2 Literature Review 117
5.3 Data and Research Hypothesis 118
5.4 Cost Function Results 129
5.5 Cost and Production Function Duality 137
5.6 Production Function Estimation 143
5.7 Production Function Results 149
5.7.1 Cost and Production Function Duality Results 155
5.8 Conclusion 158

6 Train Service Costs in Urban Rail Systems 161

6.1 Introduction 161
6.2 Literature Review 163
6.3 Data and Research Hypotheses 166
6.4 Results 178
6.5 Conclusion 188

7 Station Service Costs in Urban Rail Systems 190

7.1 Introduction 190
7.2 Literature Review 191
7.3 Data and Research Hypotheses 193
7.4 Results 202

7.5 Conclusion 209

8 Rolling Stock Maintenance Costs in Urban Rail Systems 212

8.1 Introduction 212

8.2 Literature Review 213

8.3 Data and Research Hypotheses 217

8.4 Results 232

8.5 Conclusion 243

9 Conclusions 246

9.1 Main Findings and Contributions to the Existing Literature . 246

9.1.1 General Findings and Contributions 246

9.1.2 Cost-Specific Findings and Contributions 251

9.2 Policy Implications 253

9.3 Limitations of the Study 257

9.4 Directions for Future Research 259

10 Bibliography 262

11 Annex 282

11.1 Total and average cost elasticities 283

11.2 Reduced Operating Costs Functions - GMM 284

11.3 Production function - Comparison estimations 286

11.4 Detailed definition of explanatory variables 289

11.5 Econometric models with age categorisation 294

11.6 Econometric models without input quantities 299

11.7 Breusch-Pagan Test Results 304

11.8 SFA Efficiency Index 305

List of Figures

1.1 Key characteristics of urban rail systems. 21

1.2 Corridor capacity per hour by urban transport mode (Breithaupt, 2010) 22

1.3 Population density and urban rail systems 24

1.4 Scope of the research - Aggregate operating costs and other operating costs in the CoMET-Nova framework. 25

1.5 Operating cost breakdown for CoMET-Nova metros (2014) . . 30

2.1 DEA example: production frontier with inputs X and Y. . . . 44

2.2 Comparison of OLS and COLS 50

3.1 DEA types under different technical assumptions 68

3.2 Comparison of OLS, COLS and MOLS frontiers 81

3.3 Distribution of error, efficiency and overall error term in the SFA 85

3.4 Comparison of SFA, COLS and OLS methods 87

4.1 CoMET and Nova metro members (January 2015). 101

4.2 Scope of the research - Aggregate operating costs and other operating costs in the CoMET-Nova framework. 102
4.3 Average CoMET-Nova costs 1994-2015. 103
4.4 CoMET-Nova data framework 111
4.5 Completeness of data and number of observations. 112

5.1 Aggregate operating costs and other operating costs of a typical CoMET-Nova metro operator. 119
5.2 Operating costs per car kilometre. 121
5.3 Operating costs per passenger journey. 122
5.4 Operating costs per train hours. 123

6.1 Train service costs and other operating costs of a typical CoMET-Nova metro operator. 166
6.2 Train service costs per car kilometre. 169
6.3 Train service costs per passenger journey. 170
6.4 Train service costs per train hour. 171

7.1 Station service costs and other operating costs of a typical CoMET-Nova metro operator. 194
7.2 Station service costs per passenger journey. 196
7.3 Station service costs per station. 197

8.1 Rolling stock maintenance costs and other operating costs of a typical CoMET-Nova metro operator. 218
8.2 Rolling stock maintenance costs per car kilometre. 221
8.3 Rolling stock maintenance costs per car. 221

8.4 Average rolling stock maintenance cost per car km depending on age (in US$ PPP). All other cost determinants held constant. 240

11.1 SFA Efficiency indeces for operating costs per car kilometre . . 306

11.2 SFA Efficiency indeces for operating costs per passenger journey 308

11.3 SFA Efficiency indeces for operating costs per train hour . . . 309

List of Tables

2.1 Main rail cost efficiency research 58

3.1 Cost efficiency estimation methods and examples. 63

4.1 Urban rail systems of the CoMET consortium 106

4.2 Urban rail systems of the Nova consortium 107

5.1 Operating costs dependent variable summary statistics 120

5.2 Operating costs explanatory variable summary statistics . . . 124

5.3 Operating Cost Model Results - Random Effects 132

5.4 Car kilometre Sys-GMM Production Function Results 150

5.5 Passenger journey Sys-GMM Production Function Results . . 151

5.6 Train hours Sys-GMM Production Function Results 152

6.1 Train service dependent variable summary statistics 169

6.2 Train service explanatory variable summary statistics 172

6.3 Train Service Cost Model Results - Random Effects 180

7.1 Station service dependent variable summary statistics 195

7.2 Station service explanatory variable summary statistics 199

7.3 Station Service Cost Model Results - Random Effects 204

8.1 RS Maintenance dependent variable summary statistics 220
8.2 RS Maintenance explanatory variables summary statistics . . 222
8.3 RS Maintenance Cost Model Results - Random Effects 234

11.1 Operating Cost Reduced Model Results - GMM 285
11.2 Car kilometres - Production functions 286
11.3 Passenger journeys - Production functions 287
11.4 Train hours - Production functions 288
11.5 Operating Cost - Explanatory variables definition 290
11.6 Train Service - Explanatory variables definition 291
11.7 Station Service - Explanatory variables definition 292
11.8 Rolling Stock Maintenance - Explanatory variables definition . 293
11.9 Operating Cost Model Results - Age groups 295
11.10Train Service Cost Model Results - Age groups 296
11.11Station Service Cost Model Results - Age groups 297
11.12RS Maintenance Cost Model Results - Age groups 298
11.13Operating Cost Model Results - No input quantities 300
11.14Train Service Costs Model Results - No input quantities . . . 301
11.15Station Service Cost Model Results - No input quantities . . . 302
11.16RS Maintenance Cost Model Results - No input quantities . . 303
11.17Breusch-Pagan Lagrange Multiplier Test 304
11.18Operating Cost Model Results - SFA 307
11.19SFA Efficiency Index Descriptive Statistics 308

List of Acronyms

ATOC	Association of Train Operating Companies
BRT	Bus Rapid Transit
COLS	Corrected Ordinary Least Squares
CoMET	Consortium of large metros
CRS	Constant Returns to Scale
DEA	Data Envelop Analysis
Determ.	Deterministic estimation technique
DOLS	Displaced Ordinary Least Squares
DRS	Decreasing returns to scale
FDH	Free disposability hull
FRH	Free replicability hull
GDP	Gross Domestic Product
GMM	General Method of Moments
IRS	Increasing returns to scale
KPI	Key Performance Indicator
MCF	Multiproduct Cost Functions
MDBF	Mean Distance Between Failures
MDBF RS	MDBF caused by RS incidents

MOLS	Modified ordinary least squares
MPI	Malmquist Productivity Index
Obs.	Observations in the dataset
OLS	Ordinary Least Squares
Pax.	Passenger Journeys
Param.	Parametric estimation technique
PF	Production function
PPP	Parity Purchasing Power
RS	Rolling Stock
RTD	Returns to Density
RTS	Returns to Scale
RTSC	Rail and Transport Strategy Centre
SDEA	Stochastic DEA
SFA	Stochastic Frontier Analysis
StoNED	Stochastic nonparametric envelopment of data
TC	Total Cost
TFP	Total Factor Productivity
translog	transcendental logarithmic
VC	Variable Cost
VRS	Variable Returns to Scale

Notation and Definition of Variables

α	Fixed effect or unobserved heterogeneity
β	Variable coefficient or model intercept (i.e. β_0)
ϵ	Estimation error term
γ	DEA estimation economic assumptions
δ	Covariance between a model variable and an omitted variable
μ	Particular unobserved efficiency rate
ν	Random variation of the error term
ω	Predictable or observed productivity shocks
$\wedge$	Scale parameter of DEA estimation
A	General productivity of production function
a	Labour output elasticity
b	Capital output elasticity
E	Total electricity consumed
E^i	Input efficiency for firm i.
FL	Fleet size as number of vehicles
i	Cross sectional unit or panel individual
C	Costs, dependent variable of cost function
K	Capital production input

L	Labour production input
NL	Network length
Q	Output (cost functions)
q	Order of the moment conditions
r	Rent, price of capital
ST	Number of stations
T	Production function and feasible production set
t	Particular time period
w	Wage, price of labour
x	Generic input of a production function
Y	Production output
y	Generic output of a production function
Z	GMM set of instruments

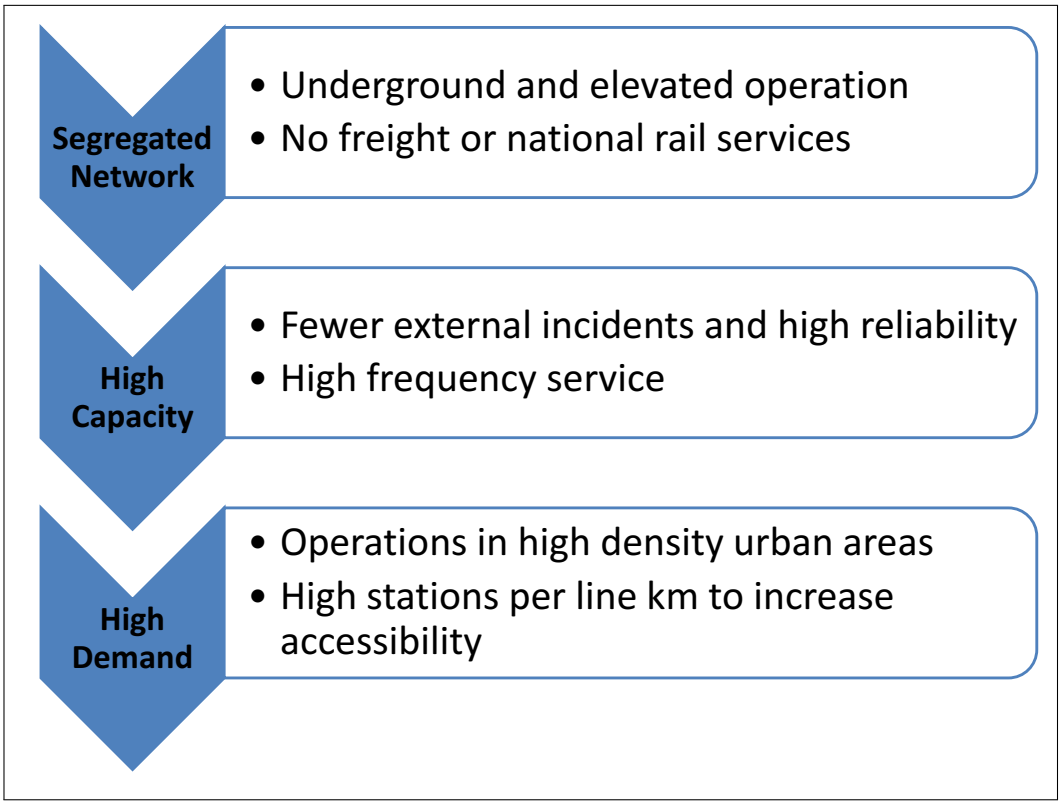
1 — Introduction

1.1 Research Background

Cities are at the core of economic development. In 2012, the largest 300 metropolitan areas contained 19% of the world's population but almost half of the world's GDP (The Economist, 2012). However, despite the productivity advantages derived from agglomeration economies, urban growth can be limited, and even brought to a halt, due to congestion costs (Venables, 2007). Therefore, urban transportation, and particularly metro systems, are critical to sustain urban growth and economic development.

Metro systems are characterised by the combination of a segregated network, with high capacity and demand as shown in Figure 1.1. A defining characteristic of a metro is the use of the network without sharing the right of way with other rail operators (i.e. suburban rail or national rail services). Most of the metros not only have a particular network but also operate substantial sections of the network underground or, to a lesser extend, in elevated rail lines. The exclusive use of the rail network enables the metro to achieve a more reliable service with less external incidents caused by the combination of different operators on the same line. Likewise, higher reliability of the rail

Figure 1.1: Key characteristics of urban rail systems.



service make it possible to reach a higher service frequency, with up to one service every 80-90 seconds in the busiest metros. This higher rail service frequency, and the particular vehicle layout to maximise passenger capacity in the trains, enables the metro to offer unparalleled passenger capacity in densely populated areas as shown in Figure 1.2. This extraordinary supply of capacity is matched by high passenger demand which characterises the high levels of passenger ridership in most metros worldwide. Metros tend to operate mostly in high density urban areas and, additionally, metros also have a low average distance between stations. This low distance between stations increases the accessibility to the system and fosters the passenger

demand needed to meet the supply of capacity.

Figure 1.2: Corridor capacity per hour by urban transport mode (Breithaupt, 2010)

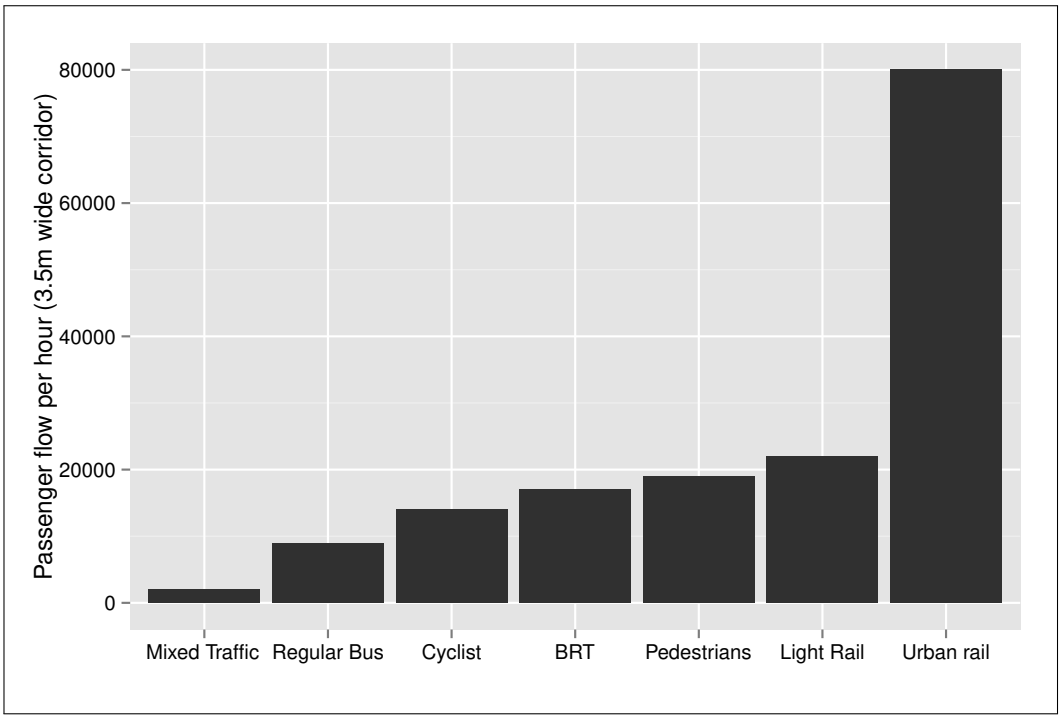


Figure 1.2 shows that the particular layout of the metro rolling stock and the high frequency of the service cause that a metro line can usually transport more than 80,000 passengers per hour. Other modes of transport using the same area, like light rail and walking, can barely reach a quarter of the urban rail capacity (Breithaupt, 2010). Additionally, both light rail and walking are substantially slower than urban rail. In terms of speed, only bus rapid transit systems (BRT) operating on segregated traffic can compete with urban rail systems in terms of speed (Hensher and Golob, 2008). Besides this, despite the fact that the initial investment for BRT is usually less than a quarter of the urban rail initial investment costs (Zhang, 2009), in the long-term

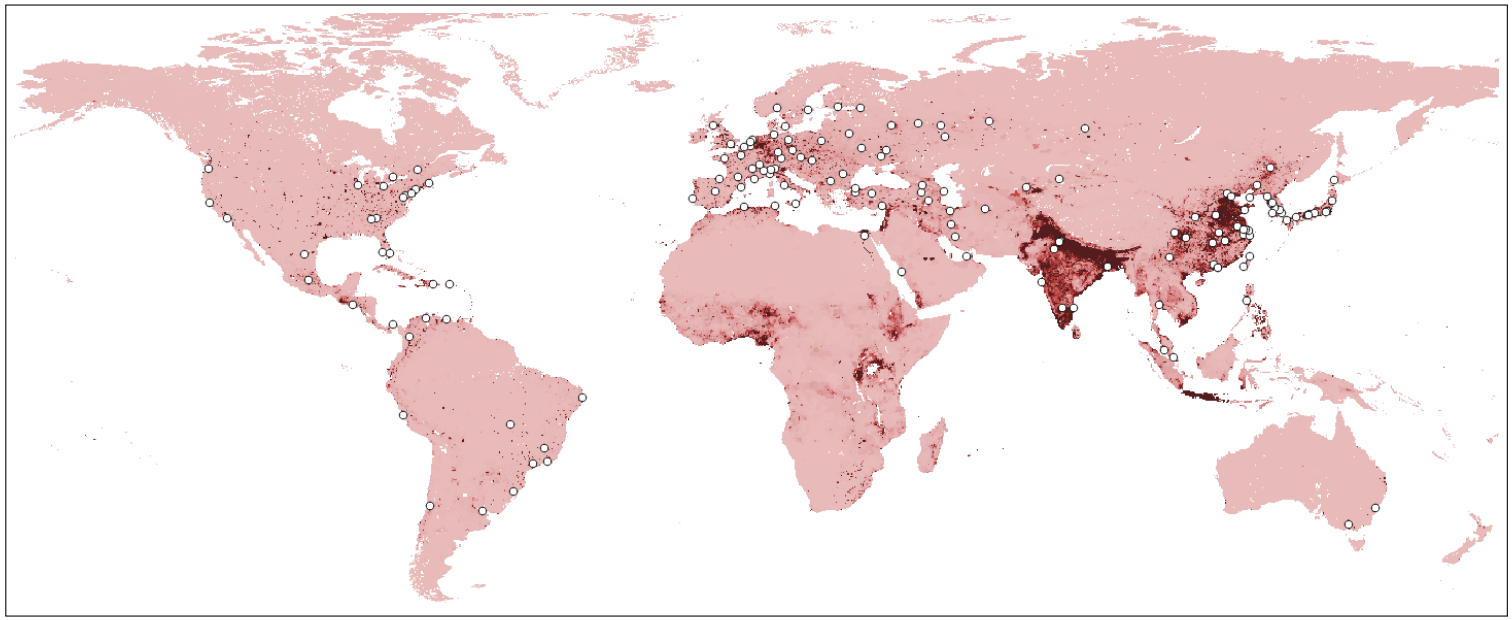
the BRT maximum flow capacity is also substantially lower than the urban rail system (see capacity comparison in Figure 1.2 by Breithaupt, 2010). Therefore, urban rail systems usually have a monopoly in high capacity, high frequency and fast urban transport. Consequently, unless economic growth is unexpectedly decoupled from transport (Armstrong and Preston, 2011), an efficient urban rail system is essential for development. Only those cities with a urban rail system can reach the highest levels of population density and benefit from the agglomeration economies without the costs of congestion.

According to the international organisation for public transport (UITP, 2014), there are almost 150 metro systems in the world with more than 20 new urban railways which started operating in the last 5 years. Moreover, there are at least two dozen currently under construction (The Economist, 2013), mainly in India and China. Based on the population density world map in Figure 1.3, there are still a number or large, densely populated areas without urban rail systems, particularly in Asia and Africa. Likewise, existing urban rail systems keep adding new lines or extending existing ones, especially in developing countries (Railway Gazette, 2014). Thus, despite the fact that metros are already essential for urban transport, their role in urban mobility is bound to become even more important in the near future.

1.2 Scope of the Research

Despite the ever-growing relevance of the urban rail systems, the academic literature on metro cost efficiency is scarce and usually focused on an aggregate analysis of the cost efficiency of the metro.

Figure 1.3: Population density and urban rail systems

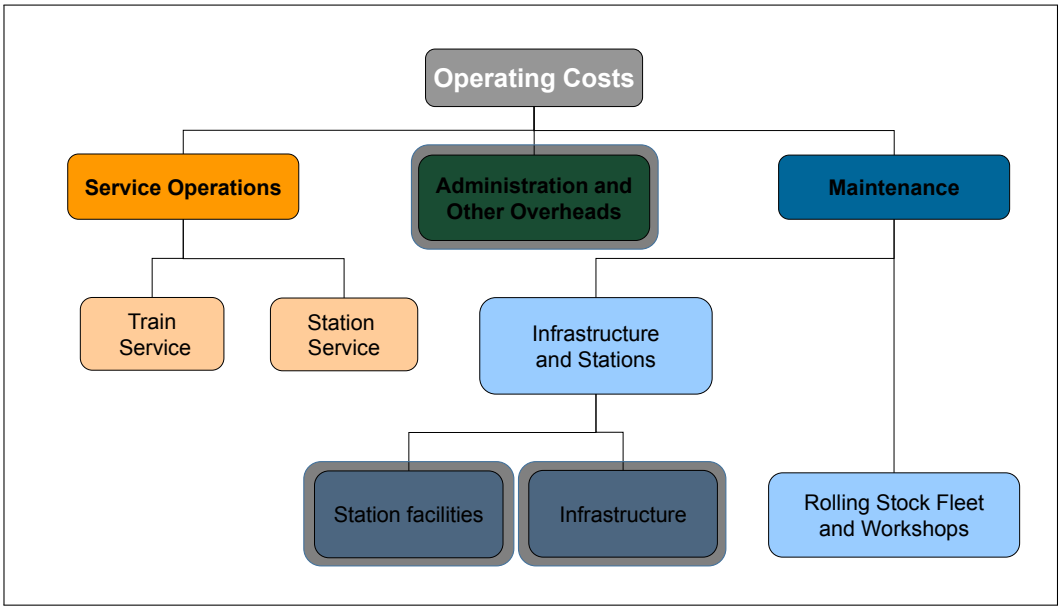


Source: Urban rail systems from UITP (2014) and world densities from CIESIN (2010).

In this research the cost efficiency of the urban rail systems is assessed both at the aggregate and disaggregate level in order to yield a general but also a detailed overview of the metro cost efficiency.

The operating costs considered are shown in Figure 4.2 corresponding to the operating costs breakdown followed by the CoMET and Nova group data base.

Figure 1.4: Scope of the research - Aggregate operating costs and other operating costs in the CoMET-Nova framework.



The aggregate operating cost models are explained in Chapter 5 whereas the lower level of train service, station service and rolling stock maintenance costs are developed in Chapters 6,7 and 8 respectively.

However, not all the operating cost divisions have been included in this thesis due to data availability as shown dark shaded in Figure 4.2. Some of the operating costs figures in the CoMET-Nova data set are not fully comparable due to the following characteristics:

- Infrastructure maintenance costs was excluded, as sometimes the network belongs to another company and the metro company is just the operator, with no real data on infrastructure costs. This impedes the accurate reporting of infrastructure maintenance costs by the metros and compromises comparability between metros.

- Station maintenance costs also present data inaccuracies as some metros are not in charge of the station maintenance or it is difficult to allocate the costs to the metro (e.g. interchange stations with other transport modes or operators). Furthermore, there are huge differences between the small suburban stations and the large, deep underground stations, which make this cost category difficult to model without additional variables on station characteristics that are not available in the CoMET-Nova dataset.

- Administration costs are difficult to split between metros and their matrix organisations. Some metros are standalone companies whereas other metros are divisions of metropolitan transport entities. In this latter case, it is difficult to allocate administration costs to one entity or another as the costs are not completely comparable between different metros. Likewise, administrative structures tend to be less comparable than other operational divisions.

The exclusion of certain cost categories limits the scope of the research and constraints the reach of its conclusions. However, and as explained above, not all the data items in the CoMET-Nova framework are fully comparable and the results from the models with less comparable data may be inaccurate.

Consequently, the conclusions of this research only apply to the non-shaded areas which, in any case, represent most of the operating costs of a metro as seen in the next section on research aims and objectives (Figure 1.5 on page 30).

1.3 Research Aims and Objectives

This thesis pursues several research aims and objectives. First, this study critically reviews the existing literature on urban rail cost efficiency as shown in Chapter 2. Additionally, and due to the scarcity of research for urban rail cost efficiency, the chapter also reviews the main literature on rail cost efficiency. The review of the main findings in the academic literature is followed by the appraisal of the econometric methods applicable for the study of rail cost efficiency. Chapter 3 examines the different approaches to study cost efficiency with particular interest in its advantages and disadvantages for the present case of urban rail cost efficiency.

The main aim and contribution of this thesis is to estimate the key determinants of urban rail cost efficiency. This includes both external or structural factors outside the metro control (e.g. electricity prices) but also internal factors under control of the metro management (e.g. labour outsourcing). The results of this analysis are shown in Chapters 5 to 8. The thesis also aims to contribute to the understanding of the metro operating costs by comparing the aggregate operating costs determinants shown in Chapter 5 to the key cost determinants for particular divisions of the metro operations shown in Chapters 6 to 8.

Additionally, the thesis complements the estimation of cost functions with production functions in Chapter 5. These functions are estimated in order to evaluate the duality theory in metros and assess whether metros operate as optimizing firms.

Overall, the conclusions of this PhD dissertation cast light on the information asymmetry concerning metro cost efficiency. The objective of this thesis is to reveal the main determinants of costs efficiency for metro operators by means of econometric analysis on a group of more than 20 metro operators worldwide. Thus, the econometric analysis supplies an alternative to assess the cost efficiency of urban rail systems in the absence of market competition. This econometric analysis is performed on a panel of more than 20 metros worldwide. Currently, there are few networks or organisations that bring metro systems together and analyse their performance systematically, due to the local scope of the metro systems. Besides this, the monopolistic properties of urban rail systems also imply that for most regulators it is not possible to find many comparable peer systems in the same region or country and, consequently, any systematic comparison of performance should be done at an international level. One of the few networks that links the most important metros in the world has been set up by the Rail and Transport Strategy Centre (RTSC) at Imperial College London. The RTSC manages two metro consortia (CoMET and Nova) comprising more than 30 urban railways with more than 20 years of experience in performance research. The 32 metros participating in the CoMET and Nova consortia collectively transported 22.8 billion passengers in 2014, more than three times the world's population. In economic terms, the CoMET-Nova metros managed a combined operating

cost of US$28.4 billion in 2014 Parity Purchasing Power US$, excluding any asset upgrade, investment or extension costs.

Figure 1.5: Operating cost breakdown for CoMET-Nova metros (2014)

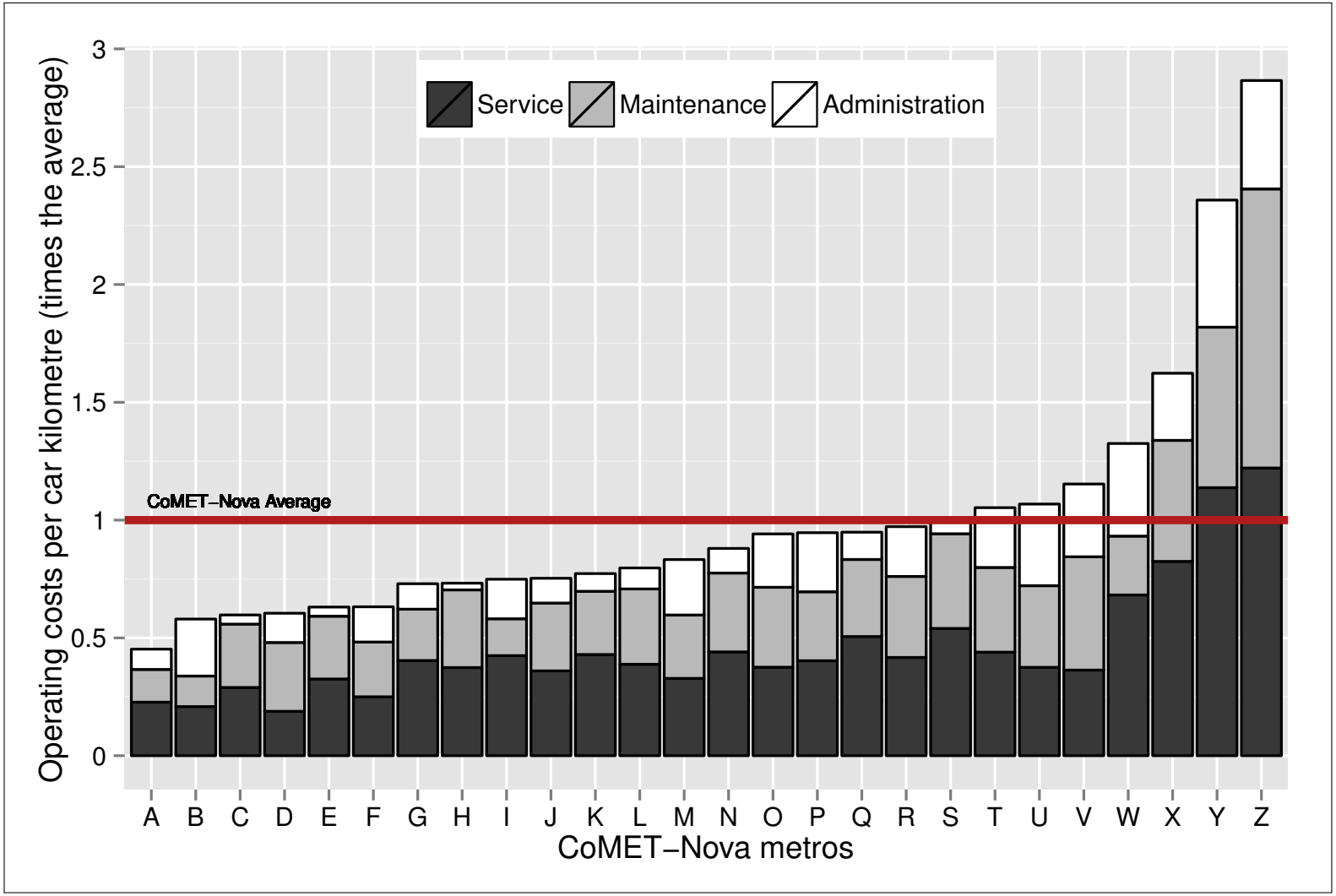


Figure 1.5 demonstrates how operating costs for the CoMET-Nova metros show a wide variability, both in absolute and relative terms. The operating costs per car kilometre range from less than half to almost three times the the CoMET-Nova average operating costs per car kilometre; and these values control for differences in prices between countries by means of indexing costs to Parity Purchasing Power (PPP) [1]. A similar variation is found when comparing the operating costs for other outputs such as train hours and passenger kilometres.

The aim of this research is to understand why some metros are more cost-efficient than others at producing output such as car kilometres, passenger journeys or train hours. This is achieved by assessing the factors determining cost efficiency, regardless of whether they are the responsibility of the metro or outside of metro management control. This would increase the accountability of the metro operators to the public transport authorities, while enabling the transport authorities to comprehend the factors that impair metro cost efficiency. Overall, this can optimise the use of resources and subsidies as well as increase the mutual understanding between the urban rail operator and the public transport authorities, leading to better urban rail regulation.

[1] A detailed explanation on Parity Purchasing Power US$ can be found in section 4.3 on CoMET-Nova financial comparability on page 113.

1.4 Contribution of this Research

The previous section describes the social relevance of the research and outlines the aim of this thesis; to understand the factors that make a metro more cost-efficient at producing a certain output. Most of the research on railway cost efficiency has focused on the freight and regional rail services and, despite its social and economic relevance, urban rail efficiency has received much less attention from academic researchers. This research fills the academic gap on urban rail cost efficiency and also adds additional findings that apply to overall rail transport. In particular, the main academic contributions to the understanding of rail cost efficiency are the following:

1. Collate a data set of more than 30 urban rail systems worldwide based on the CoMET-Nova consortia. This data set adds new urban rail systems and new operational data to the corpus of rail cost efficiency research. This also includes disaggregated data on the operations and maintenance of urban rail systems.

2. Test new hypotheses on detailed operational characteristics based on the disaggreagated data set. These hypotheses would add empirical evidence to the academic literature on the effect of key cost drivers that have seldom been tested (e.g. role of labour outsourcing in cost efficiency).

3. Identify cost determinants at aggregate and disaggregate operational levels. Most of the academic literature focuses only on the overall operating costs whereas the effect of the same cost driver can vary between

different areas of metro operations. This research tests whether the effect of a particular cost driver (e.g. wages) varies between different areas of operations such as train service or rolling stock maintenance. Likewise, the research also considers particular cost drivers for the sub-categories of the operating costs.

4. Estimate the effect of the economies of scale and density on the operating costs of metros. Likewise, identify whether the economies of scale and density are homogeneous across all the operational sections of the metros or whether different parts of the operations present different economies of scale or density. In either case, clarifying the role of economies of density and scale within the urban rail systems would lead to substantial regulatory and policy implications.

5. Apply econometric techniques that have not yet been used either to model the cost efficiency of the urban rail systems nor to model the data set of the CoMET-Nova consortia.

Overall, this research aims to fill the gap in the academic literature on urban rail, and particularly on urban rail cost efficiency. The conclusions based on this research can be useful for industry practitioners in terms of how to optimise the costs of the rail operator, but also for rail sector regulators, as a way to better understand rail operator cost drivers and produce better policies and regulations.

1.5 Academic Publications and Conferences

Thesis chapters published or submitted for publication:

1. *Determinants of Train Service Costs in Metro Operations* published in the *Transportation Research Record: Journal of the Transportation Research Board: Public Transportation. Passenger Rail and Terminals* - Volume 2 No.2534 in November 2015.

2. *Determinants of Rolling Stock Maintenance Cost in Metros* published in the *Journal of Rail and Rapid Transit* in October 2015.

3. *Determinants of Operating Costs in Metro Operations* in submission.

The thesis chapters have also been accepted for presentation at several international conferences on transport economics:

1. Metro Operating Costs: Main Patterns and Determinants presented at *Transportation Research Board* - Lectern Session 17-0404 on *Right Sizing Rail Transit Service* in Washington D.C. in January 2017.

2. Determinants of Rolling Stock Maintenance Cost in Metros presented at the *World Conference on Transport Research* - Presentation Session on *Rail Engineering* in Shanghai, July 2016.

3. Econometric Benchmarking of Metro Operating Costs: Methods and Applications presented at the *World Conference on Transport Research* - Presentation Session on *Assessment and Evaluation of Logistics Systems* in Shanghai, July 2016.

4. Determinants of Station Service Cost in Metros presented at *Transportation Research Board* - Lectern Session 16-2142 on *Catchment Areas, Connectivity, Crime Avoidance, Costs, and Capacity: 5 C's of Passenger Station Planning* in Washington D.C. in January 2016.

5. Determinants of Metro Operating Costs: An Econometric Analysis at the *Workshop on the economics and econometrics of transit modes* in the École Nationale des Ponts et Chaussées (ENPC) in Paris, December 2015.

6. Econometric benchmarking of metro operating costs: methods and applications presented at the *European Transport Conference* - Seminar Session on *Local Public Transport* in Frankfurt, September 2015.

7. Determinants of Train Service Costs in Metro Operations presented at *Transportation Research Board* - Lectern Session 15-4287 on *Rail Transit Efficiency and Congestion Management* in Washington D.C. in January 2015.

8. Determinants of Rolling Stock Maintenance Cost in Metros presented at *Transportation Research Board* - Lectern Session 15-4269 on *Innovations in the Passenger Rail Equipment Industry* in Washington D.C. in January 2015.

I would like to express my thanks for the feedback I received from the reviewers of my publications and conference presentations. Their comments have helped me to improve the final version of this thesis.

2 — Literature Review

This section reviews the main academic contributions to the literature on railway cost efficiency. Firstly, section 2.1 introduces the concept of cost efficiency and how to measure efficiency in the rail sector. Secondly, section 2.2 establishes the scope of the literature review, after which section 2.3 reviews the main contributions of the literature. Lastly, the conclusions in section 2.4 describe where this research stands with respect to the previous literature and the main academic contribution to the field.

2.1 Concept of Cost Efficiency

The concept of efficiency has been central to economic research over time, as it relates to the ability of a certain firm or industry to increase output without drawing on more resources. The contemporary research on efficiency starts with the work by Farrell (1957) on the measurement of productive efficiency. Until his work appeared, the efficiency was measured by more or less intuitive definitions of efficiency, whereby a firm was deemed to be efficient if increasing the output of a product would entail the decrease in the output of another product (Koopmans et al., 1951). Likewise, there were measures of partial

productivity of factors (e.g. coefficient of resource utilization by Debreu, 1951) which are close to current KPI schemes. However, despite partial efficiency measurements and general benchmarking can be very useful and they are still being used for efficiency assessment by regulators (e.g. Office of Rail and Road in the United Kingdom as shown in Smith et al., 2010), this literature review will focus on econometric methods to assess rail efficiency.

A common approach to assess railway cost efficiency is by means of cost functions. A cost function is a mathematical function that shows how costs (i.e. C) vary depending on several factors, usually output (Q) and input prices. For instance, the cost function in Equation 2.1 depends on the level of output (Q) and the input prices; wages w for labour L and rent r for capital K. Likewise, the input quantities (L and K) are also a function of the quantity produced (Q) and the relative prices of the inputs (w and r).

$$C(Q, w, r) = wL(Q, w, r) + rK(Q, w, r) \tag{2.1}$$

The description of rail costs by means of cost functions usually suits the operating behaviour of railway operators where the level of output is set by regulators and most railways operate as cost-minimizing firms (Nash and Smith, 2007).

2.2 Scope of the Review

This section reviews the main contributions to the field of urban rail efficiency, primarily based on cost functions. However, given that the existing

literature on urban rail cost efficiency is scarce, this review also includes the most important research on cost efficiency for the rail sector.

The research on transport cost efficiency is relatively abundant and comprises other modes of transport beyond rail such as urban bus (e.g. Farsi et al., 2006, Iseki, 2010), air transport for both airports (e.g. Barros, 2008, Liebert and Niemeier, 2013, Oum et al., 2003, Tovar and Martín-Cejas, 2010) and airlines (e.g. Barbot et al., 2008, Merkert and Hensher, 2011, Ouellette et al., 2010). Likewise, the research on maritime transport and port cost efficiency is also profuse (e.g. Clark et al., 2004, Cullinane et al., 2006, Wanke et al., 2011). Although the research on the cost efficiency of other transport modes has been considered, it has only been cited occasionally, when there has been a substantial contribution in terms of new approaches or methods to assess cost efficiency (e.g. SFA approach in Holmgren, 2013, Karlaftis and Tsamboulas, 2012).

The following literature review is limited to railway, and particularly urban rail, transport cost efficiency. This is due to two main reasons. Firstly, railway cost efficiency literature already constitutes a large corpus of academic research which embodies the main theoretical and methodological advances in the field of cost efficiency. Adding the research on cost efficiency for other modes of transport would be mainly repetitive. Secondly, urban rail and railway operations are operationally similar, as they both entail a homologous asset structure, with a large share of fixed assets, and presumably comparable economies of density and scale. However, other modes of transport or other transport operations can be substantially different. Some operators do not have the same fixed assets with large fixed costs (e.g. bus

operators) whereas other transport operators have different output that is not comparable to the railway output (e.g. airport cost efficiency). For these two reasons, the cost efficiency research of other transport modes has not been cited except in the case of original major contributions. A comprehensive review of transport cost research can be found in Oum and Waters II (1996) and De Palma et al. (2011).

Regarding the rail sector, Oum et al. (1999) cover a wide selection of studies estimating the productivity and efficiency of the rail sector by means of index, data envelopment analysis and other econometric methods. Section 2.3 follows its example as it deals with the academic contributions for each method. The particular details of the methods presented are discussed in Section 3.

2.3 Approaches to Assess Rail Cost Efficiency

2.3.1 Total factor productivity

In terms of cost function research on railway efficiency, the seminal study by Caves, Christensen and Swanson (1980) on U.S. railroad costs for the period between 1951 and 1974 is a fundamental one. This study was based on the flexible cost function for multiproduct firms developed previously by Caves, Christensen and Tretheway (1980), which enabled the simultaneous consideration of multiple inputs and outputs. The contribution of Caves, Christensen and Tretheway (1980) expanded on the multiproduct cost functions (MCF) developed by (Diewert, 1971) which, although they set no restrictions on the

elasticities of substitution between input factors, still implied constant returns to scale between costs and output levels. Therefore, Caves, Christensen and Tretheway (1980) developed a generalised translog (i.e. transcendental logarithmic) multiproduct cost function which enabled the MCF to present non-constant returns to scale in the relationship between total costs and output levels. Likewise, it also allowed for zero values for one or more outputs of the model, meaning the sample could include rail companies with freight service, passenger services or both.

Even though the initial study by Caves, Christensen and Tretheway (1980) included an initial application of the generalised translog multiproduct cost function to a set of 41 railway firms, it was the following study (Caves, Christensen and Swanson, 1980) which proved to be more influential in academic railway research. In this later study, Caves, Christensen and Swanson (1980) estimated a translog cost function while considering the multiproduct nature of the railway firms. Consequently, the study included labour, track, equipment and fuel as inputs, but it also considered several variables as outputs: ton-miles, average length of haul, passenger-miles and average length of trip. This multi-input and multi-output approach was applied again (Caves et al., 1981*b*) to assess the productivity growth of US railways (1955-1974) and the existence of economies of scale. The productivity growth of the US railways was found to increase around 2%-3% annually despite the estimates based on variable costs were higher than the total cost estimates. Regarding the economies of scale and considering the total and variable cost models, Caves et al. (1981*b*) found evidence of strong, statistically significant economies of scale if output increase came from larger trip

lengths or increased haul. However, economies of scale are lower and statistically insignificant if the increase in output is derived from an increase in the length of haul.

One of the main considerations to bear in mind when assessing the economies of scale in the rail sector is the network characteristics of the production of the output. As the output is produced in a network, an increase in output may be due to an increase in the traffic over a fixed network or it could also be due to the expansion of the network itself while keeping output per network kilometre constant. In the first case, the efficiency gains derived from the increase of output production on a fixed network would be considered economies of density, and not economies of scale. This characteristic was already assessed by Keeler (1974) who estimated a model and included the network length as a variable. Keeler (1974) found substantial unexploited economies of traffic density for most railroads but constant long-run returns to scale. Therefore, due to the economies of traffic density, increasing capacity in the same network led to less than proportional increases in costs.

These findings have been corroborated by later research. For instance, McGeehan (1993) determined the cost structure of Irish railways 1973-1983 following the method of Caves et al. (1981*b*). The results indicated substantial economies of density, and, when given a fixed network capacity, unit costs increased less than proportionately as output increased.

Another significant contribution to the assessment of railway productivity was developed by Cantos et al. (1999) with the analysis of a sample of 17 European rail operators for the period 1970-1995. The main contribution of the research was not only comparing the evolution of the total factor productiv-

ity growth, which had already been done for the US-Canada railways (Caves et al., 1981*a*), but also to decompose the productivity changes. In this case, Cantos et al. (1999) broke down the total factor productivity (TFP) growth into two components; efficiency change and technological progress. The main contribution to the TFP growth in the European railways was the technological change, particularly for the 1985-1995 period. Besides this, Cantos et al. (1999) corroborated previous findings (Hensher et al., 1995, Oum and Yu, 1994) and concluded that electrification of the network, managerial autonomy and lower subsidies were also correlated with higher railway efficiency. Moreover, there was evidence that higher managerial autonomy and lower subsidies also led to more technological progress. These findings were in line with both previous (Caves et al., 1981*a*) and following research (Laurin and Brozec, 2001) on how regulation influences railway productivity.

Aside from the seminal papers using TFP on railway efficiency, and the abundant application of their techniques to other cases, until recently there have been few papers using the TFP approach on urban rail systems (Mackett and Sutcliffe, 2003). Sometimes the analysis of urban rail systems is based on case studies, where there is no comparison between different metro systems. Likewise, other studies are qualitative without entailing econometric estimations (Babalik-Sutcliffe, 2002, Sampaio et al., 2008) due to data unavailability.

However, there are still some recent urban rail contributions. Graham (2008) analysed the efficiency of a sample of 89 urban railways worldwide for the year 1996. Graham (2008) estimated a translog production function finding that urban rail operation type (i.e. underground/subsurface), city

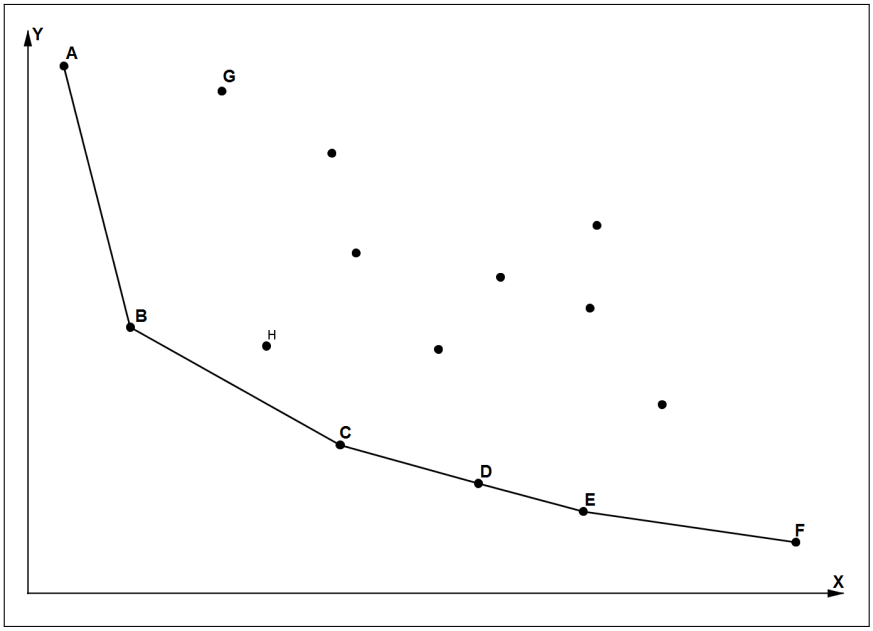
economic development and population density positively determined urban rail productivity. Besides this, the TFP was decomposed in order to identify the effect of firm-specific technology. The TFP also allowed for an assessment of the effect of non-constant returns to scale; however, it was concluded that operators face constant returns to scale but increasing returns to density. This results corroborated previous research (Graham et al., 2003*a*) on 17 urban rail metros which were characterised by constant returns to scale and also significant returns to density due to the large fixed components of urban rail costs.

2.3.2 Nonparametric: data envelopment analysis

The main nonparametric approach is the data envelopment analysis (DEA). The DEA is a deterministic method based on computing a production frontier by means of linear programming techniques. It is deterministic, as all the variation from the data is assumed to contain information about the rail operator, meaning no random noise is incorporated in the estimation. Regarding the frontier, it is essentially formed by wrapping a hull around the observed data (Greene, 2008) whereby the firms would be located at varying distances from the frontier. Those closer to the frontier would be more efficient, as the frontier itself is formed by the best practice that is observed overall. Likewise, the larger the number of observations, presumably the more accurate the calculation of the production frontier would be, as the frontier is calculated based on the firms that are being observed. A simplified example of the DEA methodology can be seen at Figure 2.1 and a detailed

explanation of the method is offered in section 3.1.1 on data envelopment analysis.

Figure 2.1: DEA example: production frontier with inputs X and Y.



The first applications of DEA to the study of efficiency were formulated by Charnes et al. (1978) in order to assess the efficiency of non-profit organisations and it has substantially evolved since then (see a review in Seiford, 1996, Tsai et al., 2015). The DEA approach was introduced in the railway sector by a piece of research on the effects of government intervention and subsidisation on railway efficiency (Oum and Yu, 1994). The research studied 19 European and Japanese OECD railways focused on passenger rail services for the period between 1978 and 1989. Oum and Yu (1994) estimated a gross efficiency index from the panel data with a DEA approach and, secondly, they regressed the efficiency index on a set of operating characteristics such as subsidy levels, managerial autonomy, average load per train, traffic

density and percentage of electrified network. As the efficiency index is constrained between 0 and 1, the second regression is constrained to lay within those boundaries. The results found that systems with high managerial autonomy and lower subsidies tend to be more efficient. Besides this, Oum and Yu (1994) also reported that the effect of the variables on the efficiency of the railways depends on the output considered. When the output considered is the passenger kilometres, high passenger train loads tend to increase railway efficiency. However, if the output is measured as train kilometres, high passenger loads tend to decrease efficiency. Likewise, there are also variables that tend to increase overall efficiency such as electrification of the network, the managerial autonomy and the lower levels of subsidy. More importantly, Oum and Yu (1994) noticed that the efficiency measures are not an accurate measurement of efficiency if the estimations do not control for differences in the operating and market environment.

Other research on European railway efficiency (Cowie and Riddington, 1996) compared the differences between the efficiency scores obtained by the DEA method and the corrected ordinary least squares (COLS, explained in section 3.2.1). By means of comparing different measurements of efficiency, Cowie and Riddington (1996) concluded that no accurate measurement of efficiency is possible and advised against the use of translog functions to assess railway efficiency when the estimates are not reconciled with the underlying theoretical framework. Cowie (1999) also studied the performance of 57 public and private Swiss railways to estimate the effect of ownership on the performance of the operators. The DEA results suggested that private operators tend to have higher technical and managerial efficiency which

corroborated previous research on ownership, managerial autonomy and efficiency (e.g. Oum and Yu, 1994).

An article by Cantos et al. (2002) expanded the applications of the DEA analysis of railway cost and revenue efficiency by means of including input price data. The availability of input price data enabled the researcher to break down the railway efficiency into technical efficiency, allocative efficiency and a scale component. The initial distinction between the two concepts was made by Farrell (1957), where technical efficiency refers to the relationship between the observed firm production and the maximum ideal or potential production. On the contrary, allocative efficiency refers to the circumstance where, given certain input prices, the costs cannot be minimised further by a change in the input quantities to achieve a certain output. According to Cantos et al. (2002), the cost and revenue efficiency of the European railways studied remained constant for the 1985-1995 period. However, by decomposing the inefficiencies, it was found that the technical inefficiencies are the most important. Likewise, Cantos et al. (2002) also found that the most cost-efficient firms also tend to be the most revenue-efficient. Lastly, Cantos et al. (2002) corroborated the findings by Oum and Yu (1994) and their previous research (Cantos et al., 1999), where greater autonomy and financial independence of management is linked to greater cost and revenue efficiency.

The group of seminal contributions to the development of DEA for the assessment of rail efficiency was focused on railway firms and not on urban rail systems. However, lately there have been an increasing number of contributions for urban rail systems. Graham (2008), in the same study of 89

urban railways previously mentioned, also estimated DEA models on the efficiency of the urban rail systems under the assumption of constant (CRS) and variable returns to scale (VRS). Graham (2008) compared both DEA estimates to the previous TFP results and found a consistent pattern in the match of both approaches. The comparisons of CRS DEA with TFP and the VRS DEA with TFP comparisons follow an exponential pattern which explained 75% of the variation with respect to the TFP estimates.

The DEA has also been applied to analyse how the ownership structure can influence the technical efficiency of 15 urban rail systems (Jain et al., 2008). Jain et al. (2008) found higher DEA efficiency estimates for private rather than for public metros, implying higher efficiency in private systems, despite not having controlled for other determinants of efficiency that may have been causing the difference in efficiency estimates.

More recently, Santos et al. (2013) assessed the performance of 37 European metros by means of performance indicators as well as DEA estimates on efficiency. The first DEA estimations used the network length (i.e. total km of track), the number of vehicles and the total number of employees as inputs. The outputs considered were the passengers transported and the number of vehicles per kilometre. The estimates on efficiency were firstly obtained by means of a DEA-bootstrap approach where the efficiency estimate is obtained by comparing the bootstrap frontier to the original input, in accordance with Simar and Wilson (2000). Secondly, these efficiency estimates are regressed to estimate the influence of operational and environmental variables. These environmental variables are the number of stations, the network length (i.e. network kilometres), the GDP per capita of the city and the population. It

was found that both higher GDP per capita and larger populations tend to increase the metro efficiency as in Graham (2008). On the contrary, a larger number of stations and longer networks tend to decrease efficiency.

Lastly, Tsai et al. (2015) analysed the determinants of technical, allocative and cost efficiency for a set of 20 urban rail systems in the 2009-2011 period. The analysis was performed with a two-stage approach. The first stage entails DEA models determining the efficiency of the urban systems in the sample. The linear programming-based estimation of the efficiency considers several inputs and outputs and it is assumed that the rail operators try to minimise the inputs for a given level of output; these results are also bootstrapped following the Simar and Wilson (2000) approach in order to smooth and correct the estimates. At the same time, the ratio of the service output to inputs is not held constant as it can vary with the size of the firm, effectively allowing for non-constant returns to scale. In a second stage, the efficiency estimates were regressed on population density, number of stations, average distance between stations and type of rail operations (i.e. heavy versus light rail operations). Tsai et al. (2015) found that urban rail systems with high efficiency are also those with high demand levels measured as passengers per network length. Besides this, the system efficiency increases with the network length and the number of stations. Heavy rail usually leads to poorer efficiency levels whereas the population density increases cost efficiency significantly.

2.3.3 Parametric: deterministic and stochastic

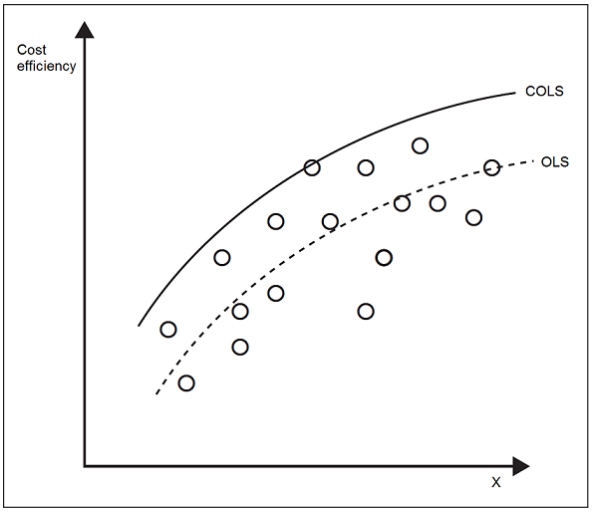
Parametric methods are characterised by the fact that the models assume a certain production or cost function type. The parametric approach can be broken down into two main groups depending on how the data variation is interpreted by the models; deterministic or stochastic approaches.

Deterministic econometric methods

Urban rail cost efficiency has also been studied by means of parametric deterministic methods, where all possible data noise is suppressed and any data variation is interpreted as rail efficiency variation. The simplest approach in this category is the corrected ordinary least squares (COLS) approach as shown in Figure 2.2. In this case, the regression line is shifted to cross the observation of the rail operator for which the costs have been overestimated the most; in other terms, the frontier line is shifted to cross the observation with the largest negative residual in absolute terms.

One of the first applications in the railway sector was made by Perelman and Pestieau (1988) to assess the efficiency of 19 rail operators during the period 1970-1983. The estimation of efficiency followed a two step approach. Firstly, an index of efficiency was estimated with a COLS model and secondly, the indexes were regressed on operational variables. The operational variables included network length per square kilometre, percentage of electrified lines, track kilometres per network length, average passenger journey and average haul length. Deprins and Simar (1989) developed a similar study on passenger and freight railway operators for the period 1970-1983. The re-

Figure 2.2: Comparison of OLS and COLS. Source: Adapted version from Bogetoft and Otto (2010, pg. 202).



search confirmed the results by Perelman and Pestieau (1988) and they found that higher average number of track by lines, more electrification of the lines and higher ratios of passengers per train tend to be related to higher rail efficiency. On the contrary, low density networks measured as the kilometres of line per hundred square kilometres, are prone to lower rail efficiency.

Coelli and Perelman (1996) continued the research comparing the estimates on railway efficiency obtained by means of COLS and stochastic frontier analysis (SFA) for a panel of 17 European railways in the period 1979-1983. Coelli and Perelman (1996) included multiple inputs and outputs in the estimation of the efficiency estimates. They concluded that the COLS approach is more convenient as the SFA efficiency estimations are not reliable under large variability of the data sample and flexible functional forms (i.e. translog). However, Coelli and Perelman (1999) also found that the COLS and DEA relative efficiency rankings were very similar, although, unfortu-

nately, the efficiency index estimates were not regressed with any operational variable to explain performance.

Gathon and Perelman (1992) assessed railway efficiency for a set of 19 European railways for the period between 1961 and 1988 by means of panel data models (i.e. random effects and fixed effects). In this case, the individual effects of the rail firms were considered the indicators of rail efficiency. Unfortunately, the explanatory variables included in the model are scarce; only the percentage of network electrification and an index of managerial autonomy were used. The research found that management autonomy and electrification increased efficiency whereas higher loading reduced rail efficiency. These conclusions have been reinforced by later research (e.g. Cantos et al., 1999, Oum and Yu, 1994, Sanchez and Villaroya, 2000) and they were also corroborated again (Gathon and Pestieau, 1995) by means of a Displaced Ordinary Least Squares (DOLS) approach. Gathon and Pestieau (1995) used the same panel of railways (Gathon and Perelman, 1992) but they also included explanatory variables on population density and hilliness of each country which ultimately were statistically insignificant.

Another deterministic measurement of railway efficiency (Filippini and Maggi, 1993) estimated a translog cost function which also controlled for the economies of density and scale. The research used a panel of 48 Swiss railway companies over 4 years in order to estimate the effect of subsidies and ownership. The results showed that ownership had no influence on efficiency and that subsidy levels were positively correlated to higher cost efficiency, which contradicts subsequent research on the effect of both factors (Oum and Yu, 1994). Filippini and Maggi (1993) also found significant increasing

returns to density in railway costs which corroborated previous research on railway cost efficiency (Braeutigam et al., 1982, Friedlaender et al., 1993, Harmatuck, 1979).

More recently, Farsi et al. (2005) used an unbalanced panel of 50 small Swiss railway companies from 1985 to 1997, resulting in over 600 observations in total, to measure cost efficiency in network industries. The estimation method was based on a conventional fixed and random effects panel data model. However, they introduced a variation following Greene's (2004) true random effects models. In particular, the models assumed a constant firm-specific heterogeneity but allowed for variation, both over time and between cross units, in the sources of inefficiency. Therefore, the random effects model combined the conventional random effect approach with a skewed stochastic term representing inefficiency. The specification of the model followed a Cobb-Douglas functional form which eased the estimation of the function but also reduced the flexibility (e.g. scale economies were assumed to be a fixed relationship regardless of the output level). Overall, Farsi et al. (2005) estimated five models. The first three were a fixed effects, a random effects and a cross-sectional pooled frontier model disregarding the panel dimension. The fourth model was an extension of the third with an additional firm-specific random effect. Lastly, the fifth model assumed that the unobserved cost differences between the firms were constant over time and were determined by the network-related unobserved characteristics rather than inefficiency. Besides the comparison of efficiency estimates, which according to the authors remained inconclusive, Farsi et al. (2005) also estimated economies of scale and density, finding significant economies of both types

with all the methodologies. They also found significant variations in their estimates which related to the different treatment of the unobserved network effects in the different models.

With regard to urban rail systems, parametric deterministic estimation methods have not been used to assess urban rail cost efficiency. However, there has been recent nonparametric deterministic research on urban rail efficiency (Tsai et al., 2015).

Stochastic econometric methods

There are also stochastic parametric methods, such as the stochastic frontier analysis (SFA), that have been used to assess rail efficiency. The first application (Kumbhakar, 1988) estimated the technical and allocative efficiency for a panel of 13 US railroads for the period 1951-1975. Furthermore, the input-specific technical-inefficiency functional form used by Kumbhakar (1988) did not assume that inefficiency caused a equiproportional reduction of the productivity of all the factors, as in earlier SFA models (Aigner et al., 1977). However, Kumbhakar (1988) not only estimated which inputs are inefficient (Lovell and Sickles, 1983) but also disaggregated the effect of each input on inefficiency. This enabled Kumbhakar (1988) to find that labour caused 80% of the overall technical inefficiency and capital and fuel were responsible for the remaining 20%. Likewise, Kumbhakar (1988) also estimated a 5.5% cost of the allocative inefficiency, which makes reference to the fact that costs could be minimised further by changing the input quantities to achieve a particular output.

Parisio (1999) also used the SFA approach to estimate technical and allocative inefficiencies in a panel of 8 European railways for the period 1973-1989 by means of a translog cost system which modelled the minimal cost. The error terms of the model were used to estimate the allocative inefficiencies, which are a function of cost share residuals and other parameters of the model. Parisio (1999) found negligible allocative inefficiencies, implying that the mix of inputs was close to the cost-minimising optimal choice. On the contrary, technical inefficiency was significant with 1.2% to 3.7% estimated over cost. This implied that railway efficiency could be improved by reducing the inputs committed to production (e.g. lowering service in less profitable branches of the network) rather than by adjusting the production factor shares.

Cantos and Maudos (2000) used a stochastic frontier cost function to estimate the efficiency and productivity levels of 15 European railways for the period 1970-1990. Cantos and Maudos (2000) followed a similar approach to Gathon and Pestieau (1995), except that they estimated a function of the railway operating costs with an SFA approach. The results showed that technical progress was the leading force of railway productivity growth in Europe between 1970 and 1990 whereas the efficiency levels remained constant. Likewise, economies of scale were found to have a negligible role in the evolution of railway productivity, which corresponds with previous research on mild returns to scale (e.g. De Borger, 1992*a*, Keeler, 1974, Kim, 1987). Finally, Cantos and Maudos (2000) also found that managerial independence and finance autonomy are linked to higher railway efficiency, corroborating previous results by Oum and Yu (1994) and Gathon and Pestieau (1995)

among others.

A more recent SFA example on railway efficiency (Couto and Graham, 2008) estimated the efficiency of 27 European railways for the period 1972-1999. Following previous research (e.g. Parisio, 1999), Couto and Graham (2008) estimated a translog cost system distinguishing between technical and allocative efficiency. Allocative efficiency was obtained by comparing cost share equations with cost functions, where the existence of differences between the two suggested the existence of allocative inefficiencies. Overall, the allocative inefficiencies accounted for an average 7% increase in variable cost, whereas technical inefficiency was responsible for a 6.5% excess cost.

Even though there have been several applications to railways, the applications of this method to urban rail systems have been limited to technical efficiency research (Canavan et al., 2015). Other research on urban cost efficiency with SFA methods has been focused on urban transit agencies (Holmgren, 2013, Karlaftis and Tsamboulas, 2012) and, to the best of my knowledge, not to urban rail operating cost efficiency.

2.4 Literature Review Conclusions

This section reviews the abundant corpus of research on rail efficiency and, particularly, cost efficiency (see also Table 2.1 on page 58). The studies reviewed show an evolution in both methodological and argumentative terms. Firstly, the methods have evolved from the early total factor productivity approaches and data envelopment analysis towards parametric methods, such as panel regression and stochastic frontier analysis. Besides this, the argu-

ments and hypotheses tested by the research have also evolved as more data from the railway operations has become available.

However, the existing corpus of research is still meagre in the case of urban rail efficiency. There are few significant studies on urban rail cost efficiency and almost none using parametric methods. Besides this, the variables used for the assessment of rail efficiency, and urban rail efficiency in particular, have been limited due to lack of data availability. Some explanatory variables that are easily available have been tested repeatedly in the last few decades (e.g. ownership, subsidy levels, network length) but other variables more influential on rail cost efficiency have not been considered in the literature (e.g. outsourcing of labour, rolling stock characteristics, driver productivity).

Besides the lack of variety in the explanatory variables, there is also a lack of variety in the dependent variables. For instance, there is plenty of evidence on total variable costs and total cost efficiency, but there is no evidence of the cost efficiency of the different operational dimensions of a railway. For example, the findings on economies of density and economies of scale always refer to aggregate measures of costs, whereas these economies may be very different depending on the operational dimension considered. Thus, assessing the economies of density in the station service costs and the economies of scale in the rolling stock maintenance costs may yield different estimates to those aggregate measurements estimated to date.

Consequently, given the gaps identified in this literature review, this research aims to make the following academic contributions:

1. Expand the current literature on rail cost efficiency by measuring the

cost efficiency of urban rail systems.

2. Use advanced analytical methods, both parametric and stochastic, which have been seldom used in the assessment of urban rail efficiency.

3. Test the effect of new explanatory variables on rail efficiency that have not been tested before due to data unavailability.

4. Assess the cost efficiency for the different sections of operations (e.g. train service, station service, rolling stock maintenance) and compare how the different cost factors (e.g. economies of density or scale) compare to the aggregate estimates in the literature.

Table 2.1: Main rail cost efficiency research

Author(s)	**Dependent Variable**	**Functional Form**	**Technique**	**Data**	**Country**	**Relevant Findings**
Cantos and Maudos (2000)	Total Cost	Translog	SFA	1970-1990	Europe	Negligible RTS. Financial autonomy and managerial independence increase efficiency.
Cantos et al. (1999)	MPI	Non-Param.	DEA	1970-1995	Europe	Managerial autonomy and less subsidisation lead to higher efficiency.
Cantos et al. (2002)	Total Cost	Non-Param.	DEA	1985-1995	Europe	Decompose inefficiencies into technical, allocative and scale. Technical inefficiencies are the most important.
Caves et al. (1980)	Total and Variable Cost	MCF	Determ.	1951-1974	US	Productivity growth trend over time.
Caves et al. (1981*b*)	Variable Cost	Translog	Param. and Determ.	1955-1974	US	Negligible RTS and low capacity utilisation.
Coelli and Perelman (1999)	Multi-output	Translog	COLS and SFA	1979-1983	Europe	COLS preferred to SFA when using translog and large variability of data sample.
Couto and Graham (2008)	Total Cost	Translog	SFA	1972-1999	Europe	Allocative and technical inefficiencies increase costs by 7% and 6.5%.

Table 2.1 – *Continued from previous page*

Author(s)	**Dependent Variable**	**Functional Form**	**Technique**	**Data**	**Country**	**Relevant Findings**
Cowie (1999)	Output (train km)	Non-Param.	DEA	1990-1997	Switzerland	Higher efficiency with private operators and managerial autonomy.
Cowie and Riddington (1996)	Output (pax km)	Translog	DEA/COLS	1992	Europe	Unsuitability of translog functions for efficiency analysis.
Deprins and Simar (1989)	Technical efficiency	Cobb-Douglas	COLS	1970-1983	World	Low-density networks lead to lower efficiency.
Farsi et al. (2005)	Total Cost	Cobb-Douglas	Pooled-OLS, RE, FE and True RE	1985-1997	Switzerland	Significant RTD and RTS. Inconclusive comparison of efficiency estimates between different econometric approaches.
Filippini and Maggi (1993)	Total Cost	Translog	DOLS	1985-1988	Switzerland	Null effect of ownership on efficiency. Higher subsidy improves cost efficiency.
Gathon and Pestieau (1995)	Labour required	Param.	RE and FE	1961-1988	Europe	Management autonomy and electrification increase efficiency output per labour unit.
Gathon and Pestieau (1995)	Output (passenger and freight)	Translog	DOLS	1961-1988	Europe	Population density and hilliness of the country have negligible effects on efficiency.

Table 2.1 – *Continued from previous page*

Author(s)	**Dependent Variable**	**Functional Form**	**Technique**	**Data**	**Country**	**Relevant Findings**
Graham (2008)	Output (Car km)	Translog PF	DEA/TFP	1996	World	RTD and null RTS. Urban density and surface operations increase efficiency.
Graham et al. (2003*b*)	Output (pax. Journeys)	Cobb-Douglas	Param. and Determ.	1996	World	Null RTS but high RTD. Effect of RTD on TFP and labour productivity.
Hensher et al. (1995)	TFP Index	Param.-TFP	Param. and Determ.	1971-1992	Australia	Substantial RTD and managerial role important for increasing efficiency.
Jain et al. (2008)	Output (Car km)	Non-Param.	DEA	1992-2002	World	Higher efficiency in private metros.
Keeler (1974)	Short-run Total Cost	Cobb-Douglas	Param. and Determ.	1969-1971	US	Strong RTD and null RTS.
Kumbhakar (1988)	Variable Cost	Param.	SFA	1951-1975	US	Decomposes rail efficiency into input and technical inefficiency. Labour identified as the source of most of rail inefficiency.
McGeehan (1993)	Variable Cost	Translog	Param. and Determ.	1973-1983	Ireland	Strong RTD. Railways cost function proven to be not homothetic and not homogeneous.

Table 2.1 – *Continued from previous page*

Author(s)	**Dependent Variable**	**Functional Form**	**Technique**	**Data**	**Country**	**Relevant Findings**
Oum and Yu (1994)	Efficiency Indices (DEA)	Non-Param.	DEA	1978-1989	OECD	Electrification, lower subsidies and managerial autonomy increase efficiency. Conclusions depend on output.
Parisio (1999)	Total Cost	Translog	SFA	1973-1989	Europe	Negligible allocative and significant technical inefficiencies.
Perelman and Pestieau (1988)	Efficiency Index	Cobb-Douglas	COLS	1970-1983	World	Electrification and higher demand lead to higher efficiency.
Sanchez and Villaroya (2000)	Total Cost	Translog	SFA	1970-1990	Europe	Technical progress causes main productivity gains. Management autonomy in decision making raises efficiency.
Santos et al. (2013)	Output (pax journeys)	Non-Param.	DEA	2006	Europe	Higher GDP per capita and population lead to more efficiency. Larger networks and more stations lead to lower efficiency.
Savage (1997)	Long-run cost	translog	3-stage Least Squares	1985-1991	US	Large RTS, particularly for maintenance and long-run costs.

Table 2.1 – *Continued from previous page*

Author(s)	**Dependent Variable**	**Functional Form**	**Technique**	**Data**	**Country**	**Relevant Findings**
Tsai et al. (2015)	DEA/Indices Efficiency	Non-Param.	DEA	2009-2011	World	Population density, network size and stations increase efficiency.
Viton (1980)	Efficiency Index	Translog	Frontier	1960-1970	US & Canada	RTD in small operators and negative RTD for big ones. Small sample.
Viton (1993)	Short-run costs	Translog	Frontier	1960-1970	US & Canada	Operators already efficient. Negative RTD for all operators. Small sample (72 obs).

Note: See list of acronyms and definitions in page 18.

3 — Methodology for Cost Efficiency Estimation

This section reviews the econometric techniques that have been applied to estimate railway cost efficiency in the academic literature. The ever-growing range of methods used in the literature on efficiency estimation tend to be broken down into four categories as shown in Table 3.1 (based on Bogetoft and Otto, 2010, Coelli et al., 2005, Murillo-Zamorano, 2004).

Table 3.1: Cost efficiency estimation methods and examples.

Methods	**Deterministic**	**Stochastic**
Parametric	Corrected Ordinary Least Squares (COLS)	Stochastic Frontier Analysis (SFA)
Non parametric	Data Envelopment Analysis (DEA)	Stochastic Data Envelopment Analysis (SDEA)

These four categories are defined depending on two criteria. The first criteria is based on how data deviations explain changes in efficiency. In deterministic methods, all the deviations of the data from the model estimates are interpreted as pure efficiency or inefficiencies, whereas stochastic methods allow for data randomness and noise. Secondly, Table 3.1 breaks down the methods into parametric, i.e. models with predetermined parame-

ters (e.g. production function), and nonparametric, i.e. methods with fewer assumptions on the functional form of the models, which let the data define the shape of the estimated functions.

This section excludes the basic approaches to measuring railway efficiency such as partial productivity ratios, even though they are still applied by rail regulators and rail firms (e.g. Office of Rail and Road in the United Kingdom as shown in Smith et al., 2010). These ratios relating inputs to outputs (e.g. train kilometres / staff hours) are easily understandable and can yield insights into rail operations. However, the ratios ignore the impact of substitution effects between inputs, they may be biased if outsourcing activities and inputs are not included and they also assume a perfect technical efficiency. Besides this, rail operations have multiple outputs that can be produced simultaneously (e.g. train kilometres and passenger kilometres), thus the partial productivity ratios always yield an incomplete description of the operator efficiency (Nash and Smith, 2007).

Sometimes different inputs and outputs can be combined to give a holistic assessment of the rail performance (Denny et al., 1981). In this case, the ratio is formed by all inputs and all outputs, with some weighting for the different inputs and outputs. Although this weighted productivity ratio may improve partial productivity ratios, it still suffers from the same drawbacks, so it is not included in the description of cost efficiency methodologies.

3.1 Nonparametric Frontier Methods

3.1.1 Data envelopment analysis

The data envelopment analysis (DEA) has been one of the most widely used approaches for assessing the efficiency of all types of organisations. The DEA method was initially proposed by Charnes et al. (1978, 1979) and has developed significantly since then. A comprehensive survey on DEA applications can be found in Tavares (2002) while additional details and applications on the DEA approach are also available (Charnes, 1994, Cooper et al., 2007, Rao et al., 1998).

DEA is not deterministic so it entails the estimation of a technology function and, depending on this function, the calculation of organisational or firm efficiency. A production function defines how a firm turns sets of inputs into certain outputs. In many cases, the production function is determined *ex-ante* like, for instance, a Cobb-Douglas function $Y = AL^aK^b$ where Y refers to production output, L and K to production inputs, a and b to the corresponding input production elasticities and A to a general measure of productivity. In the case of the DEA, the whole function is dependent on the data and the relation between inputs and outputs is estimated as a first step of the DEA approach.

The determination of the production function is essential, as it establishes the combinations of outcomes that are feasible given certain inputs and vice versa. The frontier between the feasible and the unfeasible production sets defines an optimal production frontier to which all the firms are compared

in order to assess their efficiency.

The estimation of the production function is made by means of mathematical programming and, the resulting DEA production function usually has to comply with the following economic conditions (γ):

Disposability of inputs and outputs The marginal product of inputs cannot be negative. Thus, a certain output can always be achieved with a larger input endowment. Given a set of inputs, x, and outputs, y, where T represents the production function, which determines its feasible production set, then the free availability of inputs is

$$(x, y) \in T, x' \geq x \Rightarrow (x', y) \in T \tag{3.1}$$

whereas the free availability of outputs equals

$$(x, y) \in T, y' \leq y \Rightarrow (x, y') \in T \tag{3.2}$$

This can be combined to form

$$(x, y) \in T, x' \geq x, y' \leq y \Rightarrow (x', y') \in T \tag{3.3}$$

Convexity The average of two observations on the frontier are always included in the feasible production set. Therefore, if (x^0, y^0) and (x^1, y^1)

are included in T, and $0 \leq \lambda \leq 1$, then it is also true that

$$(1-\lambda)(x^0, y^0) + \lambda(x^1, y^1) \in T \tag{3.4}$$

Scalability The output can be scaled up or scaled down.

$$(x, y) \in T, 0 \leq \lambda \Rightarrow \lambda(x, y) \in T \tag{3.5}$$

However, the output may increase less, more or in the same proportion as the change in inputs when $0 \leq \lambda \leq 1$, $1 \leq \lambda$ or $\lambda = 1$ respectively.

Additivity The addition of feasible observations is also feasible.

$$(x, y) \in T, (x', y') \in T \Rightarrow (x + x', y + y') \in T \tag{3.6}$$

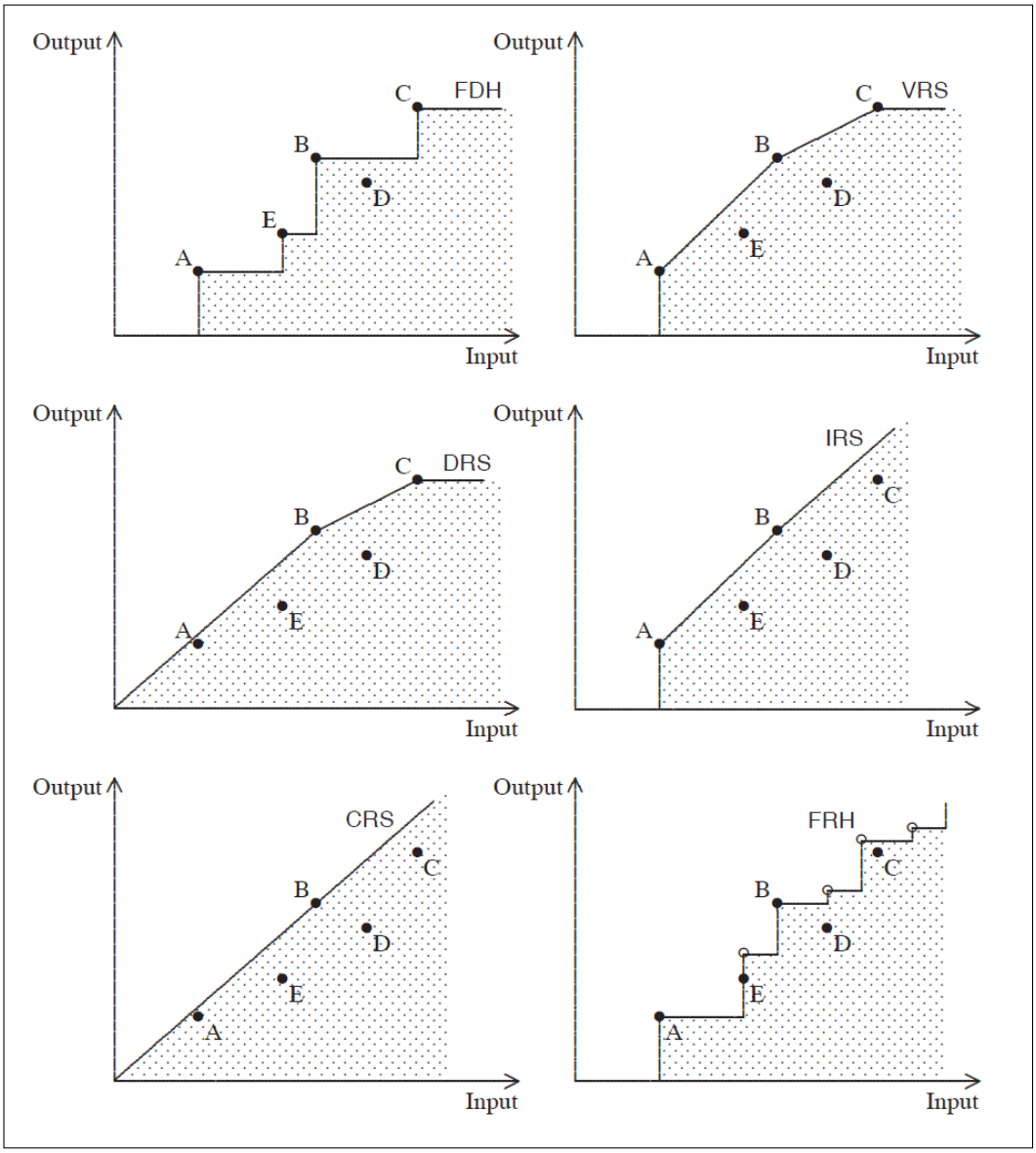
Besides complying with these assumptions that we can identify as γ, the estimation of the production function should also follow the minimal extrapolation principle. This means that the estimated production function should contain the smallest set of feasible outputs that include all the observations of the sample.

$$T^*(\gamma) = \{(x, y) \in \mathbb{R}_+^m \times \mathbb{R}_+^n | \exists \lambda \in \wedge^i(\gamma) \colon x \geq \sum_{i=1}^{I} \lambda^i x^i, y \leq \sum_{i=1}^{I} \lambda^i x^i\} \tag{3.7}$$

The original DEA model applies constant returns to scale (CRS) but there are also other alternative models. These are the free disposable hull

(FDH), varying returns to scale (VRS), decreasing returns to scale (DRS), increasing returns to scale (IRS) and the free replicability hull (FRH). Each of the formulations would yield a different optimal frontier as shown in Figure 3.1.

Figure 3.1: DEA types under different technical assumptions.
Source: Bogetoft and Otto (2010, pg. 81).



Once the frontier has been established, the second step would be to combine it with the Farrell (1957) efficiency measures in order to obtain an efficiency score for each firm.

Thus, the efficiency of firm i is a positive real number calculated on the minimum distance between the set of input-output combinations for firm i and the production function T^*_γ.

However, regardless of the validity of the DEA method as a way to evaluate railway efficiency, there are several limitations in its application to assess firm efficiency. The most significant ones, partially based in Graham (2008), are the following:

1. There is no explicit empirical assessment of the relation between inputs and output. The DEA method for modelling technical and allocative efficiency is atheoretical where there are no assumptions on firm technologies.

2. The composite measure of efficiency is a gross measure which does not control for firm operational or business circumstances, so to some extent, this could affect the comparability of the firms or units.

3. The DEA approach does not consider the characteristics of panel data, so unobserved individual effects are not taken into account.

4. The deterministic frontier methods interpret the deviations from the frontier curve as direct estimates of individual firm inefficiency. A weakness of constraining the observations to be on, or below, the frontier is when either measurement errors are present in the data or some type of

specification errors and omitted variables occur in the imposed model structure.

5. The DEA approach assumes that all deviation from the frontier is due to inefficiency and, consequently, it is very responsive to sample outliers and its estimates can be heavily distorted by extreme values (Cowie and Riddington, 1996).

6. Although some DEA methods allow for the assumption of variable returns to scale, it is still difficult to distinguish between economies of scale and economies of density in the DEA approach (Nash and Smith, 2014). However this distinction between economies of scale and density is particularly critical for the rail industry cost efficiency analysis.

7. The DEA requires large number of observations in order to be accurate, as the frontier is defined by the sample observations.

8. There is no evaluation of how changes in individual inputs can affect changes in outputs, which invalidates the DEA to break down the determinants of efficiency.

9. The DEA efficiency scores can be regressed (e.g. Tobit regression) in a second step in order to find the determinants of efficiency. However, it has been found that this Tobit regression can be serially correlated if the data set of the research does not include all the members of the population (Simar and Wilson, 2007). Besides this, more observations are needed if the researcher aims to explain the determinants of the efficiency score by means of a subsequent Tobit regression of the efficiency

scores on its determinants.

10. The DEA lacks direct interpretation of the efficiency analysis with regard to costs or unit costs such as cost elasticities. This compromises the actual implementation of the results by practitioners and rail regulators.

3.1.2 Stochastic data envelopment analysis

There have been numerous attempts to combine the advantages of the DEA approach with the control of data noise and stochastic variation as in the stochastic frontier analysis (SFA, explained below in subsection 3.2.2).

The keystone in stochastic data envelopment analysis (SDEA) is how to marry the frontier estimation with a treatment of data variation. The frontier estimation is done using DEA approaches, as it does not impose any parametric form for the function defining the frontier. However, SDEA approaches try to treat data variation as a sum of actual and random variation.

The SDEA methods try to avoid the constraint of predetermined probability distributions and ex-ante frontier functional forms while simultaneously attempting to disentangle the random noise in the data from the true efficiency variation.

Overall, the synthesis between DEA and stochastic approaches has not been widely applied for the time being as it not only mixes their advantages but also their drawbacks. Thus, the SDEA approach is more complex and harder to explain than a single DEA or SFA method while its gains over a combination of separate DEA and SFA analyses are dubious.

3.2 Parametric Frontier Methods

3.2.1 Deterministic econometric analysis

The deterministic parametric methods correspond to all those econometric approaches where, firstly, a functional form is predetermined and, secondly, all the data variation is assumed to contain information on the cross units behaviour and not data noise. This corresponds to the ordinary least squares (OLS) approach and all its derivative applications in panel data regression models.

Simple regression models tend to endure omitted variable problems which are usually not wholly solved by controlling for more variables and lagged dependent variables. This omitted variable problem is easier to handle with panel data sets: data with several observations of the same cross-sectional units ($i = 1, \ldots, I$) over several time periods ($t = 1, \ldots, T$).

The omitted variables are either constant unobserved factors or time-varying factors affecting the dependent variable. In the linear model shown in Equation 3.8, the dependent variable is explained by a constant or intercept (β_0), a set of binary variables for each period (t) to alter the intercept ($period_t$) and a set of n ($n = 1, \ldots, N$) covariates (x_{it}) with particular values for each period (t) and cross unit (i). Next, a time-constant factor (α_i) to capture non-varying characteristics of each cross unit (i.e. unobserved heterogeneity) and lastly, a time-varying error term (ε_{it}) which varies for each

cross unit (i) and time period (t).

$$y_{it} = \beta_0 + \delta_1 period_1 + \ldots + \delta_T period_T + \beta_1 x_{it1} + \ldots + \beta_N x_{itN} + \alpha_i + \varepsilon_{it} \quad (3.8)$$

The estimation of the parameters of interest, β_1 to β_n, can be done by different approaches depending on the assumptions on the unobserved heterogeneity or fixed effect (α_i). If ν represents a compound error (i.e. $\nu_{it} = \alpha_i + \varepsilon_{it}$), then there are three general approaches depending on whether α_i is disregarded (i.e. Equation 3.9), considered as a fixed value for each cross individual (i.e. Equation 3.10) or, thirdly, considered as a random variable (i.e. X in Equation 3.11).

$$\text{Pooled OLS} : \nu_{it} = \varepsilon_{it} \text{ where } \alpha_i \approx 0 \quad (3.9)$$

$$\text{Fixed effects} : \nu_{it} = \alpha_i + \varepsilon_{it} \text{ where } \alpha_i = \{i, \ldots, I\} \in \mathbb{R} \quad (3.10)$$

$$\text{Random effects} : \nu_{it} = \alpha_i + \varepsilon_{it} \text{ where } \alpha_i \sim X \quad (3.11)$$

First, a pooled OLS where the cross-sectional units are treated as if they were independently sampled units for each period, although in the case of panel data, they are the same cross units over time. This would lead to Equation 3.12

$$y_{it} = \beta_0 + \delta_1 period_1 + \ldots + \delta_T period_T + \beta_1 x_{it1} + \ldots + \beta_N x_{itN} + \nu_{it} \quad (3.12)$$

where ν_{it} is a compound error term equal to $\nu_{it} = \alpha_i + \varepsilon_{it}$. In this case, the pooled OLS model in Equation 3.12 is unbiased only if the compound error (ν_{it}), and particularly the fixed effect term (α_i), are uncorrelated with the variables of interest (x_{it}). This is equivalent to $Cov(x_{itj}, \alpha_i) = 0$ and $Cov(x_{itj}, \varepsilon_{is}) = 0$ for any time period.

Another approach would be to differentiate the observations over time, as seen in Equation 3.14, so those variables constant over time such as the fixed effect (α_i) would disappear. In this instance, strict exogeneity of $\Delta\varepsilon_{it}$ with respect to the explanatory variables is required for unbiased estimates (i.e. $Cov(x_{itj}, \Delta\varepsilon_{it}) = 0$, which requires $Cov(x_{ij}, \Delta\varepsilon_i) = 0$ for all t). Likewise, the explanatory variables (x_{it}) should neither be time-invariant for all i nor change by the same amount for every observation in order to obtain unbiased estimates.

$$y_{it} = \beta_0 + \delta_1 period_1 + \ldots + \delta_T period_T + \beta_1 x_{it1} + \ldots + \beta_N x_{itN} + \alpha_i + \varepsilon_{it} \tag{3.13}$$

$$\Delta y_{it} = \delta_1 \Delta period_1 + \ldots + \delta_T \Delta period_T + \beta_1 \Delta x_{it1} + \ldots + \beta_N \Delta x_{itN} + \Delta\varepsilon_{it} \tag{3.14}$$

Equation 3.14 corresponds to the subtraction of Equation 3.13 with the same Equation 3.13 for the previous period (i.e. $t - 1$) from Equation 3.13 shown above. Effectively, each Equation 3.14 represents data for two time periods ($t = 0$ and $t = 1$) for cross unit i.

Alternatively to first differences, another transformation to remove the

unobserved effect is the fixed effects, or within, transformation. In this case, the initial model of Equation 3.15 is averaged for each cross unit (i) to Equation 3.16

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_j x_{itN} + \alpha_i + \varepsilon_{it} \tag{3.15}$$

$$\bar{y}_{it} = \beta_0 + \beta_1 \bar{x}_{it1} + \ldots + \beta_j \bar{x}_{itN} + \alpha_i + \bar{\varepsilon}_{it} \tag{3.16}$$

The resulting variables equal, for each i, a simple mean (e.g. $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$). The time-invariant unobserved heterogeneity is still present (α_i) but it can be differentiated away by subtracting Equation 3.16 from Equation 3.15. In this case, the time invariant elements (i.e. intercept β_0 and unobserved heterogeneity α_i) disappear.

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it1} + \ldots + \beta_j \ddot{x}_{itN} + \ddot{\varepsilon}_{it} \tag{3.17}$$

Equation 3.17 corresponds to the fixed effects or within estimator which is unbiased as long as $\ddot{\varepsilon}_{it}$ is uncorrelated with all covariates (x_{itj}) for all periods ($t = 1, 2, \ldots, T.$). On the contrary, the within estimator is not biased by correlation between the fixed effect term (α_i) and the explanatory variables (x_{itn}). As in the first difference estimator, the time-invariant variables are eliminated from the models. This would imply that usual regressor variables that tend to be stable over time, such as number of stations, network length or opening hours per week, would not be able to be included in the analysis. This would constrain the explanatory power of the models and, specially, the

assessment of economies of scale in the urban rail systems as the scale of the systems is usually constant over short period of time. An alternative would be to interact the time-invariant variables with variables that change over time (e.g. time binary variables) but this would yield an estimate on how the effect of the time-invariant variable changes over time and not the base effect of the time-invariant variable itself.

The random effect estimator allows for the explanatory variables to be time-invariant, which is particularly important for urban rail systems, where certain key explanatory variables do not change over time (e.g. network length or number of stations). However, it relies on two assumptions. Firstly, the strict exogeneity explained above for the pooled OLS approach. Secondly, the assumption that the unobserved time-invariant effect is uncorrelated with the explanatory variables for all time periods (i.e. $Cov(x_{itj}, \alpha_i) = 0$). Moreover, the random effects estimator can also deal with positive serial correlation in the error terms; in this case, given Equation 3.18, where the compound error term is $\epsilon_{it} = \alpha_i + \nu_{it}$. Thus, the fixed time invariant individual effect is independent from both the predictor variables $(x_{it1}, \ldots, x_{itn})$ and the random disturbance (ν_{it}). In the random effects estimator, the fixed effect (α_i) and the random variation (ν_i) are assumed to be identically and independently distributed with zero mean, that is, $\alpha_i \sim IID(0, \sigma_\mu^2)$ and $\nu_i \sim IID(0, \sigma_v^2)$.

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_n x_{itn} + \epsilon_{it} \tag{3.18}$$

In the case of Equation 3.18, the compound error term is serially correlated across the time periods as shown in Equation 3.19, where t and s

represent different time periods.

$$Corr(\nu_{it}, \nu_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\nu^2} \tag{3.19}$$

The random effects estimator can deal with the positive serial correlation in the compound error term by means of the generalised least squares (GLS) approach. This is particularly true when the number of cross units ($k = 1, 2, \ldots, K$) is substantially larger than the time periods ($t = 1, 2, \ldots, T$) of the panel.

The GLS transformation implies the definition of λ as in Equation 3.20, where the value of λ ranges between 0 and 1 and tends to 0 as the number of time periods of the panel (T) increases.

$$\lambda = 1 - \left[\frac{\sigma_\nu^2}{(\sigma_\nu^2 + T\sigma_\alpha^2)}\right]^{1/2} \tag{3.20}$$

In this case, the GLS transformation can be applied on the initial equation 3.21 to obtain Equation 3.22

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_j x_{itj} + \nu_{it} \tag{3.21}$$

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{it1} - \lambda\bar{x}_{i1}) + \ldots + \beta_j(x_{itj} - \lambda\bar{x}_{ij} + (\nu_{it} - \lambda\bar{\nu}_i) \tag{3.22}$$

The GLS random effects estimator would subtract the λ share of the time average for each cross unit i. As λ ranges between 0 and 1 and the random

effects estimator assumes no correlation between the unobserved effect and the explanatory variables (i.e. x_{it}), these explanatory variables in Equation 3.22 can be time-invariant.

As λ is not usually known, the random effects uses an estimated value $\hat{\lambda}$ which is also know as the feasible GLS (FGLS). The estimated $\hat{\lambda}$ is based on the estimates of $\hat{\sigma_{\alpha}^2}$ and $\hat{\sigma_{\varepsilon}^2}$.In the FGLS approach, a variance matrix is formed as in Equation 3.23.

$$\Omega_{T\times T} = \begin{bmatrix} \sigma_{\alpha}^2 + \sigma_{\nu}^2 & \sigma_{\alpha}^2 & \dots & \sigma_{\alpha}^2 \\ \sigma_{\alpha}^2 & \sigma_{\alpha}^2 + \sigma_{\nu}^2 & \dots & \sigma_{\alpha}^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{\alpha}^2 & \sigma_{\alpha}^2 & \dots & \sigma_{\alpha}^2 + \sigma_{\nu}^2 \end{bmatrix} \tag{3.23}$$

This matrix ($\Omega_{T\times T}$) is used in the Generalised Least Squares approach to estimate the coefficients for the elasticity estimates. The FGLS approach of the random effects estimator presents an intermediate result between the pooled OLS and the fixed effects estimator. When the values of λ are close to 0, the random effects estimates will tend to be similar to the pooled OLS estimates. This would also entail that the unobserved heterogeneity factor (α_i) has a small or negligible effect. On the contrary, values of λ close to 1 imply a large σ_{α}^2 with respect to $\bar{\sigma_{\nu}^2}$ and, thus, the random effect estimates would be closer to the fixed effects estimates. Likewise, and following on from Equation 3.20, a panel with large time series (i.e. large T) would also tend to higher λ values and fixed effect estimates.

This can also be observed from the decomposition of the compound error

ϵ_i in Equation 3.24

$$\epsilon_{it} - \lambda\bar{\epsilon_i} = (1-\lambda)\alpha_i + \nu_{it} - \lambda\bar{\epsilon_i} \tag{3.24}$$

In Equation 3.24 it is shown that the compound error term of the random effects estimator is weighted by $(1-\lambda)$. Thus, even if the correlation between the unobserved heterogeneity factor (i.e. α_i) and the explanatory variables (i.e. x) is not zero, the FGLS approach reduces the effect of the bias.

Panel effects test

Random effects represents an intermediate approach between the pooled OLS and the fixed effects. The Breusch-Pagan test is a standard approach to distinguish whether pooled OLS or random effects is preferred. The Breusch-Pagan test (Breusch and Pagan, 1980) examines whether the variance of the individual random effects are equal to zero (i.e. $H_0 : \sigma_i^2 = 0$). This requires estimating with pooled-OLS the initial model shown in Equation 3.25.

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j + \varepsilon \tag{3.25}$$

$$\hat{u}^2 = \beta_0 + \beta_1 x_1 + \ldots + \beta_j x_j + \nu \tag{3.26}$$

$$\text{LM statistic} = nR^2_{\hat{u}^2} \tag{3.27}$$

The squared residuals of the model (i.e.$R^2_{\hat{u}^2}$) are regressed on the explanatory variables to test whether the residuals are related to the explanatory variables (i.e. Equation 3.26). The null hypothesis is that the coefficients

of the explanatory variables in Equation 3.26 are zero with a zero mean error term (i.e. ν). Lastly, an F-test is computed on this latter regression Equation 3.26. If the null hypothesis is discarded, it confirms the existence of variation across the cross-units (i.e. panel effect) and pooled-OLS should not be used for the estimation of the model.

Once the existence of panel effects has been confirmed, the choice between random and fixed effects estimation is determined by the results of the Hausman test (Hausman, 1978). The null hypothesis of the Hausman test is shown in Equation 3.28. If the null hypothesis holds, the random effects estimator is preferred.

The Hausman tests assess if there are statistically significant differences in the random and fixed effects estimates of the time-varying variables as shown in Equation 3.29. If the null hypothesis is true, H in Equation 3.29 would equal the Chi-squared distribution with one degree of freedom.

$$H_0 : Cov(\alpha_i, X_{it}) = 0 \tag{3.28}$$

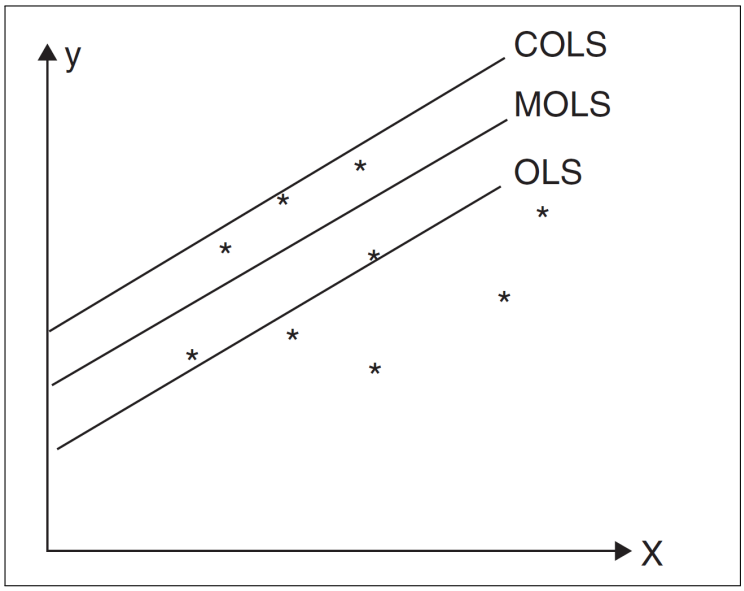
$$H = \frac{(\hat{\beta}^*_{FE} - \hat{\beta}^*_{RE})^2}{Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})} \sim \chi^2_1 \tag{3.29}$$

Regression residuals and efficiency analysis

Regarding the assessment of organisational or firm efficiency, the deterministic parametric models focus on the analysis of the residuals. The difference between the expected cost by the model and the actual cost of the firm can yield an approximation of the relative efficiency of the firm with three main

approaches: OLS, COLS and MOLS as seen in Figure 3.2.

Figure 3.2: Comparison of OLS, COLS and MOLS frontiers. Source: Adapted version from Greene (2008, pg. 106).



In the past, most of the analyses on efficiency were done by means of a corrected ordinary least squares (COLS) which is still popular in the panel data literature (Greene, 2008). The COLS takes as a reference the firm with the largest negative residual (i.e. the firm with the largest overestimation of costs) and uses this residual as a yardstick to compare the efficiency of the rest of the firms in the sample. Even though the OLS model can be unbiased and consistent, this approach is very sensitive to the existence of outliers and it can offer misleading efficiency estimates. An alternative approach would be the modified OLS (MOLS) where the variance of the negative residuals, or other higher moments of the negative residuals, are used to estimate a frontier. A common comparison of the three frontiers estimated by OLS, COLS and MOLS is shown in Figure 3.2.

Overall, the parametric and deterministic models have significant advantages over other approaches, such as:

1. The results offer a direct assessment of the relation between inputs and outputs. The interpretation of the cost efficiency of the firm is expressed in operational terms, which are intelligible and applicable estimates (e.g. output or cost elasticities). Moreover, this enables sensitivity tests with real operational data to check the validity of the estimates, something that would be more difficult with abstract measurement of efficiency (e.g. efficiency index).

2. The models leave room to control for operational and business circumstances which may explain cost efficiency and general performance. These include the disaggregated assessment of economies of density and economies of scale.

3. The statistical significance of the model estimated can be tested, given some distributional assumptions.

4. These econometric approaches can also take advantage of panel data sets and account for unobserved individual effects.

5. The GLS is particularly effective, with respect to other techniques, when the cross units ($k = 1, 2, \ldots, K$) are large and there are few time periods ($t = 1, 2, \ldots, T$) (Wooldridge, 2010).

6. These models are the least data-intensive in comparison with other techniques to assess efficiency.

7. Despite the strong assumptions required by the random effects (i.e. strict exogeneity), the effect of a fail in these assumptions can be partially moderated by the estimator.

Nevertheless, there are also significant drawbacks to the deterministic and parametric econometric approaches such as the following:

1. All the data variation is assumed to represent true cost efficiency variation and not data noise or measurement error.

2. The models rely on distributional assumptions on the error terms for estimation and statistical significance tests.

3. Some estimators like the random effects require strict exogeneity assumptions to be completely unbiased.

4. The loss of observations for some of the methods in order to differentiate away the unobserved effect (e.g. first differences estimator)

5. With the exception of the random effects estimator, most of the techniques are not suitable for assessing the impact of time-invariant explanatory variables.

3.2.2 Stochastic frontier analysis

The stochastic frontier analysis (SFA) is an increasingly popular approach used to assess firm efficiency and was first introduced by Aigner et al. (1977). This section presents the main SFA approach, although a detailed description of the approach and its numerous alternatives is available from other sources

(e.g. Baltagi, 2005, Greene, 2002, Kumbhakar and Lovell, 2003). The SFA is a parametric approach that enables certain stochastic variation in the relationship between inputs and outputs due to data noise, measurement error or even extraordinary circumstances in the production process (e.g. weather).

As in any parametric method, a certain functional form (Equation 3.30) is assumed *ex-ante* where the coefficients (β) for the parameters (x) have to be estimated. The deviation of the estimates from the corresponding actual observations of the sample is condensed in the error term (ε) for all i firms ($i = 1, \ldots, I$).

$$f(x_i) = f(x_i; \beta) + \epsilon_i \tag{3.30}$$

$$\text{where} \quad \epsilon_i = \nu_{it} - \alpha_i \tag{3.31}$$

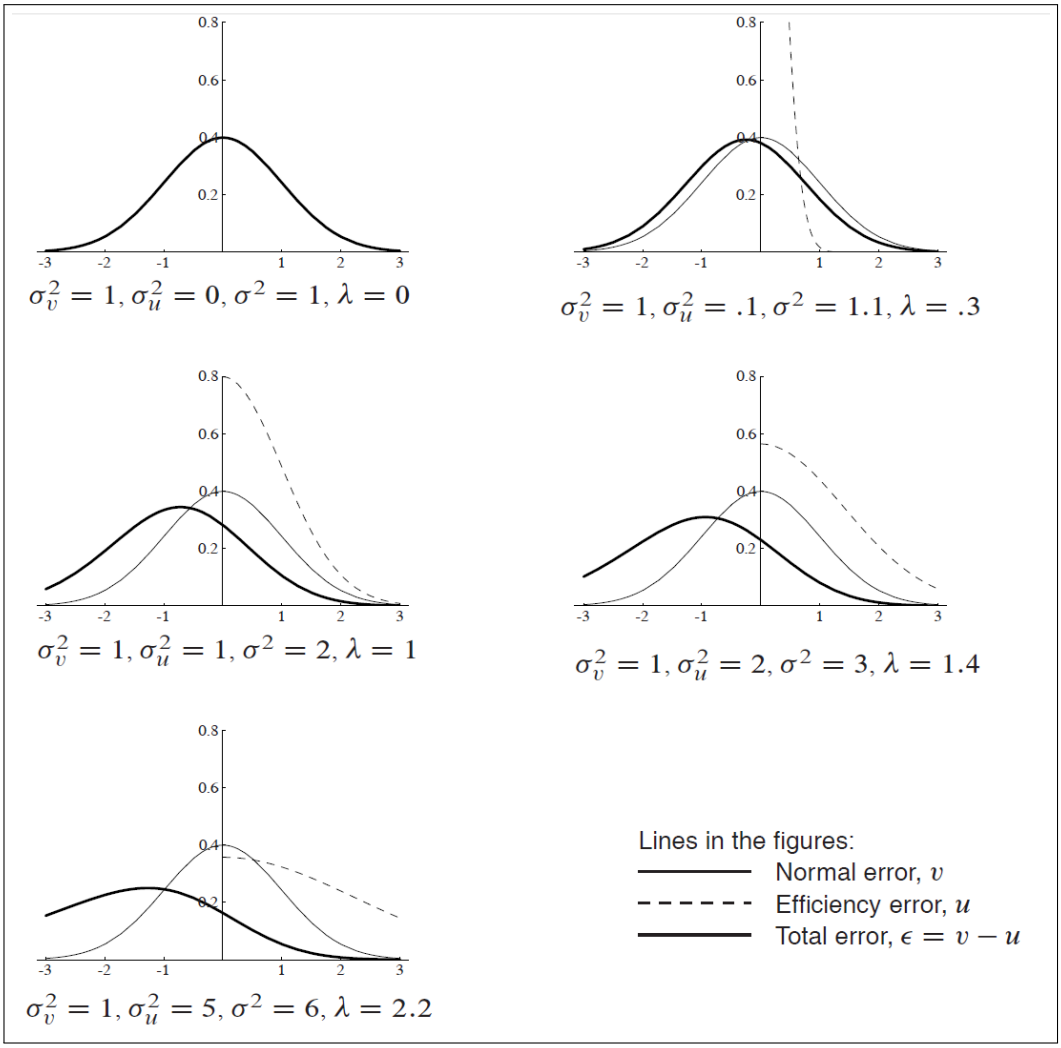
$$\nu_i \sim N(0, \sigma_\nu^2), \tag{3.32}$$

$$\alpha_i \sim N_+(0, \sigma_\alpha^2) \tag{3.33}$$

The distinctive approach of the SFA method is the characterization of the (ϵ). The divergence between the actual observations sample data and the estimated values by the model for each i firm can be due to either actual inefficiency, data noise or randomness. In the case of the SFA, the method tries to break down each of the estimated differences (model residuals, $\hat{\epsilon}$) into data noise (ν) and actual firm inefficiency (α).

The error term is assumed to be a white noise process with zero mean,

Figure 3.3: Distribution of error, efficiency and overall error term in the SFA. Source: Bogetoft and Otto (2010, pg. 206).



uncorrelated and with constant variance. The efficiency term has a truncated distribution which can only have zero or positive values, as a firm efficiency can be neither negative nor over-efficient. The error term ϵ is formed by the subtraction of random stochastic noise (ν) minus the efficiency deviation α as in Equation 3.31. The disentangle between the random noise and the efficiency term is based on the non-negative inefficiency term distribution.

The parameters to be estimated are shown in Equation 3.34 where a $\sigma_\alpha^2 = 0$ would lead to $\lambda = 0$ as in the top left chart of Figure 3.3, which equals the conventional regression approach.

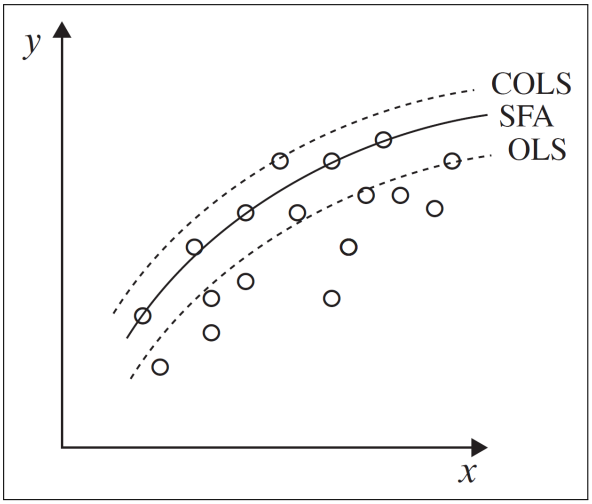
$$\sigma^2 = \sigma_\nu^2 + \sigma_\alpha^2 \quad \text{and} \quad \lambda = \sqrt{\frac{\sigma_\alpha^2}{\sigma_\nu^2}} \tag{3.34}$$

On the contrary, a large variance of the inefficiency term (α) tends to move the distribution of the error term (ϵ) towards negative values. So larger values of λ imply a more skewed total error distribution, which means that the inefficiency term is more influential in determining the total error. This is also shown in the bottom left graph at Figure 3.3.

The error and the efficiency term are independent and not correlated. Thus, for certain observations, a large positive error term may compensate the negative value of the efficiency term and set that particular firm observation beyond the optimal frontier. This would not be possible with the corrected ordinary least squares approach (COLS), where the frontier is determined by the observations with the largest positive residual as seen in the comparison of COLS, SFA and OLS methods in Figure 3.4.

This basic specification can be complemented with extensions to accommodate panel data (Pitt and Lee, 1981), so the model can also deal with unobserved characteristics of the firm that can bias the estimates (e.g. unobserved heterogeneity). More recent approaches aim to break down the total error term (ε) into more constituents (Colombi et al., 2014, Kumbhakar et al., 2014). These newer approaches, with a more detailed explanation of the total error term, also rely on distributional assumptions and increase model

Figure 3.4: Comparison of SFA, COLS and OLS methods.
Source: Bogetoft and Otto (2010, pg. 205).



complexity.

Overall, despite the significant advantages when compared to other methods, there are also significant drawbacks to its application and the method is not appropriate for the the CoMET-Nova data as shown in the Annex 11.8 (page 305). In short, the main drawbacks of the SFA approach are the following:

1. Assumption of a functional form *ex-ante* for the production function which may lead to an inaccurate representation of the production process of the firm (Karlaftis and Tsamboulas, 2012).

2. The estimates are not unbiased and consistent as the maximum likelihood estimation requires large data samples. Under small samples, the estimates have been usually found neither sensible nor reliable (Nash and Smith, 2007, pg.681).

3. The estimation method based in the maximum likelihood is not the

most efficient in small data samples, which reduces the statistical significance of the parameters estimated (Bogetoft and Otto, 2010).

4. The SFA model relies on distributional assumptions which are relatively arbitrary in order to separate the randomness of the total error term (Berger and Mester, 1997).

5. Lack of direct interpretation of the efficiency to costs or unit costs (e.g. cost elasticities). The models estimate a level of efficiency or inefficiency but the results are not directly interpretable in operational terms. A regression can be done to explain the determinants of the efficiency level but it would not offer a direct interpretation related to unit costs.

3.3 Methodological Challenges of the Research

This section reviews the main methodological challenges faced by this research. It expands on the general description of the previous sections and outlines how these challenges have been dealt using the methods described in this thesis.

3.3.1 Unobserved heterogeneity

As reviewed in subsection 3.2.1, unobserved heterogeneity (α_i) refers to the time-invariant characteristics of each particular cross unit or metro that are not captured by any of the explanatory variables (x_{it1}) included in the model. If the unobserved heterogeneity is not accounted in the model, it would be

part of the error term (ε_{it}) leading to a compound error term ($\nu_{it} = \alpha_i + \varepsilon_{it}$). The true model (Equation 3.35) would be estimated as in Equation 3.36. Unobserved heterogeneity could lead to biased estimates in Equation 3.36 if the time-invariant characteristics are correlated to the explanatory variables (i.e. $Cov(x_{itj}, \alpha_i) \neq 0$). This would be an additional assumption of exogenous variables with respect to the error term (i.e. $Cov(x_{itj}, \varepsilon_{it}) = 0$).

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_N x_{itN} + \alpha_i + \varepsilon_{it} \tag{3.35}$$

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_N x_{itN} + \alpha_i + \nu_{it} \tag{3.36}$$

The most common approach dealing with unobserved heterogeneity is to take differences of consecutive panel observations for the same cross unit. This eliminates the time-invariant coefficient of unobserved heterogeneity (α_i) although it can also eliminate time-invariant explanatory variables. In the case of this research, some time-invariant variables are critical for the assessment of cost efficiency. Some variables, such as the percentage of labour outsourcing, the network length or the number of stations may remain invariant over several years.

Panel models with time-invariant variables can be estimated by means of a random effects approach. In this case, the unobserved effect (α_i) is assumed to be uncorrelated with the explanatory variables (i.e. $Cov(x_{itj}, \varepsilon_{it}) = 0$). This assumption can be tested by means of a Hausman test (Hausman, 1978) for statistically significant differences between the coefficients of random and fixed effects estimates. However, in the case of time-invariant explanatory

variables which are critical for the analysis, the assumption of unobserved heterogeneity uncorrelated with the explanatory variables can be acceptable under certain circumstances. For instance, "if we have good controls in our equation, we might believe that any leftover neglected heterogeneity only induces serial correlation in the composite error term, but it does not cause correlation between the composite errors and the explanatory variables" (page 481 in Wooldridge, 2012).

In this particular research, the range of control variables used for the cost efficiency models is probably among the best in the corpus of academic literature. All the models included in this research had at least 5 control variables based on the confidential data reported by the metro operators to the RTSC (e.g. labour wages or outsourcing), which would substantially reduce the risk of significant unobserved heterogeneity. Likewise, one of the key research contributions of this thesis is precisely the estimation of these control variables.

3.3.2 Functional form

Any econometric research is susceptible to model misspecification. Misspecified models would not accurately describe the relationship between the explanatory variables and the dependent variable, leading to incorrect conclusions. The most popular functional form is based on the Cobb-Douglas functional form (i.e. $Y = AL^aK^b$) explained in subsection 3.1.1. The Cobb-Douglas functional form offers a more straightforward estimation and interpretation of the coefficients estimated, i.e. the production or cost elasticities

(a and b). Likewise, the estimation entails fewer parameters, which is particularly important for data sets with few observations. On the contrary, the Cobb-Douglas functional form is less flexible and assumes fixed relationships. For instance, the returns to density estimated would be constant, even though it is reasonable to assume that at some point congestion costs would probably offset density gains and, therefore, the relationship would not be linear. Other research on rail production functions (McGeehan, 1993) has found that cost functions for railways are non-homothetic and non-homogeneous, so the parameters estimated should not be constrained to sum up to 1.

An alternative to the Cobb-Douglas functional form is the transcendental logarithmic production function (i.e. translog). Translog production functions offer quadratic terms and interactions between all the parameters leading to a more flexible functional form. For instance, the estimate on returns to density would probably be non-linear (i.e. non-constant) as expected by economic theory. However, the existence of quadratic and interaction terms leads to a more difficult estimation and interpretation. Likewise, the number of parameters to be estimated grows exponentially due to the estimation of interaction terms between all the explanatory variables of the model. Consequently, it may be unfeasible to estimate translog models with many parameters if the data set is small. In the case of railway cost efficiency, a review of previous evidence showed that results based on translog functional forms can be potentially misleading (Cowie and Riddington, 1996). As a result, this research has considered these two main functional forms. However, considering the aim to test new cost drivers that have not been tested before, as well as the data set size limitations, the most pertinent functional form is

the Cobb-Douglas.

3.3.3 Omitted variable bias

Omitting a variable that belongs to the population model biases the model estimates. Given a true population model in Equation 3.37, excluding a relevant variable (i.e. β_2) would lead to underspecified model estimates (i.e. $\widetilde{\beta}_1$ as in Equation 3.38).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_{it} \tag{3.37}$$

$$\widetilde{y} = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_1 + \varepsilon_{it} \tag{3.38}$$

For Equation 3.38, the expected value of the biased estimate (i.e. $E[\widetilde{\beta}_1]$) is determined by the expected value of the true population model (i.e. β_1) but also by the relationship between x_1 and the omitted variable x_2 (i.e. δ_{x_1,x_2}) as shown in Equation 3.39. The relationship between x_1 and the omitted variable x_2 represented by δ_{x_1,x_2} is equal to the slope from the simple regression of x_2 on x_1.

$$E[\widetilde{\beta}_1] = E[\hat{\beta}_1 + \hat{\beta}_2 \delta_{x1,x2}] \tag{3.39}$$

$$Bias(\widetilde{\beta}_1) = E[\widetilde{\beta}_1 - \beta_1 = +\beta_2 \delta_{x_1,x_2}] \tag{3.40}$$

$$\delta_{x1,x2} = \frac{Cov(x_1, x_2)}{Var(x_1)} \tag{3.41}$$

The omitted variable bias is independent of the sample size, so larger

samples still present the same bias. Besides this, the actual slope parameter of variable x_2 (i.e. β_2) is unknown, although it is possible to hypothesise what effect would variable x_2 would have on the dependent variable y.

The potential number of omitted variables in an econometric model can be numerous. The best solution is to include all those variables that are hypothesised to have an effect on the dependent variable or, at least, an observable proxy variable which is related but not identical to the missing omitted variable. Given the large number of variables available in the CoMET-Nova data set, this is the first approach adopted in this thesis. All the cost efficiency models have the relevant, or at least some proxy variables, which are explained in detail in each corresponding chapter. However, due to the lack of observations and the lower quality of some variables, some explanatory variables were not included. These excluded variables are explained in the corresponding subsection of each chapter (e.g. pages 176, 200 and 229).

Another alternative is to use an instrumental variable (IV) to deal with the omitted variable bias. In this situation, given that variable x_2 is missing in Equation 3.38, it is possible to have an unbiased estimate of β_1 if there is a good instrument for variable x_1. The instrument for variable x_1 is another variable (i.e. z) that meets two conditions:

1. Exogeneity: the instrument variable (z) is uncorrelated with the error term (i.e. ε_{it}) as shown in Equation 3.38. This therefore corresponds to the assumption of $Cov(Z, \varepsilon_{it}) = 0$ which is usually based on economic theory or general conjectures.

2. Relevance: the instrument variable (z) is correlated with the endoge-

nous variable (i.e. x_1), either positively or negatively. This is summarised as $Cov(z, x_1) \neq 0$.

The instrument relevance requirement can usually be fulfilled and tested by means of standard regressions. However, the instrument exogeneity requirement is based on an assumption that cannot be easily tested (i.e. $Cov(Z, \varepsilon_{it}) = 0$). Likewise, the availability of instrumental variables outside the CoMET-Nova data set variables was very limited due to the geographical diversity of the metros studied. Given that the panel data used comes from metros from different countries across the world, it has been extremely complicated to find suitable variables to be included in the model that, simultaneously, were available for all these different metros and countries.

3.3.4 Reverse causality

Reverse causality is a potential source of bias in the estimation of regression models due to the violation of the zero conditional mean assumption, one of the Gauss-Markov assumptions for unbiased estimates. The zero conditional mean of errors assumption implies that the expectation of the error term, given the independent variables, has to be equal to zero, as shown in Equation 3.42 and 3.43. If this assumption does not hold, the regressor (e.g. x) is endogenous.

$$E[\varepsilon_{it}|x] \neq 0 \tag{3.42}$$

$$Cov(\varepsilon_{it}, x) \neq 0 \tag{3.43}$$

There are several ad hoc solutions for dealing with endogeneity. Firstly, the endogenous variables can be added to the model with one or more lag periods. This can be effective when current variable values are endogenous but past values of the same variable are exogenous. In the case of this research this alternative was not available as, due to the unbalanced characteristic of the panel, lagging one or more periods in some variables would have substantially decreased the number of observations.

Another alternative solution is to proxy the dependent variable with a variable that solves the problem of endogeneity in the model. This solution was not suitable due to two main reasons. First of all, replacing the operating cost variables would imply less precision in the conclusions of the thesis and consequently would compromise their validity. Secondly, there is no suitable proxy for the operating costs that would, at the same time, be less endogenous than the original dependent variable.

Thirdly, the best approach would be to use an instrumental variable (IV) which is exogenous but strongly correlated with the endogenous regressor. Unfortunately, this approach was not possible because of the difficulty to find suitable instruments as described above in the subsection 3.3.3 on omitted variable bias.

3.3.5 Measurement error

Measurement error refers to the lack of accuracy in the variables used in the econometric model. This is usually because the values of the variable we are interested in vary randomly from its true values. Any variable (e.g. x_1

explanatory variable) is susceptible to being reported with some degree of inaccuracy, leading to a measurement error (i.e. e_1). The measurement error is equal to the reported value for the explanatory variable (i.e. x_1) minus the true unobserved actual value (i.e. x_1^*) as shown in Equation 3.44.

$$e_1 = x_1 - x_1^* \tag{3.44}$$

$$Cov(x_1^*, e_1) = 0 \tag{3.45}$$

The measurement error is also assumed to be independent of the true unobserved value as shown in Equation 3.45. Consequently, the reported value and the measurement error have to be correlated (Wooldridge, 2012) and the covariance of these two is equal to the variance of the error presented in Equation 3.46. This leads to an attenuation bias as the estimates of β will always be closer to zero than the true population β_1 as shown in Equation 3.47.

$$Cov(x_1, e_1) = E(x_1 e_1) = \sigma_{e_1}^2 \tag{3.46}$$

$$plim(\hat{\beta}_1) = \beta_1 \left(1 - \frac{\sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} \right) \tag{3.47}$$

The impact of the measurement error in regression analysis can be moderate with high-quality data. In this research, measurement error is a minor concern given the data available shown in Chapter 4. The data for the CoMET-Nova consortia is reported by experienced urban rail professionals

who have supplied data for the same variable over several years. Besides the iterative process, the data provided follows a well-defined data handbook which details what is included or excluded for each data variable. Likewise, in the case of the dependent variables of this research, they refer to aggregate categories of more detailed data items, so the random variation of the disaggregated data items would cancel out once they are all aggregated (e.g. operating costs defined as the sum of several subcategories of operating costs).

3.4 Methodology Review Conclusions

This section reviews and explains the main methods used to assess cost efficiency of firms. The methods have been broken down into four main groups depending on whether they are parametric/non-parametric and deterministic/stochastic.

The main non-parametric approach is the data envelopment analysis (DEA), which has been popular in the railway efficiency literature. However, there are significant drawbacks to its application, as described in section 3.1.1. The main drawback is the need for large data sets in order to yield sensible estimates. Besides this, the DEA approach can hardly differentiate between economies of density and economies of scale, which is a critical distinction for the urban rail systems, and there is no evaluation on the link between inputs and outputs.

Regarding parametric methods, both stochastic and deterministic methods overcome some of the DEA drawbacks. The stochastic approaches are

increasingly popular in the literature and can deal with the randomness of the data or its measurement error. However, stochastic methods also require large data sets. Moreover, the breakdown of efficiency and data random noise is based on distributional assumptions which may be inaccurate. Lastly, the results do not offer a direct interpretation but an intermediary assessment of an abstract efficiency index.

Therefore, given the academic contribution this research intends to make and the data set characteristics and size constraints, the best option for the cost efficiency analysis is the deterministic parametric method for panel data sets such as the random effects estimator. These methods can test many variables which are available from the CoMET-Nova data sets. At the same time, these methods are able to yield sensible results when the number of data observations are low and the panel is unbalanced. Moreover, the statistical significance and the particular relation between inputs and outputs (e.g. cost elasticities) can be assessed with panel data estimators.

Overall, the choice of the most appropriate method must be decided together with the data available for the research. The CoMET and Nova data set is relatively wide, with many variables for each operator which can yield new insights in the field of cost efficiency analysis. However, it is relatively short, as the number of observations per unit is low for most of the members and some members do not have all the years available. A more detailed explanation of the data used for this research is presented in the following section.

4 — Data

The existing literature on railway cost efficiency has been constrained in particular by data availability. As explained in the conclusions of the literature review (Section 2.4), a major contribution of this research is the analysis of rail cost efficiency with new explanatory variables that have not been tested before, such as levels of contracting out, labour productivity and rolling stock characteristics.

In the case of urban rail cost efficiency, the literature is even more exiguous and the scarcity of available data even more acute. This research improves the current literature on urban rail cost efficiency by providing new results based on a new database that has not been used in other research: the CoMET and Nova database.

The database of the metro operational data comes from two consortia of metro system operators managed by the Rail and Transport Strategy Centre (RTSC), the Community of Metros (CoMET) and the Nova Group (Nova). This database is described in Section 4.1 and a map of the members can be seen in Figure 4.1. Section 4.2 describes the differences between balanced and unbalanced panel data. The next section explains how the financial data of the urban rail systems has been adjusted in order to enable international

comparisons. The different methods considered and the adopted approach in this research is explained in Section 4.3.

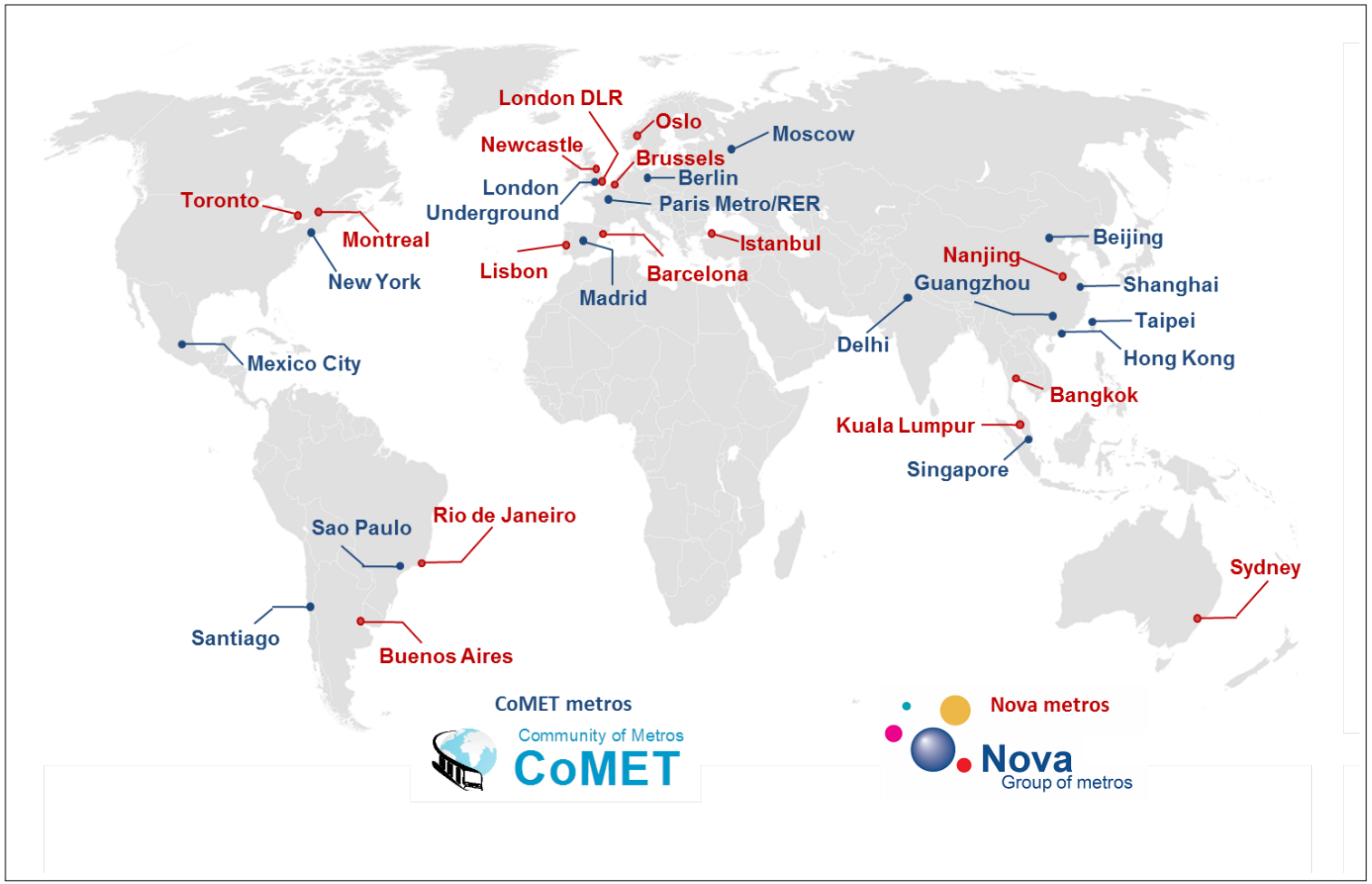
This chapter offers an overall description of the CoMET and Nova data. A detailed explanation of the data used, including descriptive statistics for each of the variables, can be found in their respective chapters. These are: section 5.3 for overall operating costs on page 118; section 6.3 for train service cost efficiency on page 166; section 7.3 for station service cost efficiency on page 193 and section 8.3 for rolling stock maintenance cost efficiency on page 217.

4.1 RTSC and CoMET-Nova Consortia

The metro operational data for this research comes from the CoMET and Nova consortia. The CoMET consortium was established in 1994 by the Rail and Transport Strategy Centre (RTSC) at Imperial College London. CoMET comprises more than a dozen of the largest metros in the world by number of passengers and network size, including world capitals such as London, New York, Beijing or Moscow among others. Subsequently, Nova was formed in 1998 for medium-sized metros following the success of CoMET. Nova members tend to be large operators although significantly smaller than the CoMET members, particularly in terms of passenger journeys.

The aim of the two consortia is to identify areas of potential improvement by means of comparing the performance of the metros. Consequently, the RTSC collects a large set of key performance indicators (KPI) related to metro performance every year. These KPIs are the backbone of the data

Figure 4.1: CoMET and Nova metro members (January 2015).



used in this research. The database spans over more than two decades and the areas of metro operations covered by the KPI range from finance, safety and reliability to efficiency and service quality. These indicators are precisely described in the CoMET-Nova KPI handbook, which is upgraded periodically to include new KPIs and redefine or clarify existing ones.

Figure 4.2: Scope of the research - Aggregate operating costs and other operating costs in the CoMET-Nova framework.

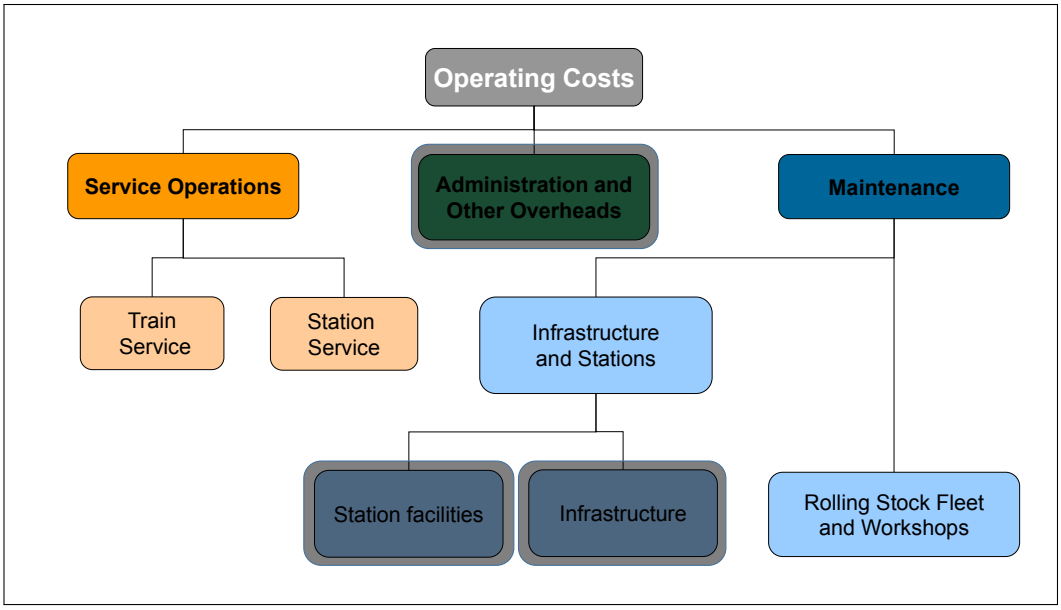
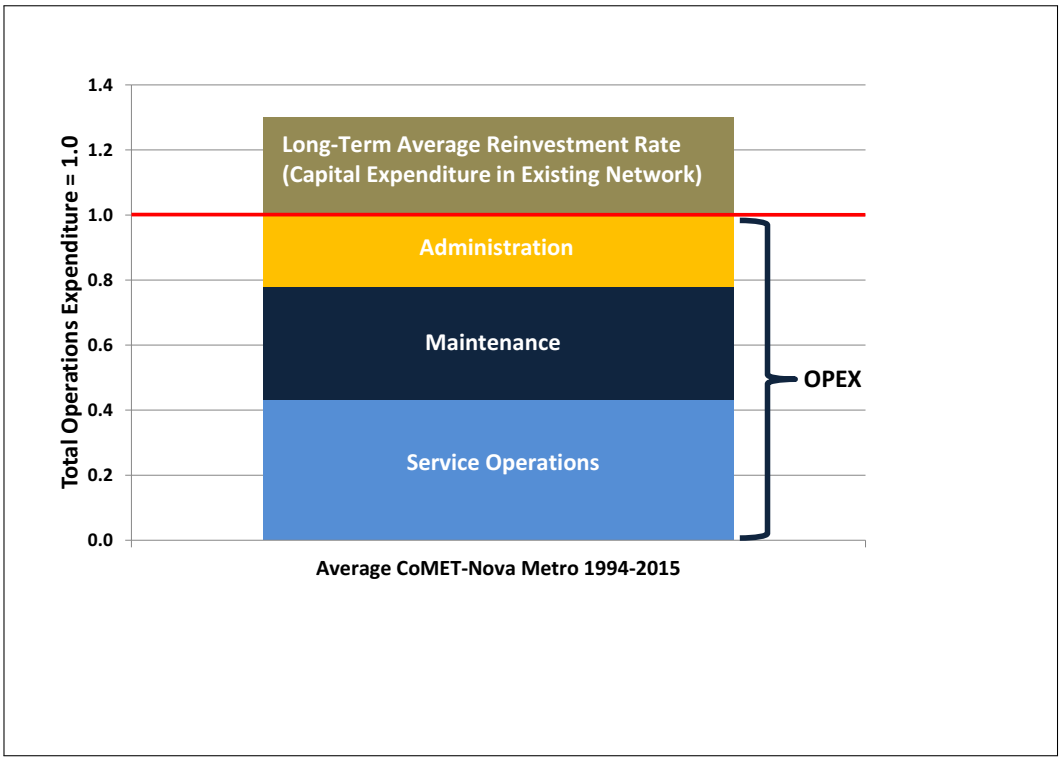


Figure 4.2 shows the operational division of the CoMET-Nova data set. This includes a breakdown of operations into service (orange boxes), maintenance (blue boxes) and general administration and other overheads (green box). This operational breakdown includes the cost data needed for the modelling of this research.

However, metros do have other costs beyond operating costs. This include the initial building costs and any corresponding debt service. Likewise, there are continuous capital expenses to further maintain the assets in a working

condition in the long term and also to improve the infrastructure endowment needed to improve the service. Figure 4.3 shows the average breakdown of costs for the CoMET-Nova metro since 1994 to 2015. The operating costs represent most of the overall costs of the urban rail systems, although it has been estimated by the RTSC that capital expenditure in existing network is around 30% to 35% of the operating costs. Therefore, despite the importance of capital expenditures, this research focuses on operating costs which represent the largest share of urban rail costs.

Figure 4.3: Average CoMET-Nova costs 1994-2015.



Regarding the CoMET-Nova consortia, the consortia comprises a group of more than 30 metros, including some of the most important metros in the world in terms of passenger demand, extension and years of service. The

basic indicators of the CoMET and Nova metros are shown in Table 4.1 and Table 4.2. The two consortia aim to be diverse in terms of metro size, age and location, which helps to ensure that the conclusions of this research can be more easily generalised to other rail systems not included in this panel of metros, as well as to other passenger rail operators.

The CoMET metros shown in Table 4.1 tend to represent large capital metropolitan areas with larger networks than the Nova members in Table 4.2, which usually represent smaller cities with less than 100 kilometres of network length. Likewise, annual demand for CoMET metros usually surpasses the 500 million passenger journeys and several members are above a billion of annual passenger journeys. On the contrary, none of the Nova members exceeds the 400 million threshold of passenger journeys per year.

The consortia members can also be grouped by age into three main categories. First of all, the old metros established before 1950 (e.g. London in 1863, New York in 1904, Madrid in 1919 or Moscow in 1935) which pioneered the development of urban rail systems. These metros are usually in world capitals, mostly in developed countries, and share common challenges such as the maintenance of old assets or the high price of labour. Secondly, a group of middle-aged metros built in the second half of the 20^{th} century (e.g. Lisbon in 1959, Mexico City in 1969, Honk Kong in 1979 or Singapore in 1987) which started the transition into mature systems and ageing assets. Lastly, the newer metros constructed in the last 20 years (e.g. Shanghai in 1993, Naples in 1993, Kuala Lumpur in 1996 or Bangkok in 2004) which are expanding their network rapidly and usually operate in growing cities of developing countries. Likewise, the consortia are also diverse geographi-

cally, with members from all the continents but Africa, as seen in Figure 4.1. The assessment of how representative are the CoMET-Nova members with respect to the metros of the world is not clear-cut due to the lack of a unique and universal definition of what is a metro. Therefore, given that it is not possible to define, unambiguously, the population of metros in the world, any assessment of how representative is a sample with respect to this population must be approximate. However, and overall, the members of the CoMET-Nova consortia represent a wide range of the different urban rail systems that operate worldwide when we consider the different city sizes, metro sizes, age of the urban rail systems and their geographical location. Therefore, this diversity of metro scale of operations, age and locations ensures a high degree of generalization in the results and findings reached in this research.

4.2 Balanced and Unbalanced Panel Data

The CoMET and Nova database provides a panel data set which offers several advantages over cross-sectional data (e.g. multiple metros observed only in one year) and time series data (e.g. one metro observed in several years). One of the main advantages of panel data is that it contains more meaningful variation - both between metro and within metro variation, while cross-sectional data and time series data contain only between and within metro variation respectively. This improves the ability of the analyst to control for unobserved heterogeneity. Panel data also increases the range of techniques available to deal with common data and estimation issues such as the endogeneity caused by correlated unobserved heterogeneity.

Table 4.1: Urban rail systems of the CoMET consortium

Metro	Country	Start of operations	Network length (km)	Number of stations	Passenger journeys (m)	Fleet
Beijing BMTROC	China	1969	225.2	171.0	848.7	1,697.5
Berlin BVG	Germany	1902	146.4	196.8	483.3	1,273.3
Delhi DMRC	India	2002	106.1	91.1	350.8	486.9
Guangzhou GMC	China	1997	172.0	116.8	717.0	1,026.4
Hong Kong MTR	China	1979	139.3	83.4	1204.0	1,434.3
London Underground	United Kingdom	1863	439.9	324.2	1106.2	4,079.0
Metro de Madrid	Spain	1919	265.6	281.9	650.0	2,006.2
Metro de Santiago	Chile	1975	89.2	95.4	547.2	824.0
Metro Sao Paulo	Brazil	1974	62.2	58.6	690.2	767.8
Mexico City STC	Mexico	1969	204.5	177.5	1547.1	3,036.8
Moscow Metro	Russia	1935	293.6	178.6	2471.7	4,523.4
New York NYCT	United States	1904	479.6	468.0	1592.1	6,310.3
Paris RATP (Metro)	France	1900	214.1	379.0	1524.3	3,509.9
Paris RATP (RER)	France	1969	115.1	66.9	470.7	1,145.4
Singapore SMRT	Singapore	1987	103.2	65.8	539.3	712.9
Shanghai SSMG	China	1993	303.5	236.4	1012.7	2,219.1
Taipei TRTC	Taiwan	1996	87.0	82.8	468.4	682.7

Note: Variables represent an average for the available years in the 2005-2012 period.
Units: Network length in kilometres, passenger journeys in millions and fleet as number of car (vehicle) units.

Table 4.2: Urban rail systems of the Nova consortium

Metro	Country	Start of operations	Network length (km)	Number of stations	Passenger journeys (m)	Fleet
Bangkok BMCL	Thailand	2004	19.7	18.0	65.5	55.9
Barcelona TMB	Spain	1924	93.4	128.8	377.4	753.5
Brussels STIB	Belgium	1976	39.9	63.0	141.0	256.4
Buenos Aires Metrovias	Argentina	1913	44.7	72.6	300.5	547.8
Metropolitano de Lisboa	Portugal	1959	38.6	50.9	177.5	338.0
Montreal STM	Canada	1966	64.8	72.3	233.2	757.1
Newcastle Nexus	United Kingdom	1980	78.0	72.8	38.7	180.0
Metronapoli	Italy	1993	11.7	12.1	25.4	66.3
Metro Rio	Brazil	1979	38.3	34.1	151.6	182.7
Toronto TTC	Canada	1954	61.9	68.0	306.4	686.3

Note: Variables represent an average for the available years in the 2005-2012 period.
Units: Network length in kilometres, passenger journeys in millions and fleet as number of car (vehicle) units.

However, one issue affecting the CoMET-Nova data set is unbalancedness. There are more data observations in recent years than at the beginning of the panel. The main reason for this is the success of the CoMET-Nova consortia and the number of metros that have joined CoMET and Nova in the last years. Overall, the data included in this report comprises an unbalanced panel data set for a 10 year period (2005-2014) with an annual observation for each member. The years preceding 2005 have been excluded as the number of metros was relatively low and the definitions of some indicators have been updated over time. Therefore, excluding years preceding 2005 does not reduce the number of observations significantly but does in fact increases data consistency.

Due to the unbalanceness of the panel data, the number of observations available for the research is far below 320; the potential maximum of 10 years for 32 metros. Firstly, the number of observations is lower because both the CoMET and the Nova group have been growing considerably over the last few years. For this research, only data for 24 members has been used and, as some members joined after 2005, some members only have data for a few years and not the whole time span of the research study (i.e. 2005-2014). Secondly, the above-mentioned processes of data verification imply that some data may be excluded from the analysis when its accuracy is dubious or it has not been confirmed by the operators. Thirdly, even though all the KPI items are precisely described in the CoMET-Nova KPI handbook, some members may not be able to provide data for certain data items due to differences between their internal accounting and the CoMET-Nova standard. In that case, the data is also excluded in order to ensure consistent comparability.

The process of translating internal reporting to CoMET-Nova KPI standards is usually developed over several years and, for this reason, some observations in the first years of some metros have been excluded from the data set to avoid data inconsistencies. Overall, this implies that growing groups tend to be more prone to present more missing data. This can significantly lower the number of observations available for the analysis, which sets an important constraint for econometric analysis as explained below.

However, the CoMET and Nova database presents a high standard of data quality as, during the years of benchmarking work carried out within the RTSC groups, a series of systematic data cleaning, verification and validation processes have been developed. The whole flow of data in the CoMET-Nova framework can be seen in Figure 4.4. The main source of standardization of the data is the CoMET-Nova KPI handbook with all the definitions of the data items collected from the members. The data is collected from the members by means of the CoMET-Nova questionnaires. Each member returns a questionnaire with the available data for all the CoMET-Nova data items. These questionnaires are assessed by the RTSC staff and any data item suspicious to be inaccurate or not compliant with the KPI Handbook is doubled checked with the corresponding CoMET-Nova members by means of follow-up questions. Next, once all the data has been verified, an annual KPI report es elaborated and disseminated to the members, which can assess again their data.

In the CoMET-Nova data framework shown in Figure 4.4 there are three main sources of feedback which lead to a continuous improvement of the data collected in terms of adding, dropping or redefining current KPI items col-

lected. First, CoMET-Nova members provide feedback on the KPI report to the RTSC staff, usually as an additional data check that the data provided is accurate. Besides this, there are two CoMET-Nova meetings per year where management staff from all the metros gather to discuss, among other issues, the data collected and its accuracy. Lastly, the RTSC has a data research and development task group who continuously improves and proposes changes to the current KPI handbook.

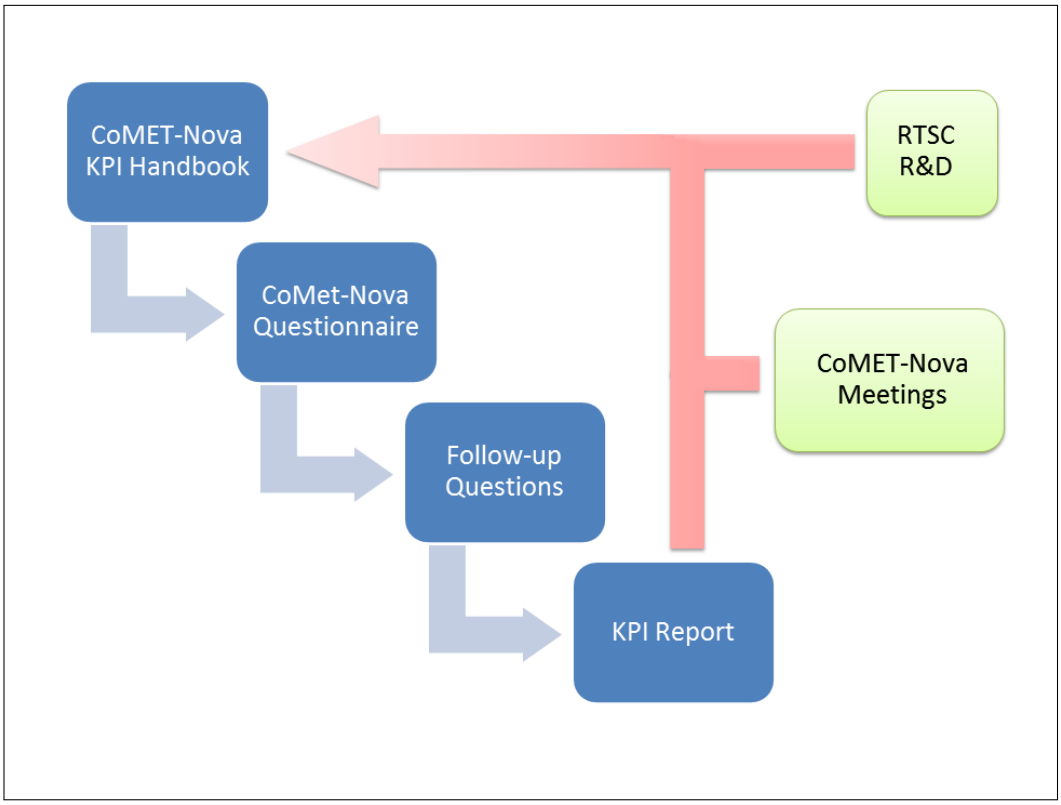
Therefore, after every annual iteration the quality of the data improves and better reflects the characteristics of the member metros. Consequently, the quality of the CoMET-Nova database ensures consistent comparability between the urban rail systems despite the size of the data available for this research may still place a constraint on the type of econometric methods that can be implemented.

Finally, mention that due to the sensitive commercial nature of the RTSC data and an existing confidentiality agreement, the data and the results presented in this research cannot compromise the anonymity of any of the CoMET and Nova members. Therefore, a detailed breakdown of the descriptive statistics by metro cannot be included in this thesis.

Regression analysis and completeness of the panel data

Regression analysis requires no missing data for any of the variables included in the regression. If one variable presents a missing value, the whole cross-sectional unit for that time period is not included in the regression analysis. Due to the data constraints described above, most of the metros have data for only 80% to 95% of all the variables in a certain year of the CoMET-Nova KPI

Figure 4.4: CoMET-Nova data framework

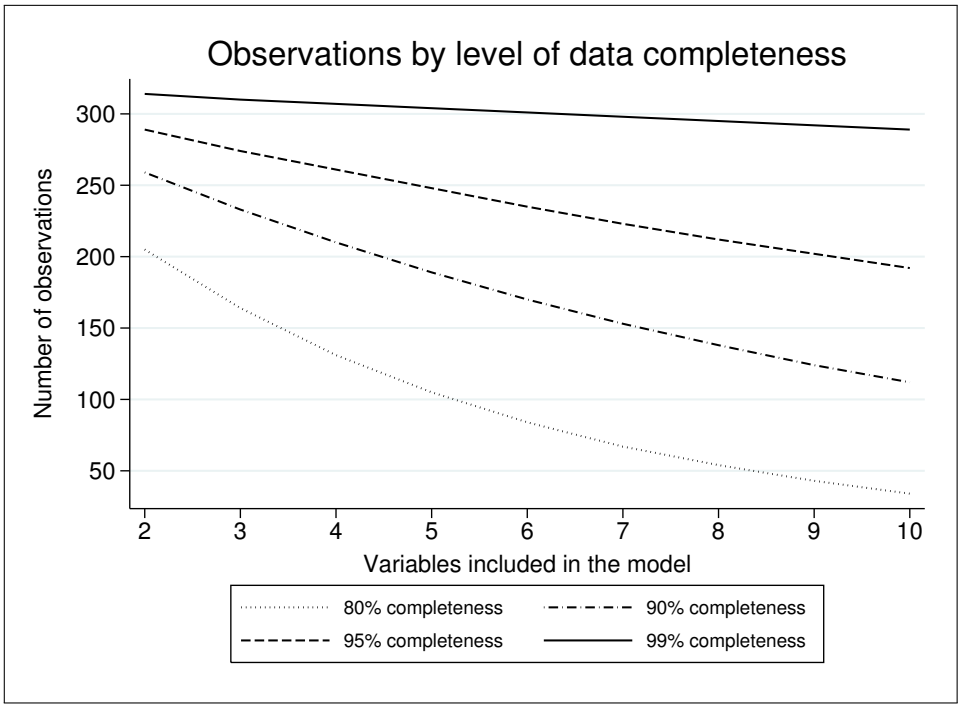


database. This imposes a considerable restriction on the number of variables that can be included in the model. Assuming a random distribution of the missing data, the observations of the final analysis would drop significantly as more variables are added to the regression.

This is particularly important when the metros have low completeness ratios, showing data for only 80% or 85% of the variables. Besides this, the effect is cumulative, so from an ideal optimal of 320 observations (i.e. 10 years from 2005 to 2014 with 24 metros), a panel with only 80% of data completeness would drop to less than 50 observations if 9 variables are included in the model, as can be seen in Figure 4.5. Likewise, given that the panel is

unbalanced and some members do not have data for the full time series, the number of observations actually available for econometric modelling is lower. On the contrary, if the average completeness level were 95%, the observations after including 10 variables in the model would be almost 200, as shown in Figure 4.5. As a result, the cumulative effect of completeness determines the number of observations and consequently influences the significance values of the explanatory variables included in the econometric models.

Figure 4.5: Completeness of data and number of observations.



Besides the effect of unbalanceness on the number of observations and the statistical significance of the explanatory variables, unbalanced panel data structures can be a source of estimation issues which can result in unrepresentative and biased model parameter estimates. For the purpose of statistical modelling, concerns arise when the incidence of non-response is

not random. If the missing observations are related to the design of the survey itself and/or are correlated with the variables of interest in the statistical models, problems may arise for statistical inference. In the context of this thesis, these problems could arise from non-zero correlation between metro cost efficiency level and the rate of missing data: worse performing metros might be less likely to appropriately collect and report data for the variables considered in the estimation of cost functions. However, after careful consideration of the data, there is no evidence that less efficient metros tend to report less data. Overall, this econometric work has tried to find an equilibrium between the importance of a balanced panel data structure and a sample size that guarantees the statistical significance and robustness of the analysis.

4.3 CoMET-Nova Financial Comparability

Financial operational data from the CoMET-Nova metros is reported in local currency units. Therefore, in order to compare performance between members, the financial data needs to be converted to a common standard. A first approach would be to apply a nominal exchange rate and convert all the local currencies into a reference currency (e.g. US Dollars or Great Britain Pounds). However, converting the financial data following the nominal exchange rate is not appropriate given the different purchasing power of each currency in its own country. Thus, a parity purchasing power (PPP) approach is adopted in order to convert currency local units of the financial variables to US Dollars.

The PPP exchange rates account for international differences in price levels and, hence, they can provide more meaningful cross-country comparisons of economic data. The PPP exchange rates equalise the cost of purchasing a given good or service in different countries and they are obtained from the ratio of the national (in local currency) prices of that good/service between any two countries. The source for the PPP exchange rates between countries has been the World Development Indicators published annually by the World Bank (2014).

It is vital to understand the influence of the PPP on the regression analysis estimates. The PPP index is a nation-wide indicator, so it refers to a national average weighted by the output of each of its constituent regions. Therefore, the purchasing power index assumes that metros are "buying" their inputs in the national market although, sometimes, the input prices vary between different locations in the same country. This effect can be more distorting in countries with huge differences between regions, such as large and developing countries. For instance, China PPP is a national average, a compromise between cities like Shanghai and much poorer rural areas in Central and West China.

This effect can be partially corrected if the metropolitan area served by the metro, or similar metropolitan areas, represents a large share of the national average, due to the larger leverage of the metropolitan areas in the national average. Likewise, although prices vary between different locations in every country, some inputs are subject to the same regulations nation-wide. For instance, labour prices may differ from city to city but the laws on minimum wage, safety or working hours are likely to be similar across the

nation, which equalises prices across the country.

Nevertheless, although the PPP approach has been chosen in this research and it is the standard for international comparisons of financial data, other parity purchasing indexes at the city level have been tested in order to increase accuracy of the PPP transformations. Particularly, the CoMET-Nova financial data was converted using the local price index reported by the Prices and Earnings Survey of the Union Bank of Switzerland (2012), which publishes price indexes for 71 world metropolises every three years. However, some of the metros in the CoMET-Nova consortia are not included in the survey, so it would lead to a loss of observations or inconsistency in the methods used to transform the financial data into PPP units. Besides this, the survey is triennial, so most of the years would have been extrapolated. For these reasons, the conversion of local currency units into comparable PPP US Dollars has been done following the PPP exchange rate published in the World Bank (2014) World Development Indicators.

5 — Operating Costs in Urban Rail Systems

Note: This chapter is an extended version of an article presented at the *Transport Research Board* - Session 17-0404 in Washington D.C. in January 2017.

5.1 Introduction

Urban rail provides a high-capacity and high-speed service in the metropolitan areas that is unparalleled by other means of transport. Metros are organisations with huge capital assets and large operating costs that seldom receive attention from the academic community. Besides this, given the lack of competition in the provision of mass and fast urban transit, most of the metros operate in monopolistic conditions, which can perpetuate an inefficient allocation of resources. Moreover, most metros are partially subsidised by public authorities, so the potential inefficiencies in metro services can have an impact beyond the metro operator and its passengers, and affect the whole community by subtracting resources from public budgets.

This chapter evaluates several hypotheses on the main cost drivers of the metros, the existence of economies of scale or density in the provision of its services and whether there is a background trend in their operating costs over time. It also estimates the production function of the main metro outputs and assesses whether metros behave like competitive firms following the duality theory between production and cost functions explained in Section 5.5.

5.2 Literature Review

The modern research on rail efficiency has a starting point with Caves, Christensen and Swanson (1980), who analysed the total factor productivity (TFP) of US rail operators using econometric techniques and multiproduct cost functions. The TFP approach has been used repeatedly to analyse the efficiency of rail operators (Cantos et al., 1999) but less often for urban rail (Graham, 2008). Besides the predominance of efficiency research on freight and national rail operators (Oum and Yu, 1994), most of the research on urban rail efficiency has been carried out using the nonparametric approach of Data Envelopment Analysis (DEA) such as the research by Graham (2008) or Santos et al. (2013). Santos et al. (2013) performed a DEA analysis with 37 urban railways while considering two types of outputs, passenger journeys and vehicle km, and found that metros with larger networks and more stations tend to be less efficient. More recent research (Tsai et al., 2015) has found opposite results with more efficient metros having larger networks and more stations. Likewise, Tsai et al. (2015) also found that high levels of passenger demand network length are fundamental for urban rail efficiency. The

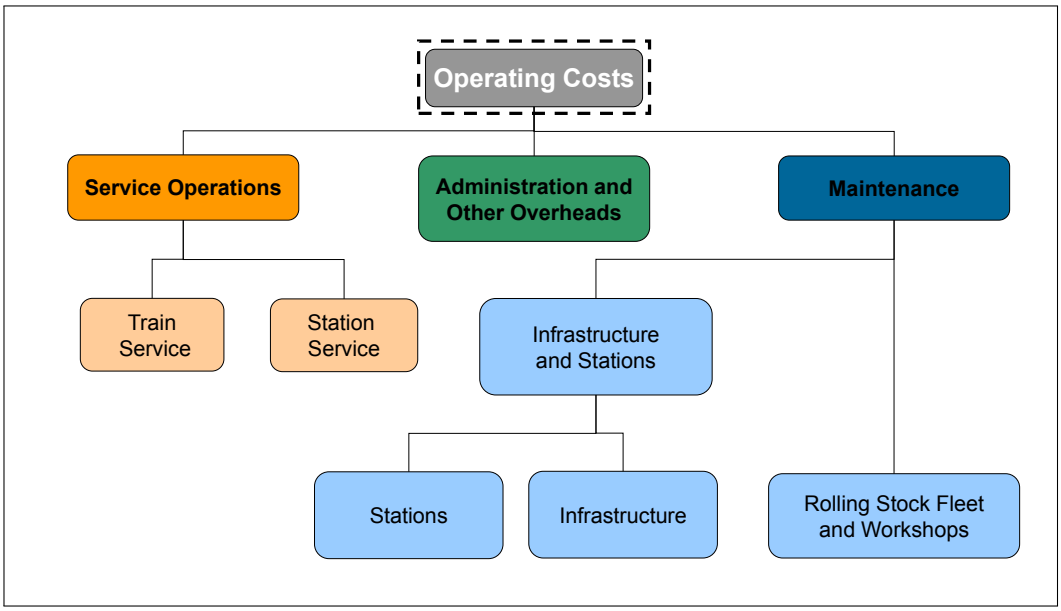
controversy surrounding the effects of large networks on the efficiency of the rail operators has been one of the main research topics in the field. Keeler (1974) already distinguished between economies of scale and economies of density in railway operators. However, there are other approaches more suitable for estimating rail economies of density and scale by means of estimated cost elasticities. For instance, Perelman and Pestieau (1988) estimated an econometric model where the performance is regressed on network length per squared km and other operational variables such as the percentage of electrified track. However, despite the relatively abundant corpus of research using regression analysis (e.g. Coelli and Perelman, 1999, Farsi et al., 2005), there are few studies on urban rail efficiency, and the most recent ones are on reliability (Canavan et al., 2015) but not the actual financial efficiency of the metros. Consequently, this study makes an important contribution by means of using new methods and testing new hypotheses in the academic field of urban rail efficiency research.

5.3 Data and Research Hypothesis

The cost functions estimated in this research are based on an unbalanced panel of 22 metros across 9 years (2005-2014) from the CoMET and Nova consortia explained in detail in Chapter 4. The two consortia have grown considerably over recent years, increasing the number of metros included in the panel but also limiting the time series available. Besides this, missing data for certain variables also reduced the final number of observations available. In this research on operating costs, 141 metro-years have been

included even though a full panel of 22 metros for 9 years would have yielded 198 observations.

Figure 5.1: Aggregate operating costs and other operating costs of a typical CoMET-Nova metro operator.



The operating costs comprise service, maintenance and administration costs as described in Figure 5.1. Service costs correspond mainly to the labour and energy costs of moving the train along the network (i.e. train service, as seen in Chapter 6) and operating the stations (i.e. station service, reviewed in Chapter 7). Maintenance costs refers to all the assets used by the metro (i.e. track, stations and rolling stock, the latter covered in Chapter 8). Lastly, administration costs comprehend all the engineering and project costs and general administration costs for the metro.

Besides these general constituents of operating costs, operating costs also include overall cost not allocated to any particular subcategory. For instance, irrecoverable input value added tax, bad debts or increases and decreases

Table 5.1: Operating costs dependent variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max
Operating costs per car km	5.63	5.39	3.29	2.72	19.89
Operating costs per passenger journey	1.19	1.1	0.56	0.63	3.04
Operating costs per train hour	1056.13	1046.43	467.73	455.17	2584.31

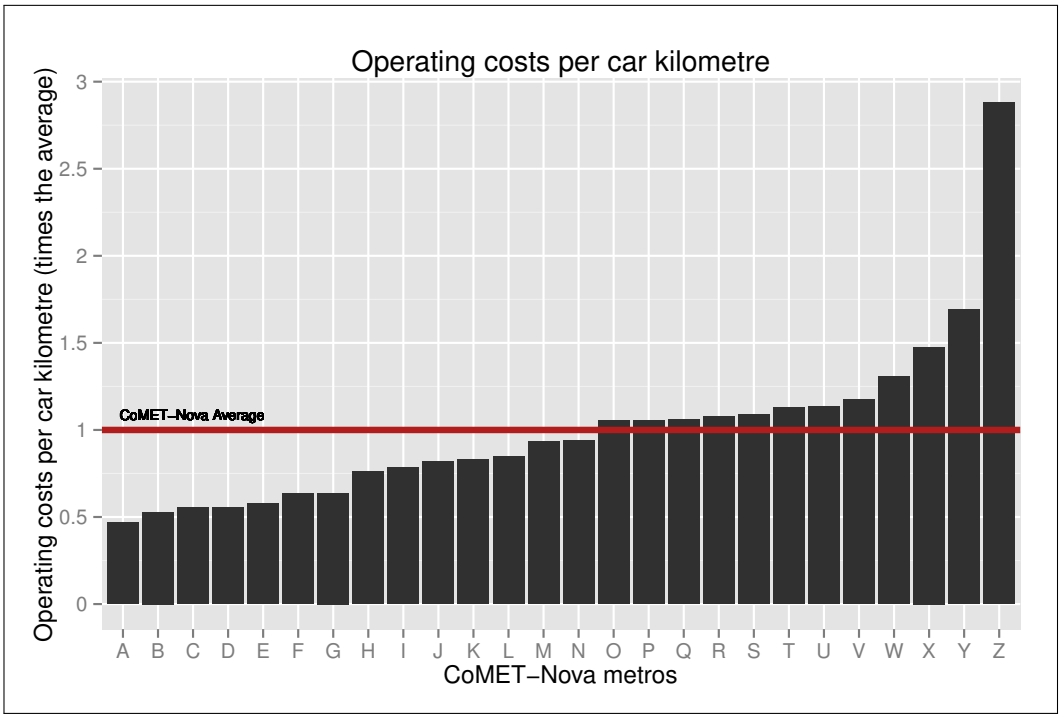
Note: Costs in Parity Purchasing Power (PPP) U.S. Dollars. Obs.: 141.

in provisions for obsolete or missing stock are also included in the operating costs category of the CoMET-Nova KPI framework. On the contrary, operating costs excludes all types of capital expenditure, either related to depreciation or asset upgrades. Likewise, interest costs and extraordinary expenses (i.e. one-off costs such as the closure of a station) are also excluded from the operating costs category.

The operating costs are normalised with three different output variables in order to enable easier comparisons between metros of different sizes. The operating costs are rationalised by car kilometres, passenger journeys and train hours as shown in Table 5.2. Figure 5.2, 5.3 and 5.4 show the same operating unit costs although the graphs have been indexed in order to respect the CoMET-Nova confidentiality agreement. In any case, the coefficients obtained in the econometric models of this chapter for the normalised costs per output can easily be transferred to cost elasticities for the total costs as explained in the Annex (Section 11.1 on Page 283).

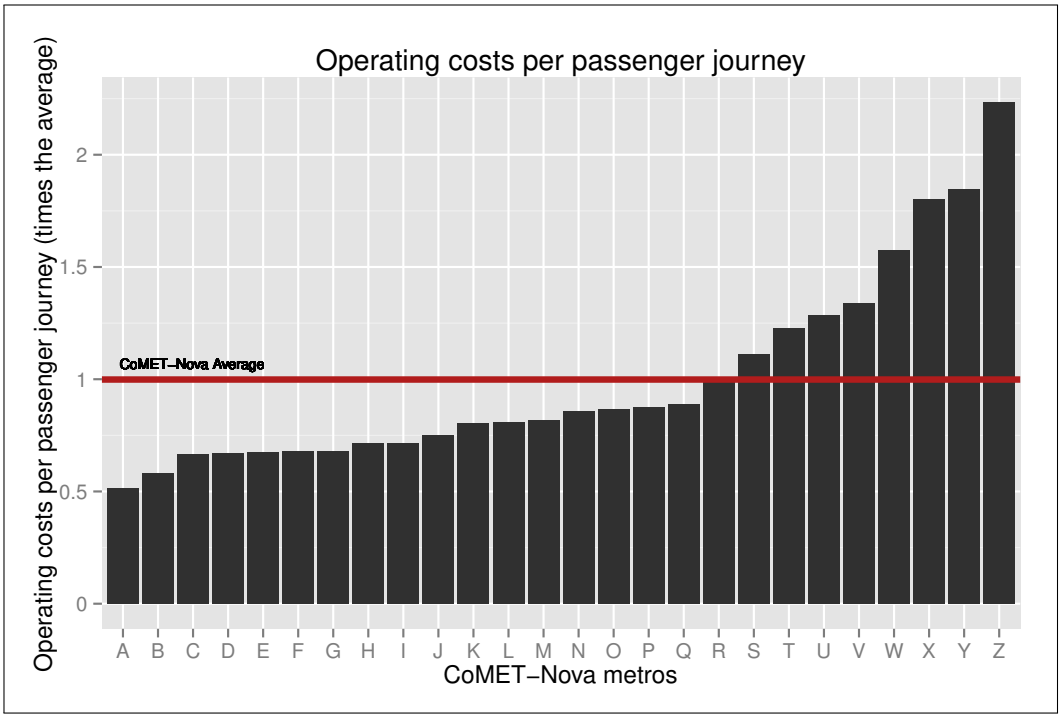
The average operating cost per car kilometre for the CoMET-Nova metros

Figure 5.2: Operating costs per car kilometre.



included in this research is 5.63 US$ with a standard deviation of 1.50 US$. Therefore, most of the metros are close to the mean, although some metros are notably expensive with a maximum of almost 20 US$ for one car kilometre in a particular metro-year. Explaining the causes of this variation is the main aim of this research. Average operating costs per passenger journey are 1.19 US$, which is slightly lower to the average cost of a fare ticket in most cities (i.e. usually around 2 US$). However, all investment and capital costs are excluded from this category, so most of the metros are, to a greater or lesser extent, subsidised by public resources. In the case of operating costs per passenger journey, the maximum is only 5 times the minimum operating costs per passenger journey, a much lower dispersion than in the case of operating costs per car kilometre.

Figure 5.3: Operating costs per passenger journey.

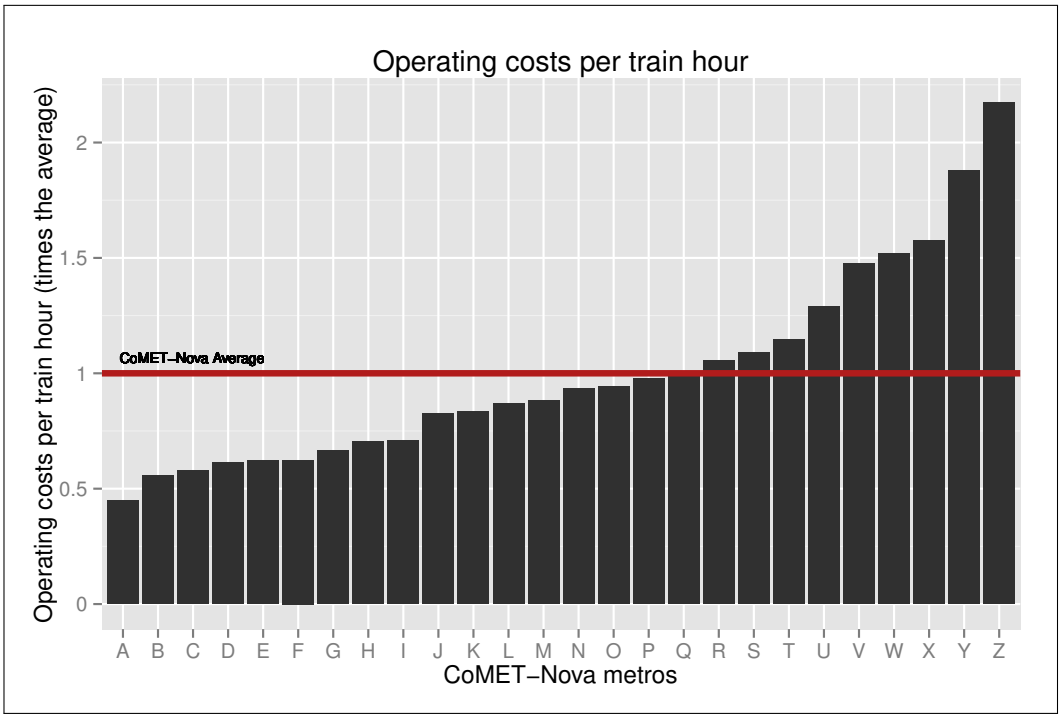


Operating costs per train hour are on average around 1000 US$, with a standard deviation of less than 500 US$. The ratio of maximum to minimum is also approximately 5, a similar dispersion rate of the extreme cases as in the case of operating costs per passenger journey.

The econometric models presented in this chapter aim to explain these operating costs per car kilometre, passenger journey and train hour with the explanatory variables shown in Table 5.2. A detailed definition of these explanatory variables can be found in the Annex (Table 11.5 on page 290).

The first explanatory variables considered are the output that normalises each of the models; car kilometres, passenger journeys and train hour. Output variables are included in the models to assess the existence of economies of density in the urban rail operations. Economies of density, or a more inten-

Figure 5.4: Operating costs per train hours.



sive utilisation of the assets, are expected in industries with large fixed costs such as the rail sector. Economies of density have been frequently found in rail services since the seminal work of Keeler (1974), although the evidence on urban rail is more scarce. Besides this, the effect of the returns to density is expected to be stronger in those outputs where the marginal cost of an additional unit of output are negligible (i.e. passenger journeys), whereas the returns to density on outputs such as car kilometres and train hours would probably be moderate.

The scale of the urban rail network is also included in the analysis by means of variables on the network route kilometres and the number of stations served by the metro. Average network length for the CoMET-Nova metros included in the analysis is 112 km, even though the largest network

Table 5.2: Operating costs explanatory variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max	Obs
Car km (m)	104.129	105.346	182.698	8.562	776.35	N = 141
Passenger journeys (m)	491.501	605.857	589.644	35.7	2572.9	N = 141
Train hours (m)	0.555	0.598	0.75	0.079	3.01	N = 141
Wages (PPP US$/h)	32.53	35.782	16.1	11.575	75.217	N = 141
Electricity price (PPP US$/MWh)	156.814	159.921	99.681	49.161	1030.105	N = 141
Total staff hours (m)	13.628	14.953	17.437	1.36	67.553	N = 141
Total electricity (MWh)	459.231	485.332	542.014	63.68	2271.5	N = 141
Network length (km)	111.987	102.59	141.189	19.74	548	N = 141
Number of stations	89.726	81	95.45	18	424	N = 141
Staff outsourcing (%)	0.275	0.319	0.188	1	0.638	N = 141
Incidents per car km	5.303	5.486	17.481	0.054	81.109	N = 141
Opening hours per week	132.376	133	9.937	110.5	168	N = 141
Subsidy (% of opex)	0.91	1.006	0.382	0.383	2.091	N = 141
Network age (years)	47.879	39	35.9	2	151	N = 141

Note: Units in PPP US$ refer to Parity Purchasing Power (PPP) U.S. Dollars.

is over 500 km long. Likewise, the metros included in the sample of this research range from 18 to 424 stations with an average of 89 stations per urban rail system. Presumably, metros with larger networks and more stations may show lower costs per output delivered, although larger metros tend to be more suburban, and consequently, use their assets less intensively. Thus, the net effect of scale on urban rail efficiency remains unclear.

As for any analysis based on a cost function, the input prices and quantities are key to understanding cost efficiency. In this analysis, both the price of labour and electricity have been considered in the econometric models. Wages refer to the whole gross cost of employed labour for the urban rail operator. On average, CoMET-Nova metros pay 30-35 US$ per hour worked, with a standard deviation of 16 US$. Electricity prices per MegaWatt/hour (MWh) show a larger variation, where the maximum price per MWh for a given metro-year is 6 times the average electricity price as shown in Table 5.2. Total input quantities used are also included in the main analysis on an hour level, which takes into account the differences between full-time and part-time employees.

The percentage of staff hours outsourced or contracted out has also been considered in the econometric analysis of cost efficiency. In theory, outsourcing would enable the metro to use more specialised labour and increase competition in labour hiring, leading to higher cost efficiency and lower costs. Some metros in the panel do not outsource any of their labour hours (i.e. minimum 0%) although the average and median percentage labour hours outsourced is 27% and 31% respectively as shown in Table 5.2.

An additional variable included in the analysis is reliability. Reliability is

measured as the ratio of incidents per million car km where an incident is a delay in service of 5 or more minutes regardless of the cause. Higher reliability would lead to a service with less interruptions and, thus, larger output (e.g. car km). Consequently, operating costs per output, such as operating costs per car kilometre, would be lower. In the case of the CoMET-Nova metros, the average incident rate per million car kilometres is around 5, so there are 5 delays in service of more than 5 minutes per million car kilometres produced. However, the standard deviation is 17, so there is a wide variation in the reliability rates of the CoMET and Nova metros.

Some factors considered in this analysis may not be related to rail service operations, but may rather be a constraint to operations such as maintenance. For instance, opening hours per week has also been included in the analysis, as longer opening hours constrain maintenance practices, leading to higher costs and also increasing the labour expenses entailed in night shifts. The metros included in the analysis open, on average, 133 hours per week; roughly 19 hours per day. However, some metros open 24 hours a day, as shown in the maximum in Table 5.2. Overall, smaller cities tend to have fewer opening hours per week, which may have an impact in terms of signalling for the lower labour costs of smaller cities.

Most public transport modes receive some sort of public subsidy. In the case of this cost efficiency analysis, subsidy is defined as the operating costs divided by the commercial revenue. This equals the revenue directly derived from the commercial operations of the metro, as non-commercial revenue is influenced by the local regulatory framework and has nothing to do with the actual operations of the rail service. In this case, higher numbers in the

subsidy variable shown in Table 5.2 would mean larger subsidies. The mean (e.g. 0.91 for a 91% ratio of cost recovery) and median (1.01) subsidy show that operating costs are usually covered by the commercial revenue, although the standard deviation in the recovery ratio is 38%. However, this implies that most metros only cover operations and that the commercial revenue would not leave any room for investment and depreciation. Likewise, it is hypothesised, in line with traditional research (e.g. Bly et al. (1980), Oum and Yu (1994), Pucher et al. (1983)), that larger subsidy levels would lead to lower cost efficiency as the metro's costs would not correspond to the value perceived in the revenues.

Lastly, the average age of the metros in the panel is 47 years, whereby age is measured as the number of years since the beginning of operations. However, the age of the metros for a given year range from 2 to 151; the latter being the case of the London Underground. Network age could be very important, as older metros may face constraints from the initial design of the network that decrease cost efficiency of the metro in the long run (e.g. track gauge or tunnel dimensions). The effect of age is measured in years in the main model although another econometric model shown in the annex also groups the metros into new, mature and old metros in order to see the effect of age by cohorts. These additional models are shown in the Annex (page 295) and explained in the results Section 5.4.

Other explanatory variables considered but not included

As part of the econometric modelling exercise, other variables have been tested that have not been included in the final cost function models. For in-

stance, capital interest rate (i.e. central bank rate) was tested as a measure of the cost of capital as lower costs of capital lead to more investment. However, urban rail systems usually receive support from government to invest and the actual price of capital is not a key factor to determine investments. Likewise, metros may invest not to minimize costs or improve profitability but for other reasons. For example, line extensions and reliability improvements may be more motivated by a neighbourhood regeneration agenda from the public authorities than by metro's optimising planning. Besides this, even if the cost of capital was not distorted by public intervention, the effect of investment may be lagged in time, so present investment may explain future cost savings. Moreover, present investment may increase disruptions and actually increase current operating costs, so the effect is not clear and longer panel data set may be needed to assess causality of investments.

Share of the metro network with driverless operation was also considered as a potential regressor. However, there is very little data on this variable and there are also different types of driverless operations. In the case of the CoMET-Nova metros, only few metros have full-driverless lines with no requirement of on board staff. Therefore, the inclusion of this variable mainly acts as a binary variable to control for these few metros with driverless lines. Consequently, as driverless operations expand to other metros, this variable would probably have an impact in metro aggregate operating costs.

Labour related factors considered in the CoMET-Nova KPI framework have also been tested such as own staff labour absenteeism. In this case, absenteeism includes both justified absences (e.g. illness, holidays or injuries) and also non-justified (i.e. lateness). However, this variable is mainly affected

by the local laws (e.g. holiday entitlement) which are out of control from the metro management. Likewise, measurement of absenteeism is not rigorously measured in some metros so the variable is not accurate enough to be included in the model.

Lastly, the aggregate operating cost model has also considered the average commercial speed as a potential regressor to explain operting costs per car kilometre and per passenger journey. However, speed is strongly correlated with the distance between stations, with longer average distance between stations being correlated with higher speeds. Therefore, due to the inclusion of number of stations and network length as regressors, speed would be mostly collinear and insignificant in the analysis.

5.4 Cost Function Results

This section presents the econometric cost models for the operating costs per car kilometre, per passenger journey and per train hour. Other econometric models are shown in the appendix, including models with age categorised into binary variables (Table 11.9 in page 295) and models without input quantities (Table 11.13 in page 300). A GMM reduced cost function model is also shown in Table 11.1 in page 285. All these models confirm the insights and coefficient estimates shown in Table 5.3 unless noted otherwise.

The coefficients for the explanatory variables reviewed in section 5.3 are presented in logarithms ($[Ln]$) so the coefficient estimates represent cost elasticities. For time variables like network age or time trend, the coefficient estimate represents the effect of an additional year. The estimated mod-

els follow the functional form shown in Equation 5.1, where the operating cost category (i.e. C) is regressed on the set of explanatory variables (x_j) described in the previous section for variables in logarithmic form (i.e. x_j) and non-logarithmic form (i.e. x_k). The models are estimated with a random effects estimator as the Breusch-Pagan test discarded the hypothesis of constant variance across entities (Table 11.17 on page 304).

$$LogC = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \sum_{k=1}^{K} \beta_k x_k + \alpha_i + \varepsilon_{it} \tag{5.1}$$

The first three independent variables in Table 5.3 assess the existence of economies of density in the aggregated metro operation costs. All three coefficients for car kilometres, passenger journeys and train hours show negative estimates, with a 99% significance level suggesting the existence of economies of density. The coefficients for car kilometres and passenger journeys are below -0.80, showing stronger economies of density than expected (i.e. -0.516 for passenger miles in McGeehan (1993)), probably due to a signalling effect. Those metros more oriented to producing car kilometres or passenger journeys may be more specialised at producing that type of output. With regard to train hours, the coefficient estimate is moderate (-0.385) but still significant at a 99% confidence level. These three coefficients follow the previous evidence in the academic literature established by Keeler (1974) and also confirm the existence of strong economies of density in urban rail operations in line with the previous evidence on returns to density for urban rail systems (Graham, 2008, Graham et al., 2003c). This can also be understood as urban rail systems being underutilised, which is sensible given the bimodal

distribution of the metro demand with two commuter peaks and a long set of valley hours.

In the estimation of the cost functions, the input-related explanatory variables are key to understanding costs. In this case, the coefficient estimates for input prices and quantities are statistically significant in almost all the cases. Wage cost elasticities are positive, ranging from 0.31 to 0.17 for the model on operating cost per car kilometre and per train hour respectively. This would imply that a 10% increase in wages is associated with a 1.7-3.1% increase in operating costs. These wage cost elasticities shown in Table 5.3 are slightly lower than expected (i.e. 0.47 of wage cost elasticity in McGeehan (1993)) as labour usually represents a higher input share for most of the metros. However, this is compensated by the positive estimates for total staff hours hired, which range from 0.54 to 0.33.

A key question is to what extent higher wages lead to higher productivity, usually by means of fewer labour hours hired. If higher wages do not increase productivity, and there is no adjustment to the labour hours hired, then the difference between the wages coefficient estimated in the model with fixed labour hours quantities (i.e. $\beta_{fixed_L_ckm}$ in Table 5.3) and the wages coefficient estimated in the model with variable labour hours quantities (i.e. $\beta_{flexible_L_ckm}$ in Table 11.13) should be null. In econometric terms this would mean that the difference between the two is statistically insignificant. Thus, the null hypothesis is that this difference between these two coefficients is statistically insignificant; $H_0 : \beta_{fixed_L_ckm} - \beta_{flexible_L_ckm} \simeq 0$.

In the case of the operating costs per car kilometre, the difference between the two cost elasticity estimates is equal to 0.159 (i.e. 0.311 - 0.152 = 0.159),

Table 5.3: Operating Cost Model Results - Random Effects

	Operating Costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.906***		
	(0.077)		
Pax. Journeys [Ln]		−0.833***	
		(0.051)	
Train hours [Ln]			−0.385***
			(0.074)
Wages [Ln]	0.311***	0.276***	0.172***
	(0.054)	(0.054)	(0.064)
Elect. Price [Ln]	0.162***	0.161***	0.113***
	(0.023)	(0.022)	(0.025)
Staff hours [Ln]	0.544***	0.482***	0.339***
	(0.057)	(0.060)	(0.081)
Total elect. [Ln]	0.471***	0.411***	0.021
	(0.076)	(0.064)	(0.076)
Network length [Ln]	0.131	0.219**	−0.187
	(0.094)	(0.088)	(0.123)
Number of stations [Ln]	−0.402***	−0.440***	−0.101
	(0.097)	(0.098)	(0.148)
Staff outsourcing [Ln]	−0.348*	−0.353**	−0.432*
	(0.177)	(0.178)	(0.256)
Incidents per car km [Ln]	0.034***	0.039***	0.043***
	(0.010)	(0.010)	(0.011)
Opening hours per week [Ln]	−0.058	0.095	−0.309
	(0.276)	(0.271)	(0.333)
Subsidy (% of opex) [Ln]	0.485***	0.536***	0.530***
	(0.035)	(0.036)	(0.040)
Network Age (years)	0.003***	0.003***	0.002
	(0.001)	(0.001)	(0.002)
Time trend	0.006***	0.006***	0.010***
	(0.002)	(0.002)	(0.003)
Constant	−11.420**	−11.897***	−17.081***
	(4.668)	(4.535)	(6.314)
Rho	0.8877	0.91	0.9823
Observations	141	141	141
R^2	0.823	0.870	0.715
Note:			*p<0.1; **p<0.05; ***p<0.01

which implies that the increase in costs when wages increase by 10% is 1.59% higher when metros have fixed labour hours. Thus, out of the initial cost elasticity estimate (0.311 in Table 5.3) only the difference (i.e. $0.311-0.159 = 0.152$) corresponds to gains in productivity associated with higher wages.

In order to reject the null hypothesis, the difference of the coefficients should be statistically significant. This is assessed by means of a t-test of the difference as shown in Equation 5.2. With regard to the operating costs per car kilometre, the t-value for the difference of the coefficients is 2.87, which is above the 5% significance level threshold for a t_{n-k} distribution with 141 degrees of freedom (i.e $t_{141,5\%} \simeq 1.977$).

$$\theta = \frac{\beta_{fixed_L_ckm} - \beta_{flexible_L_ckm}}{(\sigma_{fixed_L_ckm} + \sigma_{flexible_L_ckm})/2} = \frac{0.311 - 0.152}{(0.054 + 0.057)/2} \simeq 2.87 \qquad (5.2)$$

Consequently, and in the case of the operating costs per passenger journey, when wages increase, around half of the increase does not have any effect on cost efficiency. This may be due to higher labour productivity with some adjustment of the labour hours hired. However, approximately the other half does increase operating costs for the average CoMET-Nova metros. The results for the wage coefficient estimates for operating costs per passenger journey are very similar, with a difference of 0.169 and a t-value of 2.91. However, in the case of the operating costs per train hour, the difference is smaller, 0.95, and the t-value (i.e. 1.55) is not statistically significant at a 90% level.

With regard to other inputs, the electricity price cost elasticities shown

in Table 5.3 present positive coefficients, although they fall below 0.20, which is lower than expected. This is partially compensated by the total electricity coefficients, which are positive and above 0.40 for the operating cost per car kilometre and per passenger journey. However, the total electricity cost estimate is positive but insignificant for the train hour model. In the case of electricity, the range of prices and different efficiency levels of the rolling stock may distort the estimation of the coefficient estimates.

In addition to the analysis on returns to density, the results on the scale of the urban rail system are equally important. The scale of the network appears to have a mixed effect as shown in Table 5.3. Larger networks seem to increase operating costs, although the effects are only significant for the operating cost per passenger journey model. This indicates that larger networks are usually more suburban, leading to lower urban densities and lower passenger demand. On the contrary, a higher number of stations decreases the operating costs per car kilometre and per passenger journey. The coefficient for operating costs per passenger journey is in line with the main hypothesis, as more stations increase the accessibility of the network to passengers. However, the effect of more stations on operating costs per car kilometre is unexpected as more stations, when the network length is fixed, lead to more frequent stops, lower speeds and fewer car kilometres. With regard to the operating costs per train hour, both coefficient estimates for network length and number of stations are statistically insignificant, so there is a null overall effect. This follows the general view in the academic literature on the absence of economies of scale in rail operations.

Besides the analysis of labour wages and labour quantities, labour out-

sourcing can also lead to higher cost efficiency. In the coefficient estimates shown in Table 5.3, the outsourcing of the labour hours shows negative coefficients, following the research hypothesis shown in section 5.3. The estimates range from -0.432 to -0.348 and are significant at a 90% confidence level. This implies that outsourcing is a source of cost efficiency for most of the CoMET-Nova metros and the resulting impact can be great enough to compensate labour wage rises. However, although these results follow the expectations held by economic theory, the other result chapters of the thesis will assess whether there are substantial differences in cost efficiency gains from outsourcing depending on the operational area of the metro.

The model shown in Table 5.3 also evaluates the role of service reliability in the operating cost efficiency. The estimates show that less reliable metros, those with more incidents per car kilometre, tend to have slightly higher operating costs. The effect is moderate (0.034 to 0.043) although statistically significant at a 99% confidence level. This shows that reliability not only increases the service delivered to the passengers but it is also a potential source of cost savings.

Opening hours per week appear to have a null effect on the cost efficiency of operating cost for the CoMET-Nova metros. Although longer opening hours may set a constraint on maintenance practices and impose higher labour costs, the coefficient estimates are insignificant for all metros. On the other hand, the most significant cost driver to explain operating cost efficiency is subsidy level, measured as the share of operating costs not covered by the operating commercial revenue. The coefficient estimates are highly significant and they show that increasing the subsidy levels by 1% is linked to

about 0.5% more operating costs *ceteris paribus*. Thus, despite there being good reasons to subsidise public transport, such as economies of scale and private transport externalities (Preston, 2008), subsidies may be a strong source of inefficiencies that only represent a short-term solution while seriously compromising long-term financial sustainability. However, this thesis suggests the existence of a relation between subsidies and lower cost efficiency, although whether one causes the other is unclear. A longer panel of metros with more metro-year observations would be needed to identify the causality relation between subsidy and the increase in operating costs.

Lastly, network age presents statistically significant coefficients for operating costs per car kilometre and per passenger journey, although the effect per additional year since the start of operations is negligible. However, Table 11.9 (page 295) shows the coefficient estimates for the same models with the metro age broken down into 3 groups; old for metros inaugurated before 1945, mature for those who started operations between 1945-1980 and new metros inaugurated after 1980. This latter model shown in Table 11.9 presents large and highly significant coefficients for old and new metros, suggesting that the effect of age on operating costs efficiency may have a quadratic form. This implies higher costs for old and newer metros, in particular, between 26% to 36% for old metros and between 19% to 49% in new metros. This evidence suggests that new metros may be more expensive than expected at the beginning of operations although their efficiency may increase progressively as the metro matures. Likewise, based on this research, mature metros could anticipate that cost efficiency would deteriorate as assets become more obsolete.

Table 5.3 reports the goodness of fit of the models, which are above 0.70 for all models. This implies that the econometric models are able to explain most of the variation in the 141 observations of the CoMET-Nova panel data set. The *rho*-coefficient refers to the ratio between the standard deviation of the residuals within metros over the standard deviation of the residuals for the overall observations. In Table 5.3 the *rho*-coefficients are around 0.90, implying that almost all of the unexplained variation in the data is between metros and not within metros. Likewise, given that the econometric models explain the variation in the operating costs, these econometric models mainly explain the differences in the level of operating costs between different metros rather than the particular evolution of the operating costs for a particular metro from year to year.

5.5 Cost and Production Function Duality

Cost functions are useful for studying the efficiency of urban rail systems, although this efficiency can also be studied by means of production functions. In fact, according to economic theory, the conclusions of both approaches should be equivalent and the comparison of both approaches may provide a way to verify the accuracy of the coefficient estimates. The symmetry between the production function and the cost function for any particular firm or sector is based on the economic principle of duality. Following this principle, any characteristic observed in the production function of a particular firm or sector (e.g. economies of scale) must have its corresponding properties in the cost function of that same firm or sector. Therefore, given a certain

production function for an efficient firm, the corresponding cost function can be calculated by means of cost minimisation. Likewise, any given cost function for a particular firm or sector can yield the corresponding production function by means of a product optimisation calculation.

The economic principle of duality of the firm production and cost functions is analogous to the one that applies to the individual consumer, where the consumer's utility maximisation and expenditure minimisation are mirror images of its optimising behaviour (Varian and Norton, 1984).

In the particular case of the firm, a classical two-input labour (L) and capital (K) total cost function is shown in Equation 5.3, where w represents the wages paid to labour and r the capital rental rate. Although the explanation applies to total cost functions, it can be easily applied to unit cost functions as shown in Section 11.1.

$$C(Q, w, r) = wL(Q, w, r) + rK(Q, w, r) \tag{5.3}$$

The optimal demand for labour and capital can be derived, where optimal implies the quantity of inputs employed by a firm to minimise the cost of producing a certain amount of output. As shown in Equation 5.3, the demand for a production factor (e.g. labour, L) would depend on the level of output required (i.e. Q) and the factor price of that input factor (i.e. labour wages, w). However, equally important is the price of the other production factor (i.e. r as price of capital) as higher prices of one input would lead to more consumption of other input. Consequently, a change in one factor price would

have a direct and an indirect effect as shown in Equation 5.4.

$$\frac{\partial C(w,r,Q)}{\partial w} = \underbrace{L^*(w,r,Q)}_{\text{Direct effect}} + \underbrace{\left[w\frac{\partial L^*(w,r,Q)}{\partial w} + r\frac{\partial K^*(w,r,Q)}{\partial w} \right]}_{\text{Indirect effect}} \tag{5.4}$$

However, by means of the envelop theorem, it can be assumed that the effect on the overall costs is the direct effect itself without the addition or subtraction caused by the indirect effect (if the latter is set to nought as described in Shephard's lemma (Varian and Norton, 1984)).

From a production function to a cost function

In the case of a particular production function, such as the two-input Cobb-Douglas (i.e. $Q = AL^aK^b$), the optimal labour and capital demand would be determined by their corresponding marginal rate of technical substitution (MRTS) of the production factors shown in Equation 5.5 and 5.6:

$$\frac{\partial Q}{\partial L} = \frac{aAL^aK^b}{L} = aAL^{a-1}K^b \tag{5.5}$$

$$\frac{\partial Q}{\partial K} = \frac{bAL^aK^b}{K} = bAL^aK^{b-1} \tag{5.6}$$

From the production function side, the optimal shares of production factors can be obtained by matching the production function MRTS to the relative prices of the input factors (i.e. $\frac{w}{r}$). This match between factor MRTS and relative prices of the input factors is the keystone of the duality theory

that bridges the firm production function and the firm cost function.

$$\frac{w}{r} = \frac{aAL^{a-1}K^b}{bAL^aK^{b-1}} = \frac{a}{b}\frac{K}{L} \Rightarrow bwL = arK \tag{5.7}$$

Based on the initial Cobb-Douglas production function, $Q = AL^aK^b$, K can be defined in terms of labour (L), output (Q) and input prices (w and r) as shown in Equation 5.8.

$$Q = AL^aK^b \Rightarrow K = \left(\frac{Q}{AL^a}\right)^{\frac{1}{b}} \tag{5.8}$$

Likewise, this can be combined with Equation 5.7 to solve for L in terms of output and input prices (Equation 5.10):

$$L = \frac{arK}{bw} = \frac{ar}{bw}\left(\frac{Q}{AL^a}\right)^{\frac{1}{b}} \Rightarrow LL^{\frac{a}{b}} = \frac{ar}{bwA^{\frac{1}{b}}}Q^{\frac{1}{b}} \tag{5.9}$$

$$L = \left(\frac{ar}{bwA^{\frac{1}{b}}}\right)^{\frac{b}{a+b}} Q^{\frac{1}{a+b}} \tag{5.10}$$

Similarly, the input factor K can also be defined as a function of the input prices, w and r, and the output Q:

$$K = \left(\frac{bw}{arA^{\frac{1}{a}}}\right)^{\frac{a}{a+b}} Q^{\frac{1}{a+b}} = \left(\frac{arA^{\frac{1}{a}}}{bw}\right)^{\frac{-a}{a+b}} Q^{\frac{1}{a+b}} \tag{5.11}$$

Once the efficient demand level of the production factors has been defined in Equation 5.10 and 5.11 for labour and capital respectively, these two equations can be incorporated into the cost function equation $C(Q, w, r) =$

$wL(Q,w,r) + rK(Q,w,r)$ as shown in Equation 5.12:

$$C(Q,w,r) = w\left(\frac{ar}{bwA^{\frac{1}{b}}}\right)^{\frac{b}{a+b}} Q^{\frac{1}{a+b}} + r\left(\frac{arA^{\frac{1}{a}}}{bw}\right)^{\frac{-a}{a+b}} Q^{\frac{1}{a+b}} \tag{5.12}$$

Equation 5.12 can be simplified by means of summarising the common elements into a constant term as shown in Equations 5.13 and 5.14.

$$C = \theta w^{\frac{b}{a+b}} r^{\frac{b}{a+b}} Q^{\frac{1}{a+b}} \tag{5.13}$$

$$\text{where } \theta \equiv \frac{1}{A^{\frac{1}{a+b}}}\left[\left(\frac{a}{b}\right)^{\frac{b}{a+b}} + \left(\frac{a}{b}\right)^{\frac{-a}{a+b}}\right] \tag{5.14}$$

From a cost function to a production function

When studying the efficiency of a firm, it is usually easier to estimate the cost function than the production function. Thus, the production function can be derived from the cost function by means of the Shephard's lemma. Consequently, given a cost function $C = \theta w^a r^{1-a} Q^b$, where the exponentials of the input prices add up to one, we can find the demand for the input factors by means of the partial derivatives of the cost with respect to the input factor prices (i.e. w and r).

$$L(w,r,Q) = \frac{\partial C}{\partial w} = \theta a w^{a-1} r^{1-a} Q^b = \theta a Q^b \left(\frac{r}{w}\right)^{1-a} \tag{5.15}$$

$$K(w,r,Q) = \frac{\partial C}{\partial r} = \theta(1-a)w^a r^{-a} Q^b = \theta(1-a)Q^b \left(\frac{r}{w}\right)^{-a} \tag{5.16}$$

Equations 5.15 and 5.16 can be solved for $\frac{r}{w}$. This equivalence based on relative input prices is the key to obtaining the input quantities, which bridges the transition from the cost function to the production function of the efficient firm:

$$\frac{r}{w} \equiv \underbrace{\left(\frac{L}{\theta a Q^b}\right)^{\frac{1}{1-a}}}_{\text{Labour}} \equiv \underbrace{\left(\frac{K}{\theta(1-a)Q^b}\right)^{\frac{-1}{a}}}_{\text{Capital}} \tag{5.17}$$

Lastly, if the last two expressions corresponding to the labour and capital quantities are raised to the power of $-a(1-a)$, then we can obtain the final production function in Equation 5.18:

$$\theta^a a^a Q^{ab} L^{-a} = \theta^{a-1}(1-a)^{a-1} Q^{b(a-1)} K^{1-a} \Rightarrow Q = A L^{\frac{a}{b}} K^{\frac{(1-a)}{b}} \tag{5.18}$$

$$\text{where } A = \left[\theta a^a (1-a)^{1-a}\right]^{\frac{-1}{b}} \tag{5.19}$$

This final Equation 5.18 represents a typical Cobb-Douglas production function, which has been simplified in Equation 5.19 by means of the total factor productivity constant A.

5.6 Production Function Estimation

A production function is an equation that relates the maximum possible output for a set of certain given inputs. The most influential functional form has been the Cobb-Douglas (Cobb and Douglas, 1928) shown in Equation 5.20, where the production (Y) is a function of the total factor productivity (A), the production inputs such as labour (L) and capital (K), and the corresponding output elasticities (a and b).

$$Y = AL^a K^b \tag{5.20}$$

The implementation of production functions can have several conceptual problems such as the pertinence of a particular functional form or the accountability of the different input quality levels between different firms. Besides this, a key econometric drawback is the simultaneity (Marschak and Andrews, 1944) between the input variables and the output, as firms would employ inputs depending on the expected output (for a detailed review see Griliches and Mairesse, 1995). For instance, Equation 5.21 shows a logged Cobb-Douglas production function with output, capital and labour (y_{it}, k_{it} and l_{it} respectively). Additionally, the function also includes unobservable productivity shocks (ϵ_{it}, e.g. unexpected weather related shocks) and predictable or observed productivity shocks when the firm makes input choices (ω_{it}, e.g. management skills or expected maintenance costs). The endogeneity arises from the fact that the quantity of capital and labour hired by the firm is usually correlated with the predictable productivity shock (ω_{it}). This

would lead to inconsistency in the OLS estimates of the output elasticities (β) in Equation 5.21.

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \epsilon_{it} \tag{5.21}$$

Initial attempts to deal with endogeneity were based on fixed effects and instrumental variables. However, fixed effects implicitly assume that the forecast productivity shock is constant over time ($\omega_{it} = \omega_{it-1}$ for all t periods), which may be unlikely in many sectors. On the other hand, instrumental variables correlated with the inputs but not with the predicted productivity shock can be found (i.e. input prices), but they may pick up other effects (i.e. input quality) which bias the estimation. Consequently, nowadays production function estimation is based on newer techniques, either based on a control function approach (Levinsohn and Petrin, 2003, Olley and Pakes, 1992) or the dynamic generalised method of moments (GMM) with panel data (Arellano and Bover, 1995, Blundell and Bond, 2000).

The approach by Olley and Pakes (1992) and Levinsohn and Petrin (2003) uses previous input decisions, intermediate inputs or investment to proxy for unobserved productivity shocks. Essentially, this approach estimates the output elasticity of labour (β_l) based on two assumptions. Firstly, the assumption that capital input is decided in the previous period ($t-1$) and secondly, a strictly monotonic function of investment which is determined by the existing capital and the observed productivity shock (i.e. ω_{it}). However, it has been shown (Ackerberg et al., 2006) that in this approach the labour input demand (i.e. l_{it}) is merely a function of the observed productivity

shock and the capital demand (ω_{it} and k_{it}, respectively), which would lead to a similar collinearity problem that the approach tries to avoid.

The second alternative is based on dynamic panel estimation, which deals with the unobserved individual effects by means of taking first differences, and with endogeneity in the explanatory variables by using a Generalised Method of Moments (GMM) approach. GMM estimators specify a set of population moment conditions, with each estimator varying depending on the set of population moments used (Hayashi, 2000).

Equation 5.21 can be simplified as Equation 5.22, where x_{it} includes the explanatory variables and, if necessary, also lags of the dependent variable (i.e. $y_{it-1}, y_{it-2}, \dots$). Given that the fixed effect (ω_i) can be correlated with both the explanatory variables (x_{it}) and the error term ϵ_{it}, first differences would eliminate the fixed effect, as shown in Equation 5.23.

$$y_{it} = x_{it}^{'}\beta + \omega_i + \epsilon_{it} \tag{5.22}$$

$$\Delta y_{it} = \Delta x_{it}^{'}\beta + \Delta\epsilon_{it} \tag{5.23}$$

Here, the OLS estimator would be inconsistent if the differenced explanatory variables ($\Delta x_{it}^{'}$) are correlated with the differenced residual ($\Delta\epsilon_{it}$). Therefore, the true parameters can be estimated with GMM when the model specification includes exogenous instrumental variables, statistically independent of the error term, but correlated with the endogenous variables. For instance, Arellano and Bond (1991) add an autoregressive distributed lag (ADL) forming a dynamic panel model shown in Equation 5.24. The x_{it}

represents current and lagged explanatory variables such as capital (k), and labour (l), ω_i indicates the individual effects and ϵ_{it} is an $IID(0, \sigma_v^2)$ error term.

$$y_{it} = ay_{i,t-1} + \beta x_{it} + \omega_i + \epsilon_{it} \tag{5.24}$$

However, the estimation of the model in Equation 5.24 would show a correlation between y_{it} and ω_i which cannot be undone by fixed effects, as the transformed errors ($\epsilon_{it} - \overline{\epsilon_i}$) would remain correlated to the transformed regressors ($y_{it} - \overline{y_i}$). The estimator could remain consistent if the panel time series were long, although in the case of the CoMET-Nova data set, it is an unbalanced panel with short time series dimension.

In this case, the GMM approach, based on the assumption that the error term is $IID(0, \sigma_v^2)$ (Arellano and Bond, 1991), can be solved for the specific set of population moment conditions. In the GMM approach the exogenous instrumental variables substitute those regressors correlated with the error term (i.e. $y_{i,t-1}$) by means of exploiting the time series nature of the panel data with lagged variables of those regressors. Thus, an instrument for $y_{i,t-1}$ could be all the lags of the variable available (i.e. $y_{i,t-2}$, $y_{i,t-3}$ and successive). The same applies to the other regressor variables (x_{it}) in Equation 5.24 such as capital and labour. The longer the time series in the panel, the more

instruments are available.

$$z_i = \begin{bmatrix} x_{i1} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & x_{i1} & x_{i2} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \dots \\ 0 & 0 & 0 & \dots & x_{i1} & \dots & x_{i,T-l} \end{bmatrix} \tag{5.25}$$

The set of all instruments (Z_i) as shown in Equation 5.25 for variable z_i and the moment conditions can be summarised in Equation 5.26:

$$Z_i = (y_i, x_i) \text{ where } E[z_i' \Delta\epsilon_i] = 0 \tag{5.26}$$

If the order of the moment conditions (q) is larger than the parameters that need to be estimated, the the GMM estimator of β can be obtained by minimising the following quadratic (J) of the sum over the sample moment conditions ($E[z_i' \Delta\epsilon_i)$ in Equation 5.27. This is simplified as $\hat{\beta}_{GMM}$ in Equation 5.29:

$$J_N = \left(n^{-1} \sum_{i=1}^{n} \Delta\epsilon_i' Z_i \right) W_N^{-1} \left(n^{-1} \sum_{i=1}^{n} Z_i' \Delta\epsilon_i \right) \tag{5.27}$$

$$\text{where} \quad W_N = n^1 \sum_{i=1}^{n} (Z_i' \Delta\epsilon_i \Delta\epsilon_i' Z_i) \tag{5.28}$$

$$J(\hat{\beta}_{GMM}) = \overline{g}(\hat{\beta}_{GMM})' W_N^{-1} \overline{g}(\hat{\beta}_{GMM}) \tag{5.29}$$

There are two main panel data GMM estimators. Firstly, the difference GMM (Arellano and Bond, 1991) which uses moment conditions from the

estimated first differences of the error term. However, this approach is not suitable for persistent panel data, as the correlation between the instrument and the endogenous variables is very weak (Graham et al., 2009). Unfortunately, this is the particular case of the urban rail system data set used for this research.

However, there is a second GMM approach; the system-GMM devised by Blundell and Bond (1998), which fits the estimation of a production function with persistent panel data; a detailed illustration can be found in Blundell and Bond (2000). The system-GMM combines the differenced Equation 5.23 with the initial levels equation ($y_{it} = x_{it}^{'}\beta + \omega_i + \epsilon_{it}$) which uses lagged differences of the explanatory variables x_{it} as instruments. The vector of moment conditions, as in Equation 5.26, is shown in Equation 5.30. The set of instruments ($z_i^{'+}$) is shown in Equation 5.31

$$E[z_i^{'+}u]\Delta\epsilon_i] = 0 \text{ where } u_i = \begin{bmatrix} \Delta\epsilon_i \\ \\ \epsilon_i \end{bmatrix} \tag{5.30}$$

$$z_i^{'+} = \begin{bmatrix} z_i & 0 & 0 & \dots & 0 \\ 0 & \Delta x_{i1} & 0 & \dots & 0 \\ 0 & 0 & \Delta x_{i2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \Delta x_{i,T-l} \end{bmatrix} \tag{5.31}$$

There are two critical assumptions for the consistency of the system-

GMM. Firstly, the absence of first order serial correlation in the error of Equation 5.24 and, secondly, the exogeneity of the instrument matrix shown in Equation 5.31. The Arellano Bond test checks for the existence of first and second order serial correlation in the error terms of Equation 5.24 whereas the Sargan-Hansen test confirms the exogeneity of the instruments.

Therefore, once all the assumptions are met, the system-GMM approach is able to deal with endogenous regressors, individual heterogeneity, measurement error and persistent panel data. Furthermore, the system-GMM is compatible with a small sample (Soto et al., 2009) data set where the number of cross sectional units is limited.

5.7 Production Function Results

This section presents the production functions estimated for the three inputs considered in the cost functions of the previous section; car kilometres shown in Table 5.4, passenger journeys in Table 5.5 and train hours in Table 5.6. Each table presents a Cobb-Douglas and a translog production function estimated by system-GMM as described in Section 5.6. The Cobb-Douglas production function has also been estimated with pooled-OLS, within groups and difference-GMM as a reference for comparison. These reference models for car kilometres (Table 11.2), passenger journeys (Table 11.3) and train hours (Table 11.4) are available in the Annex (pages 286 to 288). The inputs considered are the total labour hours (L), total electricity (E), network length of the metro (NL), number of stations in the network (ST) and fleet size measured as the total number of cars (FL).

Table 5.4: Car kilometre Sys-GMM Production Function Results

	Dependent variable:	
	car kilometres	
	(1)	(2)
$\ln CKM_{t-1}$	0.518^{***}	0.361^{**}
	(0.101)	(0.141)
$\ln L$	0.195^{***}	1.190^{***}
	(0.048)	(0.327)
$\ln E$	-0.036	-0.686^{***}
	(0.103)	(0.229)
$\ln L^2$		0.383^{***}
		(0.131)
$\ln E^2$		0.286^{***}
		(0.092)
$\ln LE$		-0.689^{***}
		(0.217)
$\ln NL$	0.246^{***}	0.259^{***}
	(0.089)	(0.087)
$\ln ST$	-0.204	-0.175^{*}
	(0.126)	(0.103)
$\ln FL$	0.259^{*}	0.361^{***}
	(0.150)	(0.117)
AR(1)	-2.18 (0.029)	-1.54 (0.124)
AR(2)	-1.37 (0.17)	-0.87 (0.385)
Sargan	20.66 (0.898)	19.55 (0.969)
Observations	27	27
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5.5: Passenger journey Sys-GMM Production Function Results

	Dependent variable:	
	Passenger Journeys	
	(1)	(2)
$\ln PJN_{t-1}$	0.966^{***}	0.774^{***}
	(0.078)	(0.107)
$\ln L$	0.021	0.389^{*}
	(0.024)	(0.201)
$\ln E$	0.105	0.145
	(0.065)	(0.111)
$\ln L^2$		0.071
		(0.091)
$\ln E^2$		0.017
		(0.039)
$\ln LE$		−0.142
		(0.125)
$\ln NL$	−0.017	−0.058
	(0.069)	(0.103)
$\ln ST$	0.052	−0.013
	(0.072)	(0.093)
$\ln FL$	−0.086	0.069
	(0.076)	(0.100)
AR(1)	−2.23 (0.026)	−2.16 (0.031)
AR(2)	−1.72 (0.085)	−1.64 (0.102)
Sargan	21.17 (0.882)	22.25 (0.922)
Observations	27	27
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5.6: Train hours Sys-GMM Production Function Results

	Dependent variable:	
	Train hour	
	(1)	(2)
$\ln THR_{t-1}$	0.929***	0.428**
	(0.026)	(0.170)
$\ln L$	0.087**	0.730
	(0.037)	(0.864)
$\ln E$	−0.067	−1.209**
	(0.051)	(0.575)
$\ln L^2$		0.057
		(0.177)
$\ln E^2$		0.265
		(0.163)
$\ln LE$		−0.238
		(0.379)
$\ln NL$	0.058	0.017
	(0.065)	(0.167)
$\ln ST$	−0.019	0.406
	(0.069)	(0.249)
$\ln FL$	−0.0002	
	(0.058)	
AR(1)	−2.42 (0.015)	−2.05 (0.04)
AR(2)	0.55 (0.581)	0.4 (0.689)
Sargan	18.64 (0.947)	14.07 (0.997)
Observations	27	27
Note:		*p<0.1; **p<0.05; ***p<0.01

With regard to the diagnostic tests, the System-GMM models (Tables 5.4, 5.5 and 5.6) pass the Sargan and Hansen tests of over-identifying restrictions whereas some of the difference-GMM models are close to fail (e.g. Table 11.2). The translog models perform better than the Cobb-Douglas models at the Arellano Bond test for serial autocorrelation (AR), which is applied to the differences of the residuals. In many cases, the Cobb-Douglas specification only passes the second order test (AR2). This can also be due to the fact that the panel data set used has few cross units (i.e. metros) and is borderline with the minimum recommendation of at least 20 different cross-units (Roodman, 2009).

The models for car kilometres show the largest number of statistically significant variables and also coefficient estimates in line with the expected values based on the CoMET-Nova expertise. The lag of car kilometres is positive although smaller than the estimates for the pooled OLS in Table 11.2. This is expected, as strong metro-specific effects would tend to affect the coefficient estimates for the lagged variable, leading to upward-bias in the OLS estimates and downward-bias for the within groups estimates (Blundell and Bond, 2000). Labour presents positive coefficient estimates, although the values for the Cobb-Douglas model seem to be more accurate. With regard to the asset inputs, larger networks tend to produce more car kilometres whereas networks with more stations tend to produce fewer; this is reasonable, as more stations imply more stops and consequently lower speeds, leading to fewer car kilometres. Lastly, the fleet size increases the car kilometres as longer trains with more cars increase car kilometre production. Overall, these coefficient estimates are roughly validated by the models shown in the annex (Table 11.2

on page 286) and the cost functions estimated in this chapter. Likewise, they have also been considered sensible by industry practitioners.

The passenger journey production function estimations show fewer statistically significant variables and a poorer performance in the diagnostic tests. This is applicable to both system-GMM models but also to the alternative models included in the Annex (Table 11.3 on page 287). In this case, only labour and the lag of passenger journeys are significant. The poor fit of the production function for passenger journeys is probably due to the nature of the output considered, which is a final rather than an intermediate output such as car kilometres. Intermediate outputs resemble manufacturing production where labour and capital produce tangible outputs. The profitability of production is determined by the market, but there can be production of intermediate output, such as car kilometres without demand. On the contrary, final inputs, such as metro passenger journeys, follow a similar pattern to the service sector where the output is only produced as long as there is a demand for it. Consequently, the production function approach is better suited to model intermediate outputs rather than final output such as passenger journeys.

Lastly, the train hours production function models show few significant coefficient estimates, as in the case of the passenger journeys models. The same applies for the models estimated with pooled OLS, within groups and difference-GMM in the Annex (Table 11.4 on page 288). This is surprising, as train service hours can be considered as an intermediate output, similar to car kilometres, which would be more suitably modelled by means of production functions. However, train hours also present less significant estimates in the

cost function. Thus, it is likely that the measurement of train service hours for the CoMET-Nova metros is not as accurate as for car kilometres and the same factors affecting the cost function estimation are also affecting the production function estimation.

5.7.1 Cost and Production Function Duality Results

This subsection evaluates whether the metros comply with the theoretical framework of duality described in section 5.5, based on the cost and production functions estimated in Sections 5.4 and 5.7 respectively. According to the duality theory, it would be possible to transform a production function into a cost function while keeping all its properties (e.g. returns to density) and viceversa. The keystone of this transformation is the assumption that metros use inputs efficiently by matching the marginal productivity of the inputs with the input relative prices as explained in Equation 5.7. This would entail that metros optimise cost efficiency following market input prices.

The Cobb-Douglas car kilometres production function estimated in Table 5.4 is shown in Equation 5.32. This can be simplified to Equation 5.33 where we focus on the two inputs for which input prices are included in the operating cost functions estimated in Table 5.3, which are labour wages (w_L) and energy prices (w_E). The other inputs in the production function (i.e. CKM, $NL^{0.246}$, $ST^{-0.204}$ and $FL^{0.259}$) can be considered constant and simplified into a Phi coefficient (Φ) in Equation 5.33.

$$CKM = CKM_{t-1}^{0.518}\ L^{0.195}\ E^{-0.036}\ NL^{0.246}\ ST^{-0.204}\ FL^{0.259} \tag{5.32}$$

$$CKM = \Phi L^{0.195}\ E^{-0.036} \tag{5.33}$$

Following the duality theory and derivation shown in section 5.5, the Cobb-Douglas production function estimated in Table 5.4 and shown in Equation 5.33 would correspond to the operating cost function of Equation 5.34. The κ coefficient is based on the θ coefficient described in Equation 5.13. The other production inputs (i.e. Φ), which are considered as constant, would be included in the A coefficient of the θ description.

$$Op.Cost = \kappa w_L^{1.226} w_E^{-0.226} CKM^{6.289} \tag{5.34}$$

Given that these are the total cost elasticity estimates, the average cost elasticity estimates would only change in the cost elasticity with respect to output, which would be one unit lower as shown in the Annex (section 11.1 on page 283). The unit cost elasticity estimates for wages (w_L) shown in Equation 5.35 are similar to the estimate shown in Table 5.3; 0.226 derived from the production function versus the 0.311 estimated in the cost function. However, the production function energy price elasticity (i.e. -1.226) is substantially different to the one estimated in the original cost function (0.162 in Table 5.3) although the estimate for energy in the Cobb-Douglas production function is statistically insignificant. With regard to the output, the car kilometres derived cost elasticity (5.289) is also substantially differ-

ent to the cost function estimate (-0.906) shown in Table 5.3. Overall, the results may hint that, in the long run, duality may apply to metros in terms of labour input. However, it is uncertain whether this also applies to energy input, given that it was not statistically significant in the first estimation. In any case, there seems to be no evidence of duality in the production of car kilometres for the CoMET-Nova panel of metros.

$$\text{Op. Cost per car km} = \kappa w_L^{0.226} w_E^{-1.226} CKM^{5.289} \tag{5.35}$$

In the case of the operating costs per passenger journey, the derived cost function from the production function estimated in Table 5.4 is shown in Equation 5.36. Even though wage cost elasticity is relatively close (0.167 in the derived elasticity and 0.276 in the original cost function (Table 5.3)), the differences for energy prices and output (i.e. passenger journeys) are substantial. Overall, passenger journeys is an intermediate output and is less suitable for modelling by means of production function. Likewise, the production function had no significant coefficients beyond the lagged output and it is expected that the duality theory does not hold for the production of passenger journeys as explained above.

$$\text{Op. Cost per pax. journey} = \kappa w_L^{0.167} w_E^{0.833} PJN^{6.937} \tag{5.36}$$

Lastly, the train hours derived cost elasticities present unreasonable values, with labour cost elasticities above 4 and electricity cost elasticities below -3. As with passenger journeys, the production function variables were insignificant for almost all variables. On the other hand, the models for operat-

ing costs per passenger journey and operating costs per train hour shown in Table 5.3 showed sensible estimates. This may indicate that rail operations are more suitable to be modelled by cost functions as mentioned by Nash and Smith (2007).

Overall, urban rail systems do not seem to comply with the duality theory framework. This is probably due to the fact that metros' output production cannot be easily modelled and may be influenced by many other factors not related to efficiency. Metros may operate higher frequencies than needed to optimise their financial performance due to externalities (e.g. compete with private and more polluting modes of transport). Likewise, most of the metros are publicly owned and they may have other objectives beyond the optimisation of costs in the provision of service, such as regenerating neighbourhoods and providing mobility to deprived areas.

5.8 Conclusion

Significant and substantial economies of density are found in the operating costs, particularly in the production of car kilometres and passenger journeys. The operating cost models find positive and significant wage cost elasticities with evidence that higher wages increase productivity, but not enough to compensate the rise in costs. Consequently, metros with higher wages tend to have higher operating costs which they compensate with lower staff hours. Likewise, the models also show that labour costs can also be contained by means of outsourcing.

The effect of scale on costs is mixed, with larger networks leading to higher

costs per passenger journeys as more suburban networks operate in less populated areas. However, more stations decrease costs per passenger journey, as they may increase passenger demand by means of increasing accessibility. Overall, the results do not contradict previous research on economies of scale and density but they do offer a more detailed breakdown of the effects.

The age of the network seems to play a significant role in operating cost efficiency. However, the relation between age and cost efficiency seems to be non-linear and quadratic, as old metros (inaugurated before 1940) and new metros (inaugurated after 1980) show higher costs than mature metros. Additionally, poor service reliability tends to increase operating costs, which add up to the subtle background time trend of increasing operating costs for the metros in the panel. Lastly, subsidies are associated with substantially higher operating costs and show that short-term solutions may lead to long-term financial unsustainability. Nevertheless, this study finds an association between subsidies and lower cost efficiency but it cannot claim a causal relation between the two as the data was insufficient for a causal analysis.

The production function analysis showed a less clear picture and only the car kilometres model is relatively suitable for comparison with labour, larger networks and numerous fleets increasing the production of car km. Urban rail production of passenger journeys is less appropriate as this final output also depends on the passenger demand, whereas train hours did not present statistically relevant models. Finally, the assessment of duality by means of deriving the cost functions from the production functions showed that metros usually do not behave like competitive firms adjusting their inputs to the marginal productivity of the factors; only labour suggests to be partially

optimised. This may be due to the fact that most metro operators also consider the positive externalities of the metro network and other considerations which may distance them from the purely profit-maximising behaviour of a private firm.

6 — Train Service Costs in Urban Rail Systems

Note: This chapter is an extended version of the journal article *Determinants of Train Service Costs in Metro Operations* published in the *Transportation Research Record: Journal of the Transportation Research Board: Public Transportation. Passenger Rail and Terminals* - Volume 2 No.2534 in November 2015.

6.1 Introduction

Although urban rail operations are the result of different metro divisions, the most identifiable part of the railway system for the passenger is the train service, i.e. the actual movement of the trains throughout the operator network. Train service entails labour costs for the on-board and control centre staff and, additionally, energy costs for traction and on-board train facilities.

Urban rail systems are characterised by offering a high-frequency service and an intensive use of the rail network. Therefore, the increase of cost effi-

ciency in train service costs is particularly important for these rail operators in comparison to other low-frequency rail operators on regional and long distance passenger rail services. However, train service cost efficiency can be different depending on the actual output that is being considered. In this chapter, the train service cost efficiency is analysed for the production of car kilometres, passenger journeys and train hours. For urban rail systems, passenger journeys is the final output and both car kilometres and train hours are intermediate outputs.

The aim of the econometric analysis presented in this chapter is to provide a useful tool for rail transport planners and rail practitioners. The analysis can explain and forecast the train service costs while considering the particular circumstances of each metro. This can be particularly useful for transport planners as the econometric models explain how a change in service characteristics would affect the train service operating costs of the metro. The models can also predict how the train service costs of an average metro would change if there were a change in operations such as the introduction of air-conditioned rolling stock, the contracting out of staff or the ageing of the system. Likewise, given that many urban rail systems operate under monopolistic conditions, the models can be used by the transport authorities as a guideline to compare the cost efficiency of the metro and provide an overview of the financial performance of the rail operator.

6.2 Literature Review

The particular constituents of rail operating costs, such as the costs of circulating the trains along the network (i.e. train service costs), have not received much attention from academia, probably due to the lack of financial disaggregated data. Furthermore, within the literature on rail cost efficiency, urban rail systems have been less popular than other modes of rail, as explained in the general literature review in Chapter 2.

Nevertheless, deregularisation of the rail sector spurred a new wave of academic research on the cost structure and the variable costs of the rail operators (e.g. Brown et al., 1979, Caves et al., 1981*a*, Filippini and Maggi, 1992). The enactment of the Railroad Revitalisation and Regulatory Reform Act (1976) and the Staggers Rail Act (1980) enabled the US operators to adapt their pricing and frequency of service, and consequently, the optimisation of the train service and its costs became a source of potential efficiency gains. In particular, the economies of density and scale became a key concern (e.g. Caves et al., 1985, Kim, 1987) in relation to cost optimisation, as lower service frequencies may reduce the operating costs less than proportionally in the existence of strong economies of service density. Furthermore, the research on economies of density also considered the multi-output characteristics of the rail industry, so the economies of density and scale vary depending on whether the output considered is car kilometres or passenger journeys (e.g. Braeutigam et al., 1982).

Most of the research found notable economies of density on rail variable

costs for both large networks, such as the US (e.g. Caves et al., 1981*b*, Keeler, 1974) or Italy (e.g. Mancuso and Reverberi, 2003). However, this was also found in smaller networks like Ireland (McGeehan, 1993), the Swiss private railways (Filippini and Maggi, 1992), the Netherlands (Pels et al., 2004) and Belgium (De Borger, 1992*b*). On the contrary, most of the research found null economies of scale despite some exceptions (Friedlaender et al., 1993).

These conclusions on economies of scale and density were corroborated by academic research regardless of the econometric methodology considered in each case. Cantos and Maudos (2000) studied the cost efficiency of 15 European operators producing train kilometres in passenger service by means of an stochastic frontier approach. The stochastic frontier analysis enabled Cantos and Maudos (2000) to explain the efficiency levels by means of other variables such as the share of the network with electrified track or the managerial autonomy. Cantos (2001) also reviewed the cost efficiency of 12 European rail operators using a two-way fixed effects approach, finding low-cost elasticities with respect to passenger journeys delivered by the operators. These results on rail cost efficiency were also corroborated by later research using a data envelopment analysis (Cantos et al., 2002).

The academic proof of substantial economies of density and null economies of scale in rail service has major policy implications for the sector. According to this evidence, breaking down the provision of rail service into smaller operators would not entail any efficiency loss as the savings in variable costs are based on density of service and not on the scale of the operations. Consequently, small operators serving a particular corridor with a high frequency service could achieve high efficiency levels regardless of the size of the network

served by the operator or the size of the operator itself.

With regard to urban rail systems, there is evidence of strong returns to density and constant returns to scale (Graham et al., 2003*a*, Graham, 2008) although the research also included other operators of light rail and suburban rail operators. In the case of metros, more recent research (Santos et al., 2013) has found diseconomies of scale in operations either when scale is measured as network length or number of stations. However, this research has been contested by Tsai et al. (2015), who found economies of scale and density in a panel of 20 metros when the scale was measured by the number of stations and network length.

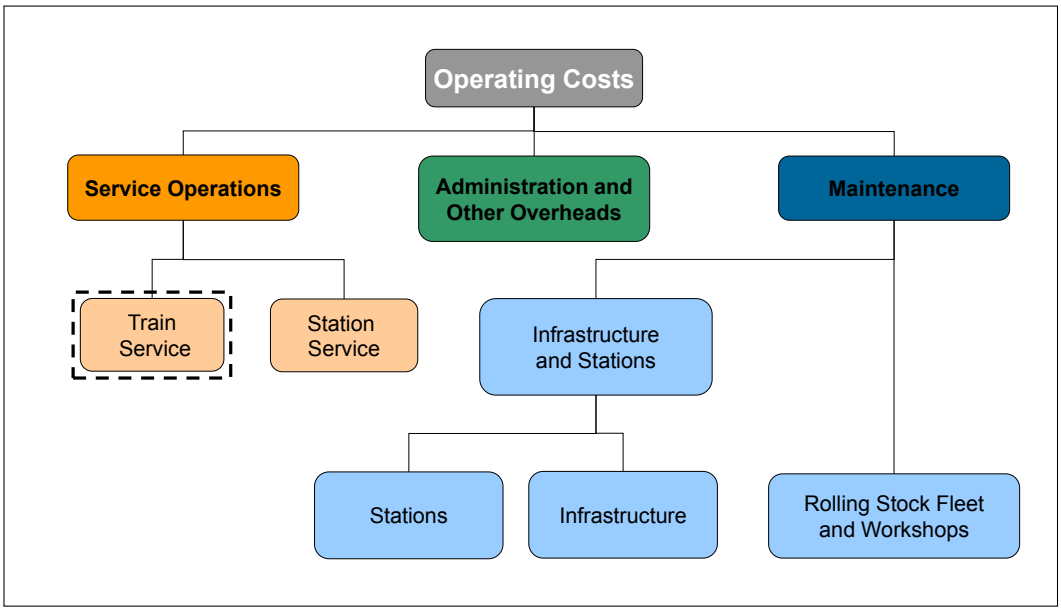
These divergences in the role of economies of scale and density for metros contrast with the general consensus on positive returns to density and constant returns to scale found in the rail efficiency academic literature. In the case of metros, this controversy could be due to the lack of more disaggregated data, so depending on which operational division is more cost-intensive for each metro (e.g. train service, station service or maintenance) the research may cast a different conclusion. Consequently, the contribution of this Chapter is to assess the role of the economies of scale and density in the provision of train services. Furthermore, the estimates are complemented by the inclusion of cost factors that have not been included before, which improve the specification of the econometric models and contribute new variables to explain the cost efficiency of the rail operators.

6.3 Data and Research Hypotheses

The data for the train service cost analysis is based on a panel of 24 metros over a period of 9 years (i.e. 2005-2014). The source of this panel is the CoMET-Nova consortia and its KPI database described in chapter 4. According to the CoMET-Nova standards, the train service costs category corresponds to one of the two subcategories of the total service costs as shown in Figure 6.1 in page 166. This train service cost category (i.e. C_{train}) is regressed on the set of explanatory variables (x_j) described in this section as explained in Equation 6.1.

$$LogC_{train} = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \alpha_i + \varepsilon_{it} \tag{6.1}$$

Figure 6.1: Train service costs and other operating costs of a typical CoMET-Nova metro operator.



In the CoMET-Nova framework, based on the KPI handbook, the train service costs analysed in this chapter include the following:

- All costs related to the drivers of the passenger service trains.
- All on-board staff or staff assigned to travel on board trains other than drivers, such as guards, conductors, customer service staff and any other staff whose normal working place is on board the train.
- All cost related to other train service operation costs such as:
  - Operations Control Centre (OCC) that supports train operations.
  - Crew management and administration (e.g. driver scheduling).
  - Service depot costs.
- All energy cost related to any of the following categories:
  - Train traction and any other train energy consumption (e.g. lighting and air conditioning).
  - Location and signalling systems.
  - Automatic train control (ATC), automatic train supervision and any other automatic train operations (ATO) system.
- All cost related to drivers of engineering trains.
- Salaries, training, overtime and employment-related fringe costs for metro staff and any contract labour cost.

Despite the comprehensiveness of the train service cost category, certain dimensions that are not critical for the operation of the trains are excluded, such as the ticketing and revenue control functions or any tasks related to maintenance of any type.

The train service costs are explained as average costs per output. The three outputs considered correspond to two intermediate outputs, car kilometres and train hours, and a final output, passenger journeys. Car kilometres comprises all the car kilometres produced which are actually operated in revenue service (i.e. trains on passenger service). Passenger journeys considers all the passengers travelling in the system, including fare evaders. Lastly, train hours correspond to the summation of all the hours in passenger service for all trains. A more detailed definition is available in the Annex Table 11.6 on page 291.

Figures 6.2, 6.3 and 6.4 present the anonymised train service cost values for the CoMET-Nova metros, whereas Table 6.1 shows the dependent variables of the econometric model alongside the descriptive statistics for each variable. Given that the financial data is not directly comparable between different countries, the local currency costs have been transformed into Parity Purchasing Power (PPP) US Dollars in order to enable international comparisons [See Section 4.3 on page 113 for further reference].

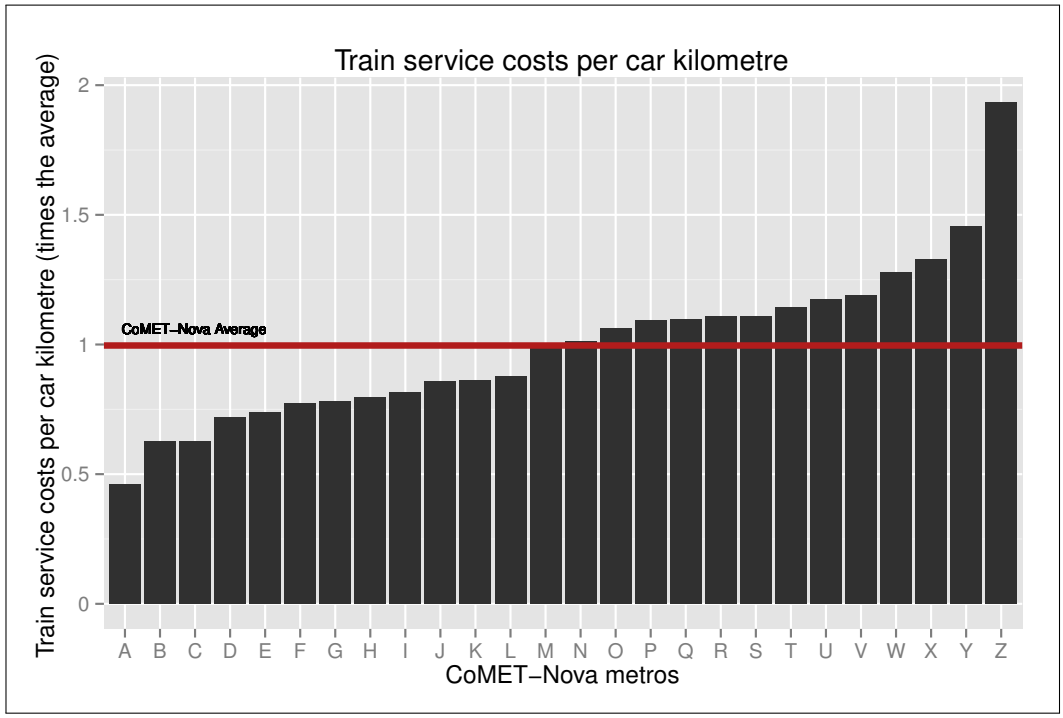
Moving a car in passenger service over the network for one kilometre costs, on average for the CoMET-Nova metros, 1.18 US$ PPP. However, there is a significant variation in these costs and, on average, the metros diverge 0.42 US$ from this 1.18 US$ average. In particular, the lowest cost is 0.67 US$ per car km, although one of the metros of the panel for a particular

Table 6.1: Train service dependent variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max
Train service costs per car km	1.18	1.17	0.42	0.67	2.49
Train service costs per passenger journey	0.25	0.24	0.12	0.1	0.62
Train service costs per train hour	214.15	207.47	75.74	112.52	457.36

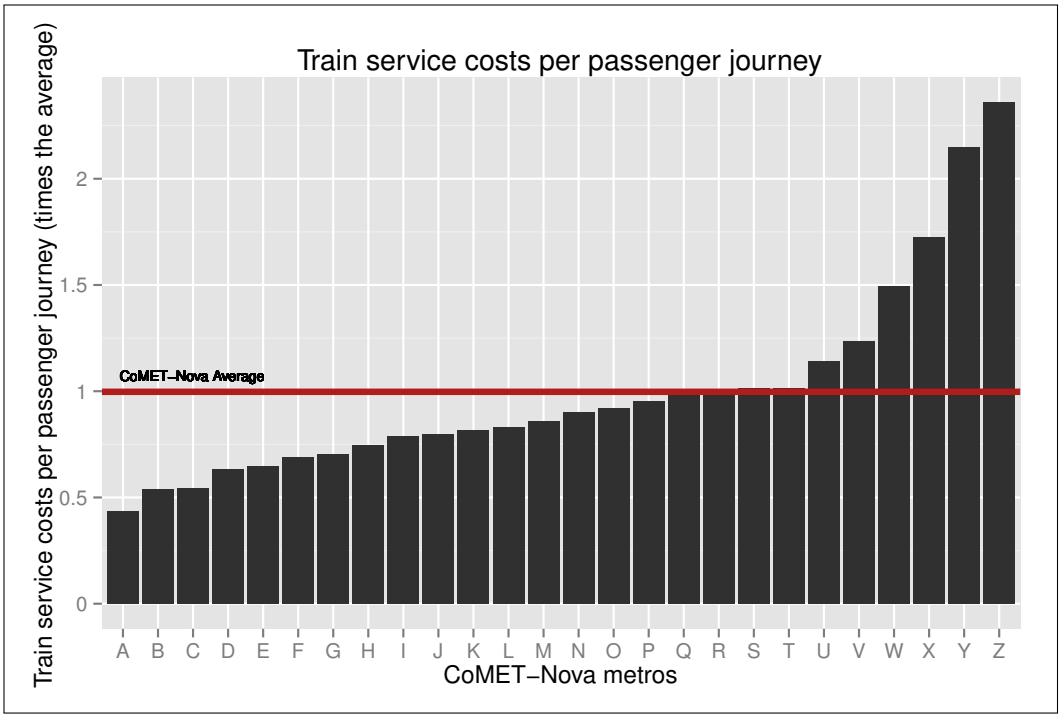
Note: Costs in Parity Purchasing Power (PPP) U.S. Dollars. Obs.: 138.

Figure 6.2: Train service costs per car kilometre.



year registered a train service cost per car km of 2.49 US$. These extremely high values are relatively uncommon due the size of the standard deviation and the median, which is slightly below the average train service costs per

Figure 6.3: Train service costs per passenger journey.



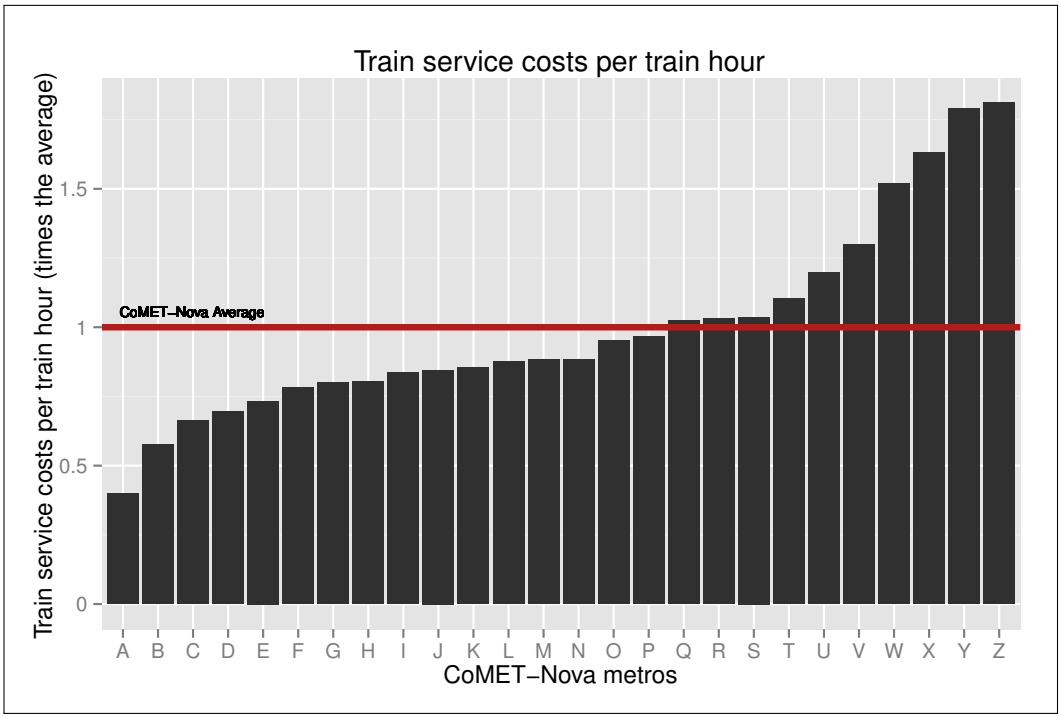
car km.

The train service costs per passenger journey follows a similar pattern to the previous costs, with similar mean (0.25 US$) and median (0.24 US$) values, ordinary standard deviation (0.12 US$) but extreme maximum values (0.62 US$). In this case, the maximum train service costs per passenger journey in the panel is roughly 2.5 times the average.

Lastly, with regard to train service costs per train hour, most of the metros have similar values to the mean, with average median values and low standard deviations. However, the extreme values are substantially different to the mean, particularly for the minimum values.

This variability in the dependent variables is explained by the econometric models shown in Section 6.4. The descriptive statistics of the model

Figure 6.4: Train service costs per train hour.



explanatory variables are shown in Table 6.2. A detailed definition is also shown in Table 11.6.

Following the research scope outlined in Section 1.5 on page 34, a main contribution of this research is to provide a detailed and disaggregated analysis of the economies of scale and density for each of the components of metro operating costs. Consequently, each of the econometric models for each of the dependent variables explained above include a regressor on either car km, passenger journeys or train hours. If the coefficient estimates for these output variables are zero, or in similar terms, statistically insignificant, there would be no economies of density in the provision of the train service. On the contrary, positive coefficients would imply that as production of car kilometres, or any of the other two outputs, increases, the train service costs per

Table 6.2: Train service explanatory variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max	Obs
Car km (m)	94.216	89.093	175.559	8.562	776.35	N = 136
Passenger journeys (m)	453.183	508.065	558.035	35.7	2463.8	N = 136
Train hours (m)	0.52	0.516	0.758	0.079	2.908	N = 136
Wages (PPP US$/h)	34.346	38.385	15.816	14.16	78.274	N = 136
Electricity price (PPP US$/MWh)	150.13	158.784	100.231	50.817	1030.105	N = 136
Staff hours - Train service (m)	1.767	1.629	3.546	0.272	13.508	N = 136
Traction electricity (MWh)	249.708	258.925	411.604	21.169	1809.5	N = 136
Outsourced train serv. staff (%)	0.012	0	0.049	0	0.328	N = 136
Driver productivity	0.48	0.489	0.124	0.268	0.8	N = 136
Network length (km)	101.406	86.2	144.825	19.74	548	N = 136
Number of stations	87.579	68	97.552	18	424	N = 136
Standardised train capacity	1129.936	990.78	510.212	339.333	2654.019	N = 136
Air conditioned RS (%)	0.551	0.546	0.462	0	1	N = 136
Network age (years)	52.779	40	38.004	2	150	N = 136
Year				2005	2014	N = 138

Note: Units in PPP US$ refer to Parity Purchasing Power (PPP) U.S. Dollars.

car km would increase. However, and particularly for the case of the train service costs per passenger journey, the expected coefficients are negative due to the fact that train service costs are relatively independent of the total passengers transported in the train.

The variability in the scale and density of operations is substantial as can be seen in Table 6.2. Most metros would run almost 100 million car kilometres, transport 500 million passenger journeys and offer roughly half a million train hours per year. However, the three density variables present standard deviations larger than the mean and the median. This contrast is expected as the panel is formed by the CoMET consortium of large metros and the medium size metros of the Nova consortium. This can also be positive for the econometric models as a larger variance in the output variables, once it is controlled by the size of operations as explained below, may contribute to the statistical significance of the variables.

Input prices and quantities have also been considered in the cost function estimation. In particular, labour wages and labour hours form part of the model in order to explain how labour-intensive the provision of the train service is across the CoMET-Nova models. The average wage in the panel of metros considered for the train service models is 38 US$ with a moderate standard deviation (15.8 US$), which also indicates the different levels of development of the CoMET-Nova cities. The wages refer to the metro average wage due to data availability and the impossibility to distinguish only train service wages from the CoMET-Nova KPI categories.

Electricity prices per MegaWatt/hour (MWh) present a lower variability than output variables but higher than for wages. The average price per

MWh is 158 US$, although the standard deviation (100 US$/MWh) shown in Table 6.2 is slightly higher than half of the average electricity price.

Input quantities for labour and traction electricity are also considered. The descriptive statistics of the variables present similar proportions to the output variables described above (i.e. car kilometres, passenger journeys and train hours). The input quantities are included in the main econometric model and the same model, without these input quantities, is also estimated in the Annex of this thesis. An analysis on the link between input productivity and input prices can be estimated by means of comparing the coefficients on the input prices (i.e. wages and electricity prices) between both models.

The train service models also include more variables related to the main input factor, labour. Firstly, the percentage of labour hours outsourced by the metro is included in order to explain the cost efficiency of the operator. Given that train service labour hours refer mostly to drivers and the control of the trains, the outsourcing levels are relatively low. The outsourcing average across the CoMET-Nova metros is just 1.2% and, most notably, the median is 0, indicating that more than half of the CoMET-Nova metros do not outsource any labour hours related to train service. The other variable on labour is driver productivity, measured as the percentage of hours spent driving over the total driver hours, for both staff and contractors, reported by the metro. This ratio takes into account the management of the spare drivers, absenteeism figures and the ability to efficiency chain different driver shifts. Consequently, the driver productivity cannot be over 1. The average driver productivity is around 0.50, so barely half of the drivers hours are actually spent driving. However, the values for a given year of a given metro

range from 0.27 to 0.80.

In order to evaluate the existence of economies of scale, two variables have been included on network length and number of stations. The average CoMET-Nova metro included in the train service models has a network of 100 km and 87 stations. However, the standard deviation is substantial, with an average deviation from these averages of 145 km and 97 stations respectively. Overall, positive coefficients for the network length and number of station variables would mean that longer networks with more stations are associated with higher train service costs per car km, passenger journey or train hour, implying negative returns to scale.

For the train service cost efficiency, the econometric models also consider variables related to the asset properties of the metro. First, the percentage of rolling stock with air conditioning as a proxy for the equipment level and complexity of the rolling stock. More technologically advanced rolling stock would probably require less train service labour, leading to lower costs. On the contrary, most of this equipment may not replace labour and may cause higher energy consumption (e.g. air conditioning) leading to higher costs. Therefore, the assumed effect of air conditioning rolling stock is ambiguous.

The standardised capacity of trains is also considered and it refers to the number of passengers that can be transported by a train, measured as the number of seats plus 4 passengers per standing square metre and all multiplied by the average number of cars per train. Presumably, larger trains would imply more passenger capacity and, thus, lower train service costs per passenger. In the panel of metros used for this chapter, the average train capacity for CoMET-Nova metros is 1129 passengers with a standard

deviation of 97 passengers. However, the largest trains in the panel of metros considered for this chapter can transport a maximum of 2654 whereas the smallest trains transport less than 400 passengers.

Lastly, the age of the system, measured as the years since the start of operations, has also been included in the models. This controls for initial design constraints that may still impose a cost for present operations. For instance, tunnel width and height, track gauge or curvature radius may impose severe constraints on train speed or capacity. In the panel of metros included in the train service models, most of the metros are above 40 years old, with an average of 50 years. However, there is a significant variation in the ages of the metros, with a standard deviation of 38 years and a maximum of 150 years for the London Underground. In the additional models presented in the Annex (page 296), the age variable has also been broken down into three main categories: new metros, for metros built after 1980, mature metros for those which started operations between 1945 and 1980, and old metros, for those which started operations before 1945.

Lastly, a time trend to control for the overall evolution of the train service costs, once all the other factors are accounted for in the previous variables, has also been added. The trend spans from the beginning of the panel time series, 2005, up to the most recent year available, 2014.

Other explanatory variables considered but not included

There are several variables that, despite the best efforts to include them and their probable role in determining train service cost efficiency, have been finally discarded. Firstly, the urban density of the area covered by the metro

was included in the models as high, medium and low density metropolitan areas. Urban density is probably a key factor in optimising the usage of the network and, particularly for the passenger journeys models, it increases the output without increasing the costs. However, the measure of the urban density can be very inaccurate as the relevant variable is not the metropolitan urban density but the density around the stations served. Furthermore, urban density is not uniform across the metropolitan area, so the same metro operator may serve different urban densities within the same metropolitan area. For instance, this is the case of the London Underground, with a high density in the core network and substantial network length in low-density suburban lines.

Another variable considered for the train service cost efficiency models was the operating train speed, as higher speeds would probably increase the efficiency of train service costs per car kilometre. In order to avoid the missing data for those metros without speed data, the average distance between stations was also considered, given that less frequent stops lead to higher speeds. However, the models already include the variables for network length and number of stations. Consequently, the average distance between stations, and thus also average commercial speed, is already indirectly considered in the model.

Lastly, among other variables, the percentage of the network with driverless operations was also considered. Driverless operations could be a key factor for train service savings, as all the driving staff hours are saved and it also increases the energy efficiency of operations. However, driverless operations are still quite rare and most of the metros do not have any driverless lines.

Furthermore, percentage of driverless network actually represents driverless lines, and these new driverless lines may not be representative of the whole metro. Therefore, a line in a high density or low density area could have outstanding train service costs per passenger journey even though it is not due to the driverless nature of the operations but to the speciality of that particular line.

6.4 Results

The econometric models shown in this section have been estimated by means of a random effects panel estimator as explained in Section 3.2.1. Table 6.3 presents the three train service cost functions estimated for each of the corresponding outputs: train service costs per car km, per passenger journey and per train hour. Each of the three estimations represents a cost function as shown in Equation 6.2, where the coefficient estimates (i.e.β_j) for the variables (i.e. x_j) represent cost elasticities.

$$LogC_{train} = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \alpha_i + \varepsilon_{it} \tag{6.2}$$

Each coefficient (β) shows the effect of that variable while holding all the other variables constant, thus, isolating the effect of the variable of interest. As for any random effect model, the coefficients represent the average response of the metros in the panel, in this case, the average effect for the CoMET-Nova metros. The estimation is done by means of random effects due to the existence of non-constant variance across entities (Table 11.17 on

page 304)

The first variables considered in Table 6.3 refer to the returns to density of the train service operations. In the case of train service cost per car kilometre and per passenger journey, the coefficients are statistically significant at a 99% confidence level. The coefficients, -0.916 and -0.913, imply that producing 10% more car kilometres or transporting 10% more passenger journeys would lower the train service cost per car kilometre or train service cost per passenger journeys. In terms of total cost, increasing the total number of car kilometres or passenger journeys would still increase the total cost as the coefficients are above -1. However, it implies that there is a very low marginal cost of transporting additional passengers and producing extra car kilometres. This is an expected result for any firm using a large share of fixed assets when operating below maximum capacity, which is the circumstance of most rail operators. Nevertheless, the spare capacity may be available at times when it is not needed (e.g. off-peak hours) so, given that the marginal cost of producing output (e.g. car km) are still positive, the metros would save costs when they operate below maximum capacity at hours with low demand (e.g. off-peak hours). The coefficients for the variables on returns to density are surprisingly large in absolute terms, so there may be a signalling or specialisation effect, where those metros with more passenger journeys or producing more car kilometres are also more efficient at it. With regard to train service hours, those metros producing more train service hours also show lower train service cost per train hour although with a more moderate, yet statistically significant, value (-0.692) than in the previous two models.

A key factor in determining cost functions is input factor prices. In this

Table 6.3: Train Service Cost Model Results - Random Effects

	Train service costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.916***		
	(0.233)		
Pax. journeys [Ln]		−0.913***	
		(0.108)	
Train hours [Ln]			−0.692***
			(0.205)
Wages [Ln]	0.597***	0.574***	0.601***
	(0.094)	(0.093)	(0.094)
Elect. price [Ln]	0.201***	0.204***	0.180**
	(0.069)	(0.069)	(0.069)
Staff hours [Ln]	0.600***	0.544***	0.489***
	(0.123)	(0.126)	(0.145)
Traction elect. [Ln]	0.342*	0.332***	0.313**
	(0.176)	(0.116)	(0.120)
Staff outsourcing [Ln]	−0.606*	−0.573*	−0.539
	(0.332)	(0.334)	(0.325)
Driver productivity [Ln]	−0.204*	−0.224*	−0.326**
	(0.118)	(0.118)	(0.142)
Train capacity [Ln]		0.043	
		(0.178)	
Network length [Ln]	0.046	0.109	0.070
	(0.211)	(0.174)	(0.184)
Number of stations [Ln]	−0.118	−0.098	−0.215
	(0.224)	(0.213)	(0.225)
RS with air con. (%)	0.003	−0.038	−0.033
	(0.115)	(0.115)	(0.116)
Network age (years)	−0.002	−0.001	−0.002
	(0.002)	(0.002)	(0.002)
Time trend	−0.007	−0.008	−0.008
	(0.005)	(0.005)	(0.005)
Constant	14.263	15.444	15.806
	(10.362)	(10.664)	(10.298)
Rho	0.7941	0.8162	0.8466
Observations	138	138	138
R^2	0.479	0.657	0.838
Note:	180	*p<0.1; **p<0.05; ***p<0.01	

case, both wages and electricity prices are consistent across the three models shown in Table 6.3. Wage coefficients range from 0.574, for the train service costs per passenger journeys model, to 0.601 in the costs per train hour model. All of them are significant at a 99% confidence level and entail that a 10% increase in wages, keeping labour hours hired and all other cost factors constant, would lead to an approximate 6% rise in train service costs. This is coherent with the economic theory as train service operations are very labour-intensive. Likewise, it also shows larger coefficient estimates than in the operating costs models.

For the electricity prices, the variables are also highly significant and show similar coefficients across the three models ranging from 0.180 to 0.204. Consequently, a 10% increase in electricity prices, holding constant the total traction energy consumed, would lead to just a 2% increase in train service costs. These coefficients represent an approximation to the share of inputs required for the train service across the CoMET-Nova metros. In the case of train service, labour is still far more important than energy, and changes in wages have a much larger effect on the final train service costs. However, this could change substantially as more metros introduce driverless lines in their networks and on-board staff is severely reduced or no longer needed.

Additionally, input quantities cost elasticities show similar results. All the values are significant at least at a 90% confidence level and the coefficients show similar proportions; only the values for the traction electricity quantities are slightly higher. Staff hours cost elasticity range from 0.489 to 0.600 whereas the cost elasticity for traction electricity ranges from 0.313 to 0.342.

The same model as in Table 6.3 is estimated without input quantities in

the Annex Table 11.14 on page 301. By means of comparing the cost elasticity coefficients for metro wages for the models with and without variable labour quantities, it can be approximated to what extent higher wages lead to higher productivity. Usually, higher wages lead to high productivity and a decrease in the total train service staff hours employed by the metro. Taking into account train service wages instead of the overall metro wages would be more accurate but, as explained in section 6.3, train service wages are not available under the CoMET-Nova KPI framework.

The keystone of the analysis is based on the wage coefficient estimates. The analysis checks if the coefficient for the model with quantities shown in Table 6.3 (i.e. β_{fixed_L}) is equal to the wage coefficient for the model without input quantities shown in Table 11.14 (i.e. $\beta_{flexible_L}$). Consequently, the null hypothesis shown in Equation 6.3 is that these two wage coefficients are similar, which means that the difference between them is not statistically significant. If the two coefficients are analogous, it would imply that when wages increase, there is no adjustment in the labour hours hired, all other factors held constant. If the coefficient for the model without fixed quantities were smaller, it would imply that metros usually offset wage increases by means of higher productivity and hiring less labour.

$$H_0 : \beta_{fixed_L} - \beta_{flexible_L} \simeq 0 \tag{6.3}$$

In order to assess if the differences between the coefficients are statistically significant, a t-test is applied following Equation 6.4. In the case of the train service cost per car km, the t-value of the difference is 0.88, so

the difference between the two coefficients β_{fixed_L} and $\beta_{flexible_L}$ (i.e. 0.086) cannot be considered to be different of zero at any conventional statistical significance level. Consequently, the null hypothesis that metros do not adjust staff hours when wages increase cannot be discarded for the train service costs per car kilometre. A similar result is shown for the train service costs per passenger journey (i.e. t-value of 0.51) and per train hour (i.e. t-value of -0.12). This would indicate that the metros do not hire less staff in train service labour when wages increase. This could be caused by the rigidities faced by the metro in order to reduce train service labour as in the case of train drivers, who usually have substantial bargaining power.

$$\theta = \frac{\beta_{fixed_L} - \beta_{flexible_L}}{(\sigma_{fixed_L} + \sigma_{flexible_L})/2} = \frac{0.597 - 0.511}{(0.094 + 0.101)/2} \simeq 0.88 \tag{6.4}$$

The models estimated in Table 6.3 also show the effect of other labour-related variables such as the percentage of outsourced train service staff and driver productivity. Table 6.3 has revealed staff outsourcing as a powerful policy to reduce train service costs. The coefficients for all the models are below -.50, implying that a 10% increase in outsourcing would lead to a 5% decrease in train service costs per car kilometre or per train hour. However, only the coefficients for the train service costs per car kilometre and per passenger journey are significant at a 90% confidence level. This is probably due to the fact that staff outsourcing in train service tasks is rare and very limited across the CoMET-Nova metros. Therefore, the lack of sufficient variance in the variable reduces its statistical significance. Adding more observations in future research, and particularly from metros outsourcing a

substantial share of their train service tasks to add data variability, would probably increase the statistical significance of the estimates.

With regard to driver productivity, the coefficient estimates for driver productivity are all negative and significant at a 90% confidence level. The coefficient estimates for the train service cost per car kilometre and per passenger journey are moderate (-0.204 and -0.224) although coherent with the research hypotheses. The coefficient estimate for the train service cost per train hour model is significant at a 95% confidence level and shows a stronger effect of driver productivity on train service costs efficiency per train hour (-0.326). Consequently, increases in driver productivity, measured as the proportion of driving time per driver hour, have a greater impact on train service costs per train hour, which is particularly coherent for the train service per hour model. The relationship between driver hours and passenger journeys or car kilometres may be more blurred than the relationship between train hours and driver hours. Consequently, this estimate (-0.302) may reflect the share of driver costs over the total cost of operating a train hour (i.e. $\approx$ 30.2%). This is also coherent with the industry expertise consulted at RTSC although the figure may vary substantially from operator to operator.

The model for train service costs per passenger journey also considered train passenger capacity as a factor to increase cost efficiency. The research hypothesis is that larger trains decrease train service costs per passenger journey. However, the relationship between train capacity and train service costs per passenger journey appears to be non-existent. The same results are found in the other models without input quantities (see Table 11.14 on page 301) and with the age groups (see Table 11.10 on page 296). This may

be due to the fact that cost per passenger journey is determined more by demand than the supply of capacity.

The models shown in Table 6.3 also considered scale variables such as network length and the number of stations served by the metro operator. For the three models the scale shows no relation with the train service costs per car kilometres, per passenger journey or per train hour. The coefficients for the two variables are barely zero and both variables are statistically insignificant for the three models. This follows the previous evidence on returns to scale where the estimates are very close to 0 (e.g. 0.003 for unit costs in Graham (2008)). Consequently, the returns to scale for the metro train service seem to be null or negligible in accordance with the literature review presented in Section 6.2.

Next, with regard to rolling stock, the percentage of rolling stock with air conditioning, which acts as a proxy for rolling stock equipment and complexity, presents insignificant estimates for the three models. Consequently, the effect of the rolling stock characteristics seems to be null, although it is most likely that the lack of more precise variables on the rolling stock characteristics, and the lack of variance over time in this characteristics, impede to obtain a more accurate analysis on how rolling stock characteristics affect train service costs.

As for all other models, a common determinant tested at all models is the network age. The coefficient estimates for this shown in Table 6.3 are not significant for any of the train service models estimated. The coefficients are almost zero and the standard errors are larger than the coefficient estimates. In Table 11.10 (Annex, page 296) the same models are estimated but the age

variable is broken down into old metros, mature and new metros where the base case is mature metros. In this case, both the old metros which initiated operations before 1945 and the new metros, those starting operations in 1988 or later, present positive estimates. New metros may show higher train service costs per car kilometre as, due to the fact that they are new, the main lines are probably central and there are fewer extensions into the suburbs where producing car kilometres is easier. Likewise, for passenger journeys, the demand for these metros may still be lower than usual for other metros as the mode share of urban transport requires several years to respond to the new capacity offered by the new metro. This results also corroborates the analysis for the overall operating costs models shown in the previous Chapter 5.

Lastly, the time trend shows a noticeable negative estimate, despite not being statistically significant. This negative estimate could be due to general technological progress in the metros that increased cost efficiency over the last years, leading to cost savings. In this case it should also be observable across other metros worldwide that are not members of the CoMET-Nova consortia. However, these cost savings could also be due to the share of good practices and benchmarking work of the CoMET-Nova metros. In the latter case, they would not be observable in other metros outside the CoMET-Nova consortia.

With regard to the assessment of the econometric models as a whole, the *rho*-coefficients shown in Table 6.3 represent the share of the variance due to the differences between metro operators out of the total variance. The total variance is composed by the sum of the variance within the metros

(i.e. variance in the variables over time) and the variance in the differences between the different metros. Consequently, for the train service models, around 80% of the model variance is due to differences between metros. This is consequential on the interpretation of the model. In this case, the model coefficients explain mainly why different metros have different train service cost levels rather than the evolution of the train service cost over time for the particular metros of the CoMET-Nova consortia.

The goodness of fit of the model is reported by the R^2 as explained in section 3.2.1. Despite the moderate sample size, 138 observations, the models are able to explain at least half of the variation for the model on train service cost per car kilometre. For the train service costs per passenger journey the variation explained is 65%, whereas for the model on cost per train hour the variation explained is up to 83%. Overall, the models explain a significant share of the variation, particularly for the train service costs per train hour model. This can be due to the fact that train hours is the most exogenous output of the three, as it does not depend on the characteristics of the network (i.e. suburban networks may be more cost-efficient on train service costs per car kilometre) and it is also independent from passenger demand and urbanisation levels, which may help to boost the efficiency of the train service costs per passenger journey.

Overall, the models left around half of the variation unexplained, probably due to exogenous variables to the metro operator that influence its performance (e.g. urban densities and train service costs per passenger journey) and other metro characteristics that could not be tested due to the lack of data (e.g. rolling stock detailed characteristics).

6.5 Conclusion

This chapter presents several econometric models identifying the main drivers of train service costs. The panel data comprises 24 CoMET-Nova metros over 9 years (2005-2014) with more than a dozen explanatory variables.

The models assess the existence of economies of density in relation to car kilometres, passenger journeys and train hours. The models find notable economies of density in the provision of train service, complementing the previous evidence in the academic literature and adding new evidence on a more disaggregated dimension. With regard to returns to scale, scale is measured in the analysis as network length of the urban rail system and number of stations in the network. The results are in line with previous research as no returns to scale have been found in the train service dimension of the operating costs. Overall, the results on returns to density and sale corroborate the previous findings in the general operating costs analysis.

The research also evaluates the main determinants of train service costs, among them several labour-related explanatory variables. Wages and labour quantities play a key role in train service costs, quantifying the relative weight in train service cost with respect to other inputs such as traction energy. Likewise, the relation between wages and labour hours hired by the metros is assessed, concluding that metros do not adjust train service staff hours when they face wage increases.

Besides this, the role of other labour variables such as labour hours outsourcing and driver productivity has also been analysed, adding new explana-

tory variables that have not been tested before in the rail academic literature on cost efficiency. For instance, both outsourcing and driver productivity increase metro cost efficiency. Outsourcing tends to have a stronger impact than driver productivity but, due to the low variance of the variable, the coefficient estimates are not significant for all the models. On the contrary, the coefficient estimates for driver productivity are lower than for train service staff outsourcing, although they are all statistically significant. Consequently, the potential gains on cost efficiency from an increase in driver productivity have also been quantified.

The complexity and equipment level of the rolling stock, included in the analysis with the proxy level of the percentage of rolling stock with air conditioning, seems to have a negligible effect on the train service cost efficiency. Likewise, train capacity seems to have no role in the cost efficiency of the train service cost per passenger journey, so larger trains do not increase efficiency in transporting passengers.

Lastly, the network age estimates show that age, once controlled for all the other explanatory variables, does not have a role in determining the cost efficiency in train service. However, this is challenged by other models where newer metros, which started operations after 1980, show slightly higher train service costs per car kilometres. Consequently, the train service models confirm the findings of the operating costs analysis with regard to how age affects metro cost efficiency.

7 — Station Service Costs in Urban Rail Systems

Note: This chapter is an extended version of an article presented at the *Transport Research Board* - Session 16-2142 in Washington D.C. in January 2016.

7.1 Introduction

Most passenger journeys in urban rail systems are characterised by relatively short journey times and distances in comparison with average railway passenger journeys. Therefore, a significant share of the time spent by the passengers in the network is at the stations, either in egress time (i.e. entry and exit from stations), access to the platforms or waiting for the train arrival. Thus, the actual travelling time on the rolling stock may not be significantly longer than the actual time spent in the stations. Furthermore , despite the wide range of staff employed by the urban rail systems, it is usually only at the station that the metro staff is visible and accessible to the passengers. Therefore, stations are fundamental for the urban rail service and also for

the passenger experience in the urban rail system.

Parallel to its importance, station service costs also represent a significant share of the total service operating costs of a urban rail system. In the CoMET and Nova metros, station service costs represents around 20% of the total service costs, but some metros with air conditioned stations, abundant elevators and escalators, complex station design or high staffing levels may reach 60% of the total service costs. Thus, for some metros, station service costs are higher than train service operating costs. Moreover, despite the great importance of stations in the passenger's experience and the operating costs incurred by the metros, there is a gap in the rail literature on what makes a rail operator more efficient in terms of station service operating costs.

Consequently, this research tries to explain this variability in the station service costs, building upon the evidence provided by the existing academic research reviewed in Section 7.2.

7.2 Literature Review

Most of the literature on rail operation costs and its determinants has generally focused on rail services where stations are not critical, such as suburban, long-distance and freight services (e.g. Cantos et al., 1999, Caves, Christensen and Swanson, 1980, Kumbhakar, 1988, Oum and Yu, 1994).

In some of this research, stations have been included in the analysis as a way to identify if there are returns to scale (e.g. Santos et al., 2013). In the case of urban rail systems, the size is usually considered as the route length

or the number of stations in the network.

In a similar manner to returns to scale (RTS), rail operations, and particularly station operations, can show significant returns to density (RTD). Economies or returns to density describe the evolution of costs when assets (e.g. stations) are used more intensively. In other words, the RTD describe the relation between inputs and outputs when the output increases but the metro size remains constant.

However, despite the prolific debate on RTS and RTD in railways, few studies have analyzed how the RTS and the RTD change over the different areas of rail operations. Almost all the research is done at an aggregate level which, although insightful, may be misleading if the effect is not considered at a more disaggregated level (i.e. station service level) of the different operations of the rail operator. In this sense and to the best of my knowledge, there is only one research study on station cost efficiency for urban rail systems by Quddus et al. (2007). This research analysed the effect of the levels of demand (i.e. total passengers, interchange and entry passenger levels) on the total station costs. It was found that an increase in demand tends to increase total costs less than proportionally, which leads to significant returns to density in station service costs, as a more intensive use of the assets increases costs less than proportionally. This is in line with previous research on demand and total overall costs in urban rail systems (e.g. Graham, 2008, Santos et al., 2013, Tsai et al., 2015) where population density and passengers per network length were also found to increase urban rail efficiency. Analogous results on the total operating costs have been found for railway operators (e.g. Filippini and Maggi, 1993, Perelman and Pestieau, 1988).

Likewise, Quddus et al. (2007) also identified other costs factors such as number of station platforms, service levels at the peak hours and use of air conditioning. All these factors were found to significantly increase the total station costs for the metros. Similarly, Santos et al. (2013) also found that a larger number of stations was related to lower cost efficiency of urban rail systems.

Overall, there have been few studies on urban rail cost efficiency, and particularly at the disaggregated level of the urban rail operator such as station costs (e.g. for train service costs see Brage-Ardao et al., 2015). This study contributes to the field by adding a more comprehensive research of the station service costs, controlling for the main cost factors used in more aggregate research on cost efficiency and also adding new explanatory variables not considered before.

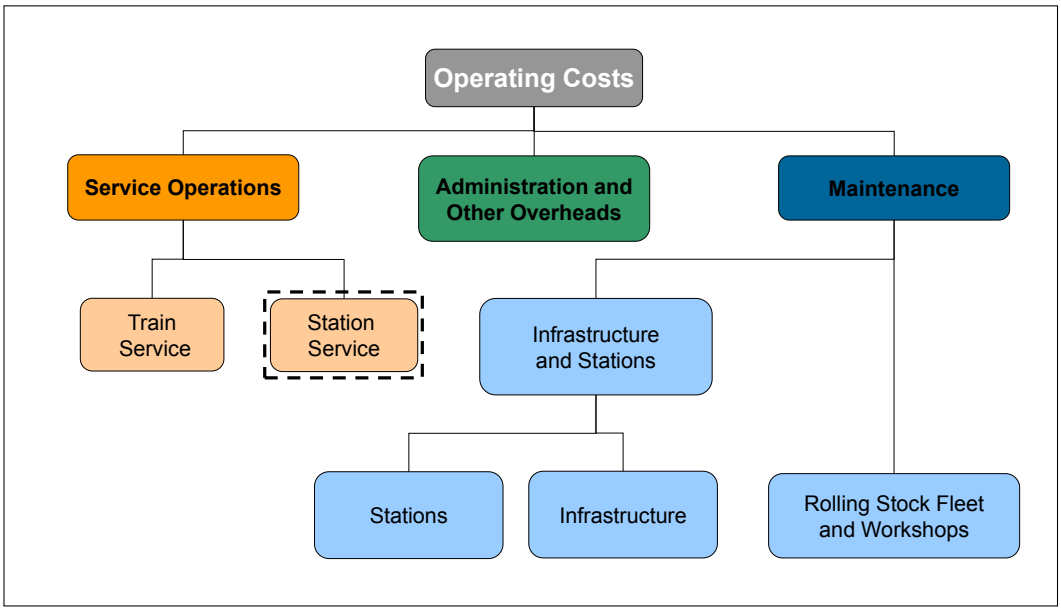
7.3 Data and Research Hypotheses

The data for the urban rail systems comes from the CoMET and Nova consortia as described in Chapter 4. The econometric models presented in this chapter used an unbalanced panel data set for a 10 year period (2005-2014) including 20 unique metros. The panel is unbalanced as some urban rail systems miss data for some years, reducing the number of observations significantly.

The dependent variable of the econometric models is station service costs which comprises the costs associated with the following tasks:

- All types of platform supervision, platform assistants and multi-functional

Figure 7.1: Station service costs and other operating costs of a typical CoMET-Nova metro operator.



duties.

- All station energy or power consumption including escalators, air conditioning and heating.
- All staff at ticket offices and gate lines.
- Salaries, training, overtime and employment-related fringe costs for metro staff and any contract labour cost working at the stations.

Within this category, there are certain activities that take place at the stations but are excluded from the station analysis, such as:

- All tasks related to cash handling, counting and transportation, ticketing and revenue control.
- Any costs from security staff.

- Station cleaning, which in the CoMET and Nova framework, is considered as part of station maintenance cost and not station service.

In order to model the station service costs, they have been rationalised by output (i.e. passenger journeys) and by scale of operations (i.e. number of stations). Table 7.1 presents the summary statistics for the station service costs per passenger journey and per station; Figure 7.2 and 7.3 show the corresponding anonymised values. The complete definitions of all the variables used in this chapter are available in the Annex Table 11.7 on page 292.

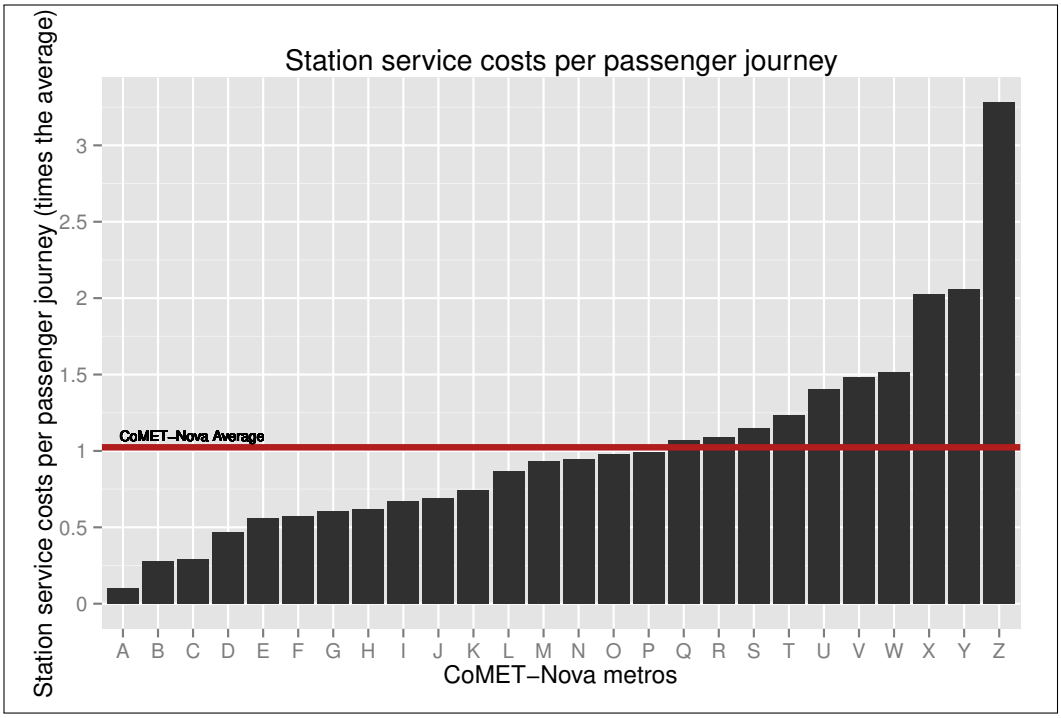
The average station service costs per passenger journey is 0.15 US$ in PPP terms, although most of the members show slightly higher costs (i.e. 0.18 US$) and there is a great diversity in the station service costs per passenger, with a standard deviation of 0.15US$ and a maximum station service cost for a given year of 0.79US$ per passenger journey. For station service costs per station, the proportions are similar, with small differences between the mean and the median but with large dispersion in the tails. This indicates that there is a group of metros close to the average (720,000US$) but also extreme cases with small suburban non-staffed stations (30,000US$) and, conversely, large underground fully air-conditioned ones (3.51 US$ million).

Table 7.1: Station service dependent variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max
Station service costs per passenger journey	0.15	0.18	0.15	0.01	0.79
Station service costs per station (millions)	0.72	0.71	2.72	0.03	3.51

Note: Costs in Parity Purchasing Power (PPP) U.S. Dollars. Obs.: 121.

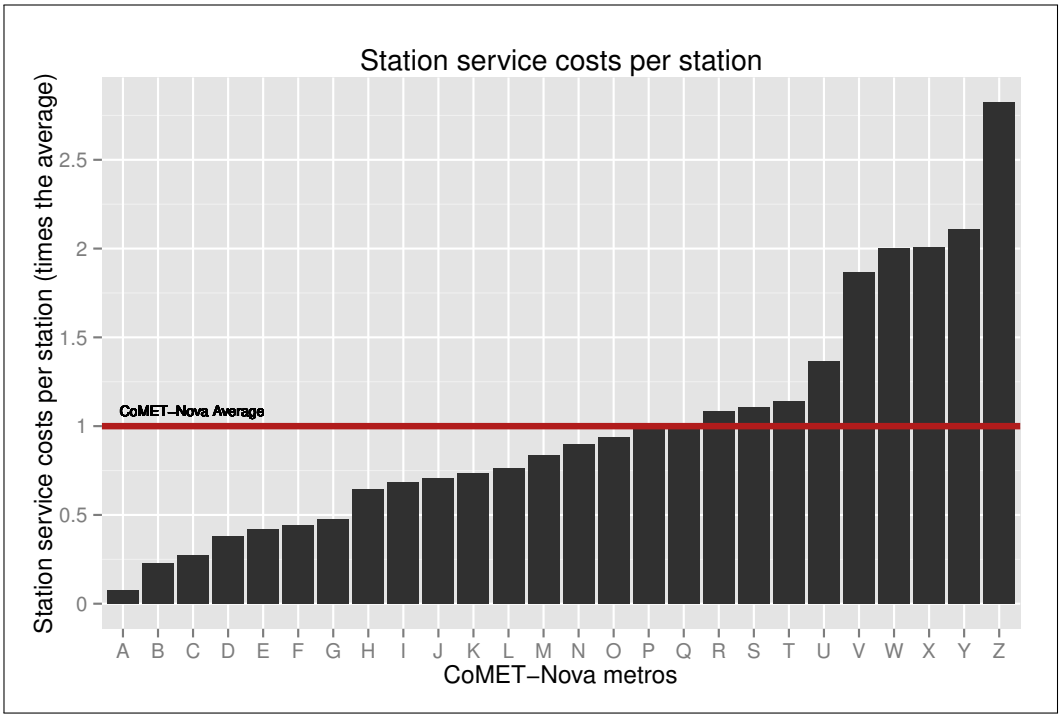
Figure 7.2: Station service costs per passenger journey.



The summary statistics for the explanatory variables are presented in Table 7.2. The first variables correspond to annual passenger journeys, in millions, and the number of stations. The demand levels of urban rail systems included in this research range from 2,573 million passenger journeys to only 29 million with an average of 381 annual passenger journeys. The median is 385 million passenger journeys, so there are several metros clustering in the average despite the large standard deviation of 639 million. In the case of stations, the smallest system only has 14 stations whereas the largest one comprises 424. The average number of stations is 81 with a standard deviation of 96, which implies a significant variation in the number of stations although not as large as for passenger journeys.

Input prices and quantities have also been considered in the modelling of

Figure 7.3: Station service costs per station.



station service costs. The average gross labour costs per hour corresponds to the overall metro average labour cost due to the impossibility of breaking down the station service wages from the CoMET-Nova KPI framework. The average gross labour cost per hour in parity purchasing power (PPP) US Dollars is almost $31, with a range from $11 to $75 depending on the metro. This corresponds to the diversity of the CoMET and Nova consortia which includes urban rail systems in world capitals and also cities from developing countries. Nevertheless, the median of the gross labour wage is $33 per hour, which is in line with the average and shows that, despite some extreme values, there is a numerous group around the mean. Electricity prices per Mega-Watt hour (MWh) also show a significant variation in extreme values ($50 to $1030 PPP US Dollars per MW hour) even though the average and

the median both converge around $150 PPP USD per MWh. With regard to input quantities, they are less informative as they measure the total labour hours and total MWh consumed. However, they are important to control for how labour- and energy-intensive the different urban rail systems are. Furthermore, by comparing the effect of input prices on cost efficiency when controlling for input quantities, it is possible to assess how changes in the input prices lead to changes in input productivity.

There are also variables related to the operations of the station service. For instance, the percentage of contracting out of station staff hours, assuming that contracting out can be a key factor in increasing cost efficiency. Thus, this variable explains the percentage of station staff hours contracted out where station hours are defined as the total hours scheduled for station management, platform supervision, part-time platform assistants, multifunctional duties and operational control of subcontractors used for station operation duties. Overall, most of the CoMET and Nova metros do not outsource the station service labour hours, as the median of the variable is zero. However, a small number of metros can outsource as much as 53% of their station service hours.

A key characteristic of the stations with impact on the station service costs, and particularly the energy consumption, is the number of lifts, escalators and travelators. Each of the variables has been tested individually and, due to the lack of observations, the three variables have been combined into one. The average number of lifts, elevators and travelators in the network ranges from 63 to 2145. This absolute number is controlled by the number of stations and passenger journeys which have also been included in the

Table 7.2: Station service explanatory variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max	Obs
Passenger journeys (m)	381.614	385.046	639.898	28.691	2572.9	N = 122
Number of stations	81.151	80	96.903	14	424	N = 122
Wages (PPP US$/h)	31.495	33.047	16.324	11.575	75.217	N = 122
Electricity price (PPP US$/kWh)	148.443	157.42	102.292	50.817	1030.105	N = 122
Staff hours - Station service (m)	3.311	3.11	5.586	0.058	28.283	N = 122
Non-traction electricity (MWh)	113.361	90.379	170.711	4.59	627.83	N = 122
Outsource station staff (%)	11.318	0	18.342	0	53.3	N = 122
Escalators and lifts	444.667	384	558.288	63	2145	N = 122
Train length	109.524	113.071	35.871	60	193.189	N = 122
Network age (years)	51.512	41	37.848	2	151	N = 122
Year				2005	2014	N = 122

Note: Units in PPP US$ refer to Parity Purchasing Power (PPP) U.S. Dollars.

model. Overall, there are significant differences between the average number of escalators, lifts or travelators (i.e. 445) and the corresponding median and standard deviation values: 384 and 558 respectively. This suggests that there is a large diversity in the different accessibility policies followed by the CoMET-Nova members.

Additionally, train length is also included as a proxy to take into account the length of the platforms as well as the approximate size of the stations. Train length can also serve as a proxy for capacity, as longer trains typically tend to have higher passenger-carrying capacity, which may lead to additional station costs (e.g. the need for more staff to manage larger passenger flows). The train length equals the average car length multiplied by the average number of cars per train. The expected effect of this variable on the station service costs is unclear. Presumably, larger stations will have higher station service costs per station but, at the same time, the station service costs per passenger may be lower.

Lastly, the age of the network is also considered, measured since the beginning of metro operations and not weighted by extensions. This variable can be a proxy for the assets condition and it can also capture the effect of some decisions adopted at the initial design of the metro that can still be affecting the cost efficiency of the metro (e.g. tunnel dimensions and its impact on service).

Other explanatory variables considered but not included

Additional variables to the ones included in the final model have been considered in this research. However, due to data availability, accuracy and

econometric pertinence, the following variables have been excluded from the analysis.

Firstly, the analysis considered demand profile data which describes the average number of passengers travelling in the network per hour. A peak ratio variable was created using this data which refers to the ratio between the highest demand level in the peak hour over the lowest demand level in the inter-peak. As a general rule, higher inter-peak ratios show more commuter-like demand, with high morning peaks and low demand in inter-peak hours. On the other hand, lower values for the inter-peak ratio show a more constant demand profile throughout the day. It is assumed that higher peak ratio values would entail higher station service costs per passenger journey as the resources used to manage the peak hours tend to be underused during the off peak time (i.e. labour shifts usually last longer than the peak). Unfortunately, the data on the demand per hour does not specify where the focus is within the network and, probably for this reason, the coefficients for this variable were never significant. Furthermore, the critical factor is not only the demand profile over the course of the day but also the network shape, how the demand breaks down into the different stations and how bidirectional it is along the different lines of the network.

Similar variables to the peak ratio were tested, such as the number of hours with more than double, three times or four times the lowest inter-peak demand. However, they suffered from the same drawbacks as the peak ratio variable described above. Besides this, many metros do not adjust their resources to respond to the fluctuations in passenger demand and station crowding. The CoMET-Nova members replied to a questionnaire on

whether they deploy more staff during peak hours in comparison with inter-peak hours or whether the station staff shifts were fixed length (i.e. 8 hours). Unfortunately, the ratio of response on these questions was very low and, furthermore, a large majority of the metros followed the same policies, so there was not enough variance to include these binary variables in the research. Another question on whether the metro uses multifunctional station staff in small stations with lower demand suffered the same limitations.

There are other variables on the station properties that were discarded due to low data availability. These correspond to the variables on the number of stations with platform screen doors and the number of stations with air conditioning. In the case of the platform screen doors, it may reduce the staffing requirements but the variance of the variable is very low as most metros do not use platform screen doors. With regard to air conditioning, the data availability was low, so including this variable would have substantially reduced the total number of observations of the model. However, it is already partly included as part of the energy consumption variable.

7.4 Results

This section presents the estimates for the station service cost function models. The dependent variable of the models, station service costs, has been rationalised with respect to output (i.e. station service costs per passenger journey) and scale (i.e. station service costs per station). Table 7.3 presents the results from the estimation of the station service costs following a random effects approach already explained in section 3.2.1 (page 72).

The appropriateness of this econometric is based on the existence of non-constant variance across entities shown by means of a Breusch-Pagan test (Table 11.17 on page 304). The econometric models have been estimated using logarithms ($[Ln]$) in order to provide cost elasticities. There are other variable units whose usage was more advisable, although the interpretation is based on the additional unit (e.g. year, metre) and not on percentage change. The model estimates shown in Table 7.3 represent the general Equation 7.1. Equation 7.1 regresses the station service cost category (i.e. C_{st}) on logged (i.e. x_j) and not logged (i.e. x_k) explanatory variables.

$$LogC_{st} = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \sum_{k=1}^{K} \beta_k x_k + \alpha_i + \varepsilon_{it} \tag{7.1}$$

The first explanatory variables considered are those related to the passenger journeys and the number of stations. Both of these variables are statistically significant in their corresponding model. The negative coefficient estimates suggest that, as the output (passenger journeys) or the scale (number of stations) increases, the unit station service costs decrease. A 10% increase in passenger journeys is correlated to a 15% decrease in station service costs per passenger journey, reflecting very significant economies of density. The variable is statistically significant at a 99% confidence level, although this significance is with respect to the null hypothesis of having a zero coefficient. However, given this large coefficient, there must be a signalling effect that explains why the coefficient is larger than 1. This signalling effect would mean that more passenger-intensive metros are usually more efficient in terms of station service costs per passenger journey.

Table 7.3: Station Service Cost Model Results - Random Effects

	Station service costs per	
	Pax. journey	Station
	(1)	(2)
Pax Journeys [Ln]	−1.517***	
	(0.174)	
No. of stations [Ln]		−1.177***
		(0.222)
Wages [Ln]	0.464*	0.463*
	(0.251)	(0.256)
Elect. Price [Ln]	−0.119	−0.131
	(0.111)	(0.117)
Station Staff hr [Ln]	0.678***	0.502***
	(0.153)	(0.143)
Non-Traction elect. [Ln]	0.326*	0.267
	(0.182)	(0.184)
Staff outsourcing [Ln]	−0.361	−0.285
	(0.389)	(0.398)
Lifts and esc. [Ln]	0.417**	0.297
	(0.198)	(0.198)
Train length	0.013***	0.009***
	(0.003)	(0.003)
Network age (years)	0.006	0.005
	(0.004)	(0.005)
Time trend	0.011	0.004
	(0.011)	(0.012)
Constant	−22.354	−8.785
	(22.394)	(23.780)
Rho	0.8447	0.8252
Observations	122	122
R^2	0.480	0.448
Note:	*p<0.1; **p<0.05; ***p<0.01	

On the contrary, in the scale measure, a 10% increase in the number of stations is shown to lead to a 12% decrease in station service costs per station. This is probably reflecting the fact that metros with more stations typically tend to have smaller and simpler stations, and vice versa. For example, newer Asian metros tend to have proportionally fewer stations than older European or North American metros but those stations are significantly larger, leading to higher station service costs per station. Furthermore, the stations of metros located in developing countries tend to serve very dense areas, which usually entail more station equipment (e.g. lifts/escalators), higher passenger volumes and air conditioning; all factors which lead to higher total station service costs.

With regard to the input factors, only the input price of labour is significant at a 90% confidence level with a coefficient close to 0.50, implying that station service costs would increase by roughly half the increment in overall metro wages. On the contrary, the price for electricity shows statistically insignificant estimates. This correctly shows the relevance of wages in determining station service costs, but also suggests that, because the costs do not increase proportionally, there are some trade-offs occurring. For example, metros may increase productivity and/or automation in response to higher overall wages, with net costs therefore not increasing proportionally to the change in wages.

The productivity effect can be assessed if the wage coefficient in Table 7.3 (i.e. β_{fixed_L}) is compared with the coefficients of the same model without input quantities (i.e. $\beta_{flexible_L}$), shown in the Annex (Table 11.15 on page 302). In the model without fixed labour quantities, the input quantities are allowed

to vary, so a wage hike could be compensated by a drop in input quantities hired (i.e. less labour hours hired). The difference in the coefficients for wages would correspond to the differences in the labour productivity, usually adjusted by means of labour hours hired. The null hypothesis shown in Equation 7.2 is that the difference between these two coefficients is statistically insignificant. As explained in section 7.3, the wages refer to the overall metro wages and not the particular station service wages due to the categorisation of costs in the CoMET-Nova KPI framework. However, the analysis still approximates which labour hours are reduced when overall wages increase.

$$H_0 : \beta_{fixed_L} - \beta_{flexible_L} \simeq 0 \tag{7.2}$$

In the case of the station service costs per passenger journey, the coefficient for the model with quantities, Table 7.3, is 0.464 whereas the coefficient for the model without quantities is 0.064. Assuming as standard errors the average of the corresponding standard errors as shown in Equation 7.3, it is possible to test if the t-value of the difference (i.e. θ) is statistically significant in a t-distribution with the corresponding degrees of freedom. In this case, the critical t-value for a 90% confidence level with 111 degrees of freedom (i.e. t_{n-k}) is 1.659. Therefore, 1.53 would fail to be statistically significant at a 90% confidence level for a t-distribution with 111 degrees of freedom, although close enough as to be indicative. This difference between wage estimates suggest that metros with higher wages tend to decrease the

labour hours hired for the station service tasks.

$$\theta = \frac{\beta_{fixed_L} - \beta_{flexible_L}}{(\sigma_{fixed_L} + \sigma_{flexible_L})/2} \simeq 1.538 \tag{7.3}$$

With regard to the input quantities in the original econometric model shown in Table 7.3, the coefficient for both quantities are positive and statistically significant for the station service cost per passenger journey, as expected from the economic theory.

Other labour coefficients, such as the level of station staff outsourcing, show negative coefficient estimates for both models in Table 7.3. However, both coefficient estimates fail to be statistically significant at a 90% confidence level. This is probably due to the small variation in the values of this variable across the members of the panel, as most of them do not outsource any of their station labour hours. More observations and more variability in the level of outsourcing would help to obtain statistically significant estimates. In this case, the station service estimates contradict the previous findings for the train service and the overall operating costs estimates on staff outsourcing.

In addition to the previous cost factors, the asset conditions and design were also considered in the analysis. The combined variable of the number of lifts, elevators and travelators has a positive coefficient, suggesting that more equipment leads to higher costs, although the variable is not statistically significant in the station service costs per station model. On the other hand, train length, which is a proxy to measure the length of the platforms in the stations, shows a significant estimate at a 99% confidence level for the station

service cost per passenger journey model. As noted earlier, this variable is considered in metres rather than in log form to aid the interpretation of the estimates. Consequently, the results suggest that if train length is increased by 19 metres, the average car length for CoMET and Nova metros, then station service costs per passenger journey are expected to increase by 1.13% (i.e. $19 * 0.013$).

Lastly, the age of the network, based on the original opening date of the metro, is not significant in any of the models. However, the variables of age are significant in the models where age is broken down into old metros, medium age metros as the base case and new metros shown in the Annex (Table 11.9 on page 295). In this case, old metros which started operations before 1940 and new metros, which started after 1980, both show higher station service costs. This pattern has been shown in the overall operating cost models and also in the train service costs analysis. Lastly, the time trend cannot discard the null hypothesis of being zero, so it seems that, once controlled for all this factors, there is no significant trend in the stations service costs for any of the models.

Finally, Table 7.3 reports the *rho*-coefficient which refers to the percentage of the variance attributed to the differences between metros. In this case, the differences between the metros predominate over the differences over time within the same metros. Up to 80% of the variance in the model is due to these differences between the metros in the panel.

Based on the 122 final observations included in the model, Table 7.3 reports the R^2, the proportion of the sample variation in the station service costs that is explained by the regressor variables in the econometric model.

The econometric model explains roughly half of the variance in the station service costs per passenger journey and around 45% in the case of the station service costs per station. These goodness of fit values suggest that the models explain a substantial part of the cost variations, but there is another half of the variation left unexplained. This is due to several factors such as the omission of the variables that are not possible to include in the model described in subsection 7.3. Additionally, given the high *rho*-coefficients, a significant part of the station service costs may be linked to metro-specific characteristics and not so much to the general explanatory variables that can be included in the model.

7.5 Conclusion

This research assesses the determinants of station service cost efficiency in urban rail systems. The analysis uses an unbalanced panel of 20 metros, members of the CoMET and Nova benchmarking group, for the period 2005-2014. The two models present cost elasticity estimates for the station service costs per passenger journey and station service cost per station explaining around 50% of the station service cost variation from the mean.

The results show significant economies of density in the station service costs per passenger journey and significant economies of scale for the station service costs per station. The economies of density are expected according to economic theory, as the marginal cost of extra passengers for the station service is usually close to zero. However, the coefficients present cost elasticities below -1, which implies that these estimates are likely to mask some metro

signalling effects. Thus, metros with more passenger journeys may tend to be more efficient with regard to station service costs per passenger journey due to other explanatory variables not included in the model. Similarly, metros with more stations seem to have lower station service cost per station; probably due to the fact that, despite economies of scale, metros with more stations have different type of stations (i.e. small and suburban).

Inputs such as labour and electricity also show a statistically significant effect on the station service costs. Labour wages and quantities are particularly significant and the results hint that higher wages lead to a substantial reduction in station labour hours. This could be potentially offset by means of outsourcing, which shows negative coefficient estimates for both models. However, staff outsourcing coefficient estimates fail to be statistically significant due to the low number of observations included in the model and the scarce variability in outsourcing policies among the CoMET-Nova metros.

Asset characteristics, such as the number of lifts, escalators and travelators seem to substantially increase the station service costs and show statistically significant coefficients. The train length, a proxy for platform and station size, also shows positive coefficients, meaning that more equipped and larger stations tend to be more expensive.

Lastly, older metro systems tend to have higher station service costs although the coefficients are not statistically significant. This is also tested by means of other metros where binary variables for old metros show significant coefficients, corroborating the findings in the overall operating costs and train service chapters. Finally, the time trend is statistically insignificant, demonstrating a lack of an overall pattern over time, once all other

factors are taken into account.

8 — Rolling Stock Maintenance Costs in Urban Rail Systems

Note: This chapter is an extended version of the journal article *Determinants of Rolling Stock Maintenance Cost in Metros* accepted for publication in the *Journal of Rail and Rapid Transit* in October 2015.

8.1 Introduction

Asset management and asset maintenance have received significant attention in the academic literature. Maintenance is relevant in all industries, but it is particularly critical for capital-intensive sectors such as railways and urban rail metros. In the case of metros, the urban rail systems require both a large physical infrastructure and valuable rolling stock assets. According to a recent research study on rail costs (Jan and Phillips, 2011), rolling stock maintenance costs over the life cycle of the rolling stock can exceed the actual acquisition costs. Therefore, it is critical for the cost efficiency of any rail operator to understand the determinants of rolling stock maintenance.

Although there is an abundant literature on rail maintenance and cost

efficiency, it is mainly focused on the cost efficiency of infrastructure maintenance, and particularly for freight and long distance railway operators. There is more data available on infrastructure maintenance due to the numerous operators in the US which maintain their own infrastructure. In Europe, the privatisation of the network (e.g. Preston, 1996) and the more recent investment plans announced by several governments (e.g. Jowit et al., 2012) have also increased the academic interest on rail infrastructure maintenance.

On the other hand, there is a profuse literature on rolling stock maintenance from a technical point of view, analysing life cycles and optimising replace timings of rolling stock constituents. However, there are few analyses on the cost efficiency of rolling stock maintenance itself. Therefore, this chapter fills the gap in the literature by focusing on rolling stock maintenance costs for urban rail systems, although some of its conclusions can also be applied to other types of passenger rail services.

8.2 Literature Review

Although the literature on maintenance and rail operations is abundant, there are still significant gaps in the understanding of maintenance in the rail industry. Most of the literature on rail maintenance has focused on track maintenance (e.g. Andersson, 2008, 2011, Marti et al., 2009, Tervonen and Idstrom, 2004), spurred on by the debate on infrastructure privatisation of Network Rail in the United Kingdom (BBC, 2007) and Amtrak in the United States (The Economist, 2011). Therefore, despite the importance of track and infrastructure maintenance, most of the academic literature on

rail maintenance efficiency disregards other forms of fixed assets (e.g. power system maintenance or station maintenance) and also other rail mobile assets like the rolling stock.

Nevertheless, there is a significant corpus of research considering rolling stock maintenance from a non-economic point of view. Most of the analysis on rolling stock maintenance refers to technical areas such as how to effectively monitor the condition of the assets in order to improve failure diagnosis (e.g. Barke and Chiu, 2005, Bladon et al., 2004). Another relevant field of research on rolling stock maintenance focuses on the prediction of failures (e.g. Kumar et al., 2008), which is essential for improving maintenance strategies and optimising replacement schedules (Wang et al., 2007). This also relates to more specific research on how to make the best use of spare provisions and the frequency of rolling stock replacements (e.g. Cheng and Tsao, 2000). However, although all this research can improve the performance of the rail operators in rolling stock maintenance, it is either too focused on some of the rolling stock maintenance tasks (e.g. improving the diagnosis of wheel damage) or it offers general research on strategic considerations that are not tested with real data from rail operators.

Although they are scarce, there are some examples of academic research on rolling stock maintenance from the cost efficiency perspective. For instance, Wang and Liao (2006) analysed the productivity growth of the Taiwan rail operator for the period 1991-2000. The research developed a multiproduct cost model where there were four categories of rail variable costs; one of these categories is the rolling stock maintenance costs. The maintenance costs for rolling stock are calculated as an index. The index equals the

division of the total maintenance expenses by the total hours of rolling stock repair and maintenance. Wang and Liao (2006) found small elasticities for all input factors including rolling stock labour prices. Therefore, changes in input factors such as rolling stock maintenance labour would have a small effect on the labour used. However, on the one hand a substituting effect was also found between labour, maintenance and materials and supplies, and on the other, a complementary relationship between maintenance and intermediate inputs. This work offers an analysis of rolling stock maintenance costs and the elasticities of the costs with respect to other factors. However, the measurement of the rolling stock maintenance costs is a proxy itself; an index assuming that the costs of rolling stock are proportional to the share of the labour working on rolling stock maintenance with respect to other maintenance tasks. Furthermore, the research is based on one operator. This is not representative of the average rail operator cost efficiency in rolling stock maintenance, due to multiple factors such as the particular type of rolling stock used by the Taiwan rail operator, the network characteristics or the particular labour agreements of the rolling stock maintenance staff at the Taiwan rail operator.

The lack of academic research on rolling stock maintenance cost efficiency can be explained by the scarcity of data publicly available. In the case of infrastructure maintenance, where the academic literature is more abundant, most of the owners and rail infrastructure managers are state-owned companies. These companies tend to be more prone to making data available for research, particularly following the debate on their privatisation. In the case of rolling stock maintenance, there are many more operators but they tend

to be private companies less prone to publicly sharing their operational and cost data for academic research purposes.

The lack of academic research on rolling stock maintenance cost efficiency has been partly covered by the research of the rail regulatory institutions. In the United Kingdom, both the Department of Transport, the Association of Train Operating Companies (ATOC) and the Office of Rail Regulation promoted a report (Jan and Phillips, 2011) on whole life costs in rolling stock assets, which analyses how rolling stock costs evolve over time. According to Jan and Phillips (2011), rolling stock maintenance costs represent 44% of the rolling stock whole life cost breakdown, whereas capital cost and operating costs represent 31% and 25% of the whole life costs respectively. The authors also reported other factors that can increase rolling stock maintenance costs, such as the number of rolling stock types in the fleet. Jan and Phillips (2011) estimate that fleet complexity measured as the types of rolling stock can increase maintenance costs by 15% to 20%. On the other hand, the report mentions other factors (e.g. train mid-life refurbishment practices) but it does not quantify their impact on maintenance costs. This study also found that internalising the responsibility for rolling stock maintenance could lower maintenance costs by up to 10% for the average train operator. However, few other factors were quantified so the effect found in the internalisation of maintenance could be due to other factors not controlled in the report. Likewise, most of the insights referred to topics such as maintenance franchise and franchising schemes rather than actual operational characteristics of the rolling stock maintenance.

Overall, it is clear from the literature review that there is a gap in the

field of academic and industry research on rolling stock maintenance costs. Therefore, this research on the determinants of rolling stock maintenance cost efficiency complements the previous research on technical rolling stock maintenance and cost efficiency on general rail maintenance.

8.3 Data and Research Hypotheses

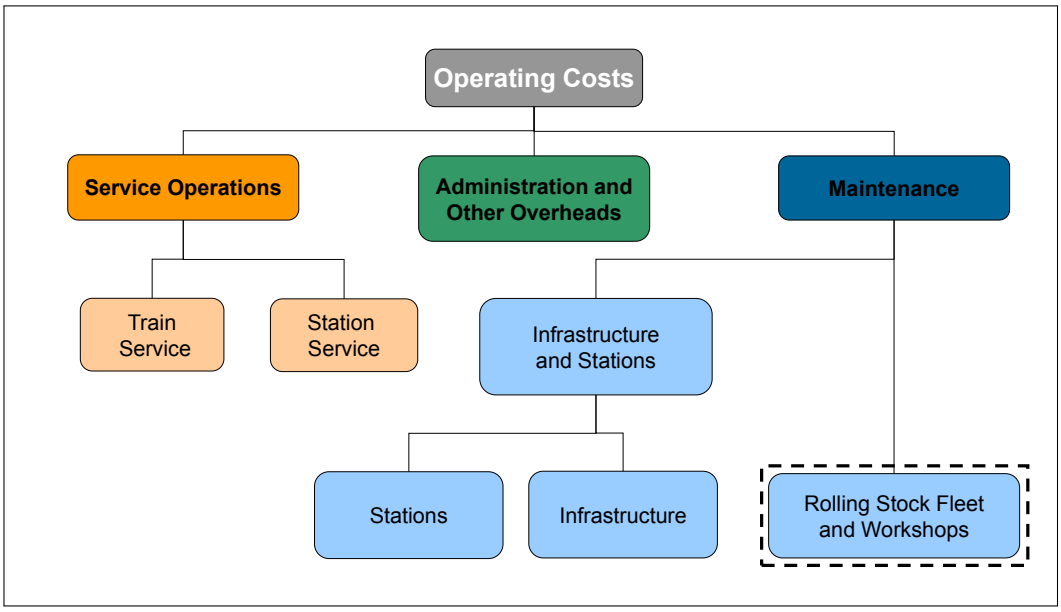
The data source for this chapter is the CoMET-Nova KPI database as explained in chapter 4 on the data used for this research. The relation between the operating costs, other operating costs constituents and the rolling stock maintenance costs is shown in Figure 8.1. The variables explained in this section correspond to Equation 8.1 where the rolling stock maintenance cost (C_{R_Stock}) is explained by a set of explanatory variables (i.e. x_j).

$$LogC_{R_Stock} = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \alpha_i + \varepsilon_{it} \tag{8.1}$$

According to the definitions of the CoMET-Nova KPI handbook, the rolling stock maintenance costs comprise the following costs:

- All maintenance procedures of rolling stock maintenance such as routine checks of key parts, lighting or maintenance of door open-close mechanisms.
- All rolling stock cleaning, including both regular and in-depth cleaning operations of rolling stock in the workshops.
- All management and support staff costs associated with the rolling

Figure 8.1: Rolling stock maintenance costs and other operating costs of a typical CoMET-Nova metro operator.



stock maintenance and the workshops.

- All maintenance of the rolling stock workshops.
- Salaries, training, overtime and employment-related fringe costs for metro staff and contract labour cost for any of the rolling stock maintenance tasks.

On the other hand, certain rolling stock maintenance costs are excluded as they represent non-recurring operational maintenance costs. Therefore, as this research is focused on ordinary rolling stock maintenance costs, the data excludes any expenditures associated with the acquisition of new rolling stock, mid-life rolling stock refurbishment and renewal of expired rolling stock assets.

In order to compare the rolling stock maintenance costs of different rail

operators, these have been rationalised by two different outputs. The two outputs considered for the rolling stock maintenance costs are as follows:

- Car kilometres: All car kilometres which were actually operated in revenue service, excluding empty stock movements, movements from depots, engineering trains, driver training runs, cancellations of scheduled runs and rail replacement bus services.

- Fleet units: Total number of cars owned or leased by the metro which are suitable for operation in normal passenger service. Due to continuous change in fleet numbers entailing the addition of new fleets or retirements of old fleet, the total fleet unit equal the average fleet units over the year.

Consequently, the two dependent variables of this research are, firstly, rolling stock maintenance costs per car kilometre, and secondly, rolling stock maintenance costs per fleet unit (or per car). The anonymised rolling stock maintenance costs per car kilometre and per car are shown in Figure 8.2 and Figure 8.3 respectively. The basic descriptive statistics of the dependent variables can be found in Table 8.1. The costs in Table 8.1 are shown in Parity Purchasing Power (PPP) units in order to enable international comparisons as explained in Section 4.3 on page 113.

The average maintenance cost per car km is 0.704 US$ PPP although the range of values range from 0.248 $ to 3.014 $ in PPP US$ PPP. Despite these extreme values, most of the members concentrated around the mean, given the low value of the standard deviation (0.315 US$ PPP).

Table 8.1: RS Maintenance dependent variable summary statistics

Variable	Mean	Median	Std.Dev	Min	Max
RS maintenance costs per car km	0.704	0.691	0.315	0.248	3.014
RS maintenance costs per car	74906	78728	25562	28846	189865

Note: Costs in Parity Purchasing Power (PPP) U.S. Dollars. Obs.: 142.

On the other hand, the annual average rolling stock maintenance costs per car for the CoMET-Nova metros included in the research panel is 74,906 US$. The values range from a minimum of 28,846 US$ to a maximum of 189,865 US$ in PPP with a modest standard deviation of 25,562 US$. In both cases, the rolling stock maintenance costs per car km and per fleet show moderate standard deviations, implying that most of the members cluster around the group average. However, there are also significant outliers that add variability to the data set and may help to obtain statistically significant coefficients in the econometric models shown in section 8.4.

The costs per output shown in Table 8.1, rolling stock maintenance costs per car km and rolling stock maintenance costs per fleet unit are explained by a set of independent variables. The summary statistics of these explanatory variables are shown in Table 8.2. A detailed definition is shown in the annex Table 11.8 on page 293.

The first explanatory variables are the same variables considered for the rationalisation of the rolling stock maintenance costs: car km and fleet units. The inclusion of these variables in the analysis enables us to test the existence of economies of scale and economies of density in the provision of rolling stock

Figure 8.2: Rolling stock maintenance costs per car kilometre.

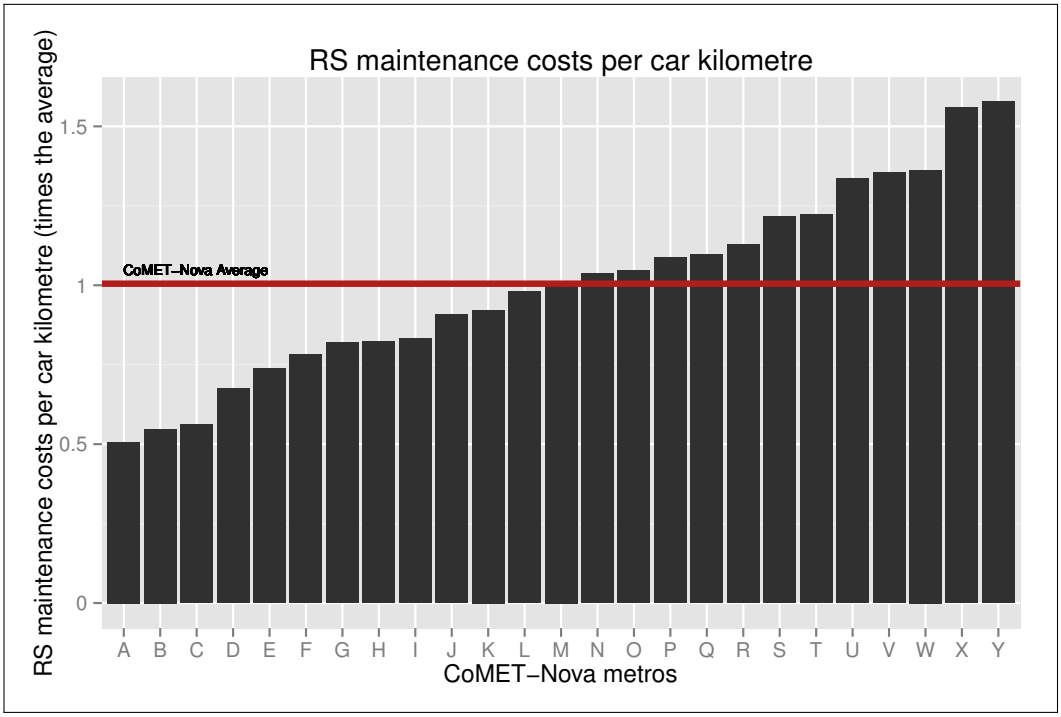


Figure 8.3: Rolling stock maintenance costs per car.

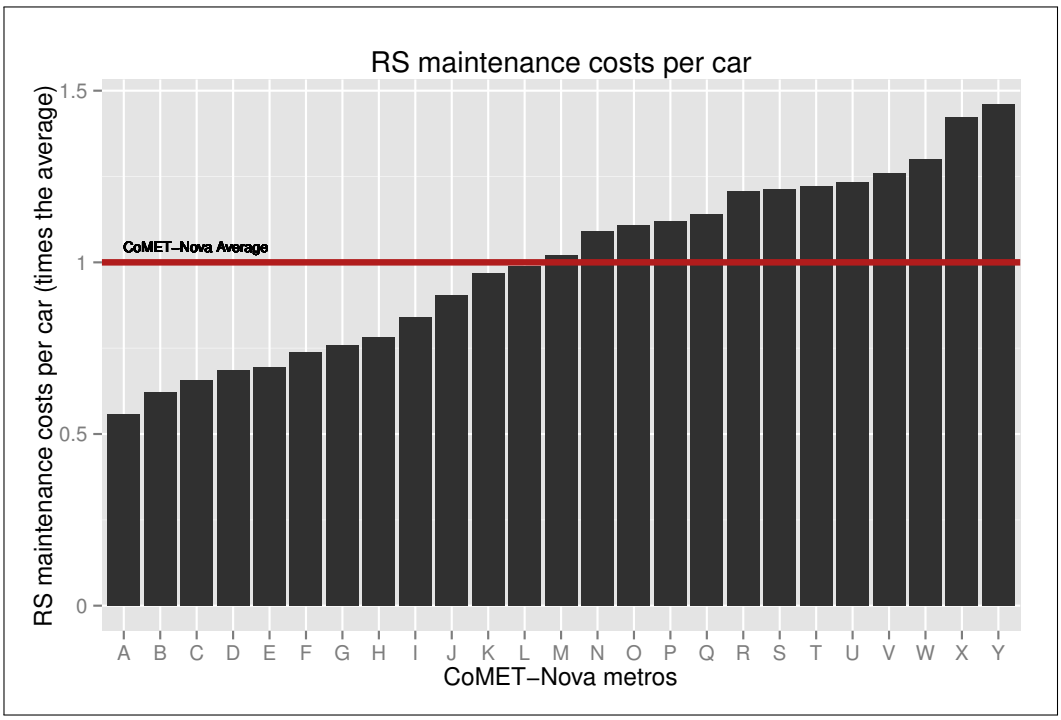


Table 8.2: RS Maintenance explanatory variables summary statistics

Variable	Mean	Median	Std.Dev	Min	Max	Obs.
Car km (m)	107.286	104.662	172.939	5.544	776.35	N = 142
Fleet size	1008.657	846.482	1572.889	78	6417	N = 142
Wages (PPP US$/h)	36.278	40.973	14.99	14.16	75.217	N = 142
Staff hours - RS maint (m)	1.571	1.644	2.537	0.068	13.3	N = 142
Outsourcing staff (%)	0.245	0.216	0.207	0	0.856	N = 142
Fleet available at peak	0.883	0.874	0.073	0.654	1.05	N = 142
Operating speed (km/h)	32.689	32.9	5.248	12	45	N = 142
MDBF - RS	159.458	178.633	72.895	0.019	392.344	N = 142
RS with air con. (%)	0.496	0.52	0.459	0	1	N = 142
Fleet age	20.334	19.647	10.934	3.406	49.754	N = 142
Network age (years)	57.535	43	36.984	10	151	N = 142

Note: Units in PPP US$ refer to Parity Purchasing Power (PPP) U.S. Dollars.
MDBF-RS refers to Mean Distance Between Failures caused by Rolling Stock.

maintenance costs. If the rolling stock maintenance costs increase less than proportionally to the increase in the fleet size, while all the other variables are held constant, this would entail that there are economies of scale in the provision of rolling stock maintenance. According to economic theory, economies of scale may arise from labour specialisation, automatisation of tasks or procurement advantages given the large size of the fleet, among other sources of efficiency gains. In this case, the average fleet size of the metros included in the data set is above a thousand cars, which results in a fleet size large enough to develop economies of scale.

Likewise, if the increase in car km leads to a less than proportional increase in the rolling stock maintenance costs, holding all other variables constant, this would entail the existence of economies of density. The sources of economies of density in rolling stock maintenance costs can arise from the difference in maintenance routines. Some maintenance routines are based on the car km produced by each car, whereas other routines are based on time or normal wear and tear (e.g. car lighting or seat replacement), independent of the car kilometres produced. Consequently, a more intensive production of car km per car may lead to a decrease in rolling stock maintenance costs per car km.

Labour-related factors, such as the metro wage level and the total staff hours of rolling stock maintenance have also been considered in this research. The average gross own staff wages for the CoMET and Nova metros considered in these models is 33.28 US$ PPP per hour, which also includes any benefit or fringe cost associated with the metro workers. Overall, there is a large diversity in the wages paid to their own employees across the CoMET-

Nova metros. This is caused by the different development levels of the cities, the differences in labour regulations between countries and also the particular labour union power for each rail operator. Unfortunately, it was not possible to include a rolling stock maintenance average gross labour cost due to the allocation of costs of rolling stock maintenance labour within the CoMET-Nova KPI framework.

As in the models presented in other chapters of this thesis, input prices are assumed to have an increasing effect on total costs. Therefore, holding all other costs determinants constant, higher labour wages and higher use of labour are expected to lead to higher rolling stock maintenance costs, and consequently, lower cost efficiency. However, this relation does not have to be proportional; higher wages may increase the substitution of labour for capital (i.e. automatisation) or may increase labour productivity, so an increase in wages could correspond to a less than proportional increase in overall rolling stock maintenance costs. Likewise, more labour-intensive metros may have higher maintenance costs but not in a proportional relation, as labour is not the only factor explaining maintenance cost efficiency.

Another labour-related factor included in the analysis is the percentage of rolling stock maintenance staff hours contracted out over the total rolling stock maintenance staff hours. According to the economic theory, outsourcing rolling stock maintenance could reduce wages and increase productivity if the contracted labour is more productive. In the case of the CoMET and Nova metros, contracting out is relatively frequent, as shown in Table 8.2, where the average contracting out is roughly 25% (0.245) of the total of rolling stock maintenance staff hours. However, there is a large standard

variation of 21% (0.207). Furthermore, several metros do not outsource any of the maintenance staff hours as the minimum is 0%, whereas some metros have most of the rolling stock maintenance staff hours outsourced, with a maximum of 86% for a particular metro in a certain year of the data set.

In the case of the rolling stock maintenance costs, it is essential to include explanatory variables related to the characteristics and operational conditions of the rolling stock which affect the tear and wear of the asset. The first variable on the operational conditions of the rolling stock is the percentage of fleet available at the peak hour. This variable is defined as the total number of cars available for use during the busiest weekday peak hour, usually during the morning peak. The number of vehicles available sets a constraint on the maintenance routines and the timing of the routines that can be performed. For instance, a high percentage of vehicles available at the peak would probably entail more maintenance tasks performed at night, increasing rolling stock maintenance costs. In the case of the CoMET and Nova metros included in this research, the average fleet available at the peak is 88% (0.883), with a small standard deviation of just 7% as shown in Table 8.2.

Average commercial speed was also considered, measured as the average scheduled speed at which trains operate in passenger service. The speed is related to rolling stock maintenance costs in several ways. Firstly, higher speeds increase the production of car km per car and certain maintenance routines are practiced on a mileage basis (e.g. brake checks every half million car km). However, there are parts of the car that are not affected by the car km produced (e.g. car lights) so higher speeds and higher car km production

would lead to lower rolling maintenance costs per car km. Likewise, speed can also proxy for other variables, such as number of stops, as higher speeds are also related to longer distance between stations and this is also likely to entail less braking and door movements.

In the case of the data set of CoMET and Nova metros included in this research, the average speed is 33.14 km/h, with a maximum of 45 km/h and a minimum of 12 km/h for a particular year and metro of the data set. However, most of the metros circulate at an average speed of 30-35 km/h, as the median is 32.9 km/h and the standard deviation is relatively low at 5.25 km/h.

Another fundamental variable related to the rolling stock maintenance cost is the rolling stock reliability, measured as rolling stock mean distance between failures (MDBF). This variable refers to the incidents caused by rolling stock failure resulting in a delay to the train service of 5 minutes or more over the total actual operated revenue car kilometres. The relation between the MDBF and the rolling stock maintenance cost is unclear. Higher rolling stock incidents leading to more than a 5 minute delay in the service may be linked to faulty rolling stock assets that require more maintenance. However, unreliable and faulty rolling stock can also be caused by deficient rolling stock maintenance. Therefore, the direction of the effect is unclear. This can also be affected by other variables, such as the rolling stock age or the operational speed, which are controlled in this analysis. Consequently, this research can clarify the actual relation between reliability and maintenance costs, given that the estimates obtained already control for the other explanatory variables which are held constant.

In terms of mean distance between failures caused by rolling stock, Table 8.2 shows that the average CoMET and Nova metro runs 160 million car kilometres before having a car failure in any of their cars. However, although the median shown in Table 8.2 is close to the average (178.63 million car kilometres), the standard deviation represents almost half of the average (72.90 million car kilometres), implying large differences between the members in rolling stock reliability.

The percentage of rolling stock with air conditioning has also been considered in this analysis as more equipment on board may lead to higher rolling stock maintenance costs. Furthermore, cars with air conditioning also tend to have other equipment so, overall, it may also proxy for the degree of car complexity in the fleet. Table 8.2 shows that the CoMET and Nova metros have an average of 49.6% of rolling stock with air conditioning. However, the standard deviation is 45.89% and there are both metros with no air conditioning in any of the cars, minimum of 0, and fully air-conditioned fleets, maximum of 100% depending on the level of development and the local climate.

A key variable for understanding the rolling stock condition, which has a fundamental influence on the rolling stock maintenance costs, is the rolling stock age. Nevertheless, the effect of the age on the rolling stock maintenance cost may not be linear. Newer fleets may have higher maintenance costs in the beginning as they have new equipment and may demand new maintenance routines. However, the maintenance costs may drop significantly after the adaptation of the new fleet. Thereafter, rolling stock maintenance costs would tend to increase progressively until the end of the life cycle due to

ageing of the assets and obsolescence, which leads to higher maintenance costs over time.

In this analysis, the rolling stock age has been measured as the average age of the car units in commercial operation. This value is averaged over the year to take into account the retirement of old units and the addition of new rolling stock. In the case of the CoMET and Nova metros, the average rolling stock age is 20.33 years with a standard deviation of 10.93 years as shown in Table 8.2. This implies a mature but not old average fleet as the average rolling stock life is around 40 years, which can also be extended with any due mid-life refurbishments.

Lastly, the age of the network is also assessed, as there could be initial decisions taken when the metro was designed that could still influence or limit current practices (e.g. tunnel dimensions and rolling stock able to run on the network, given its design). Intuitively, older networks may tend to have higher rolling stock maintenance costs due to these constraints. In the case of this research, the network age is measured as the years since the start of the metro operations. Table 8.2 shows that the average network age of the metros included in the data set is 57 years, whereas the youngest member was only 10 years old and the oldest, in the latest year available, was 151 years old.

Finally, the analysis also considered a time trend to control for the evolution of the rolling stock maintenance costs once all the other explanatory variables are held constant. This variable can account for the general technical progress or general gains in productivity across the rail industry. In our case, the trend may not be representative of the overall trend in the rail

industry, as the members included in the panel actively try to improve their rolling stock maintenance performance by means of the CoMET and Nova groups.

Other explanatory variables considered

Other variables have been considered as potential determinants of rolling stock maintenance costs, although the decision was ultimately made to exclude them from this research. Firstly, the research considered other inputs besides labour, such as electricity and capital. However, electricity is not a main input in the maintenance tasks for rolling stock. The energy consumed by the workshops is negligible in comparison with the traction energy of train service costs or the stations' air conditioning for station service costs. Capital was also considered as an input, as cheaper capital may lead to more investment and more frequent renewal of assets. However, the measurement of capital is very controversial, particularly for state-owned companies such as metros which may have access to several types of financing. Furthermore, the effect of lower capital prices is indirect and, more importantly, lagged. Therefore, the benefit of lower capital costs would only begin to have an impact on maintenance costs some years after the investment. This would reduce the number of observations available and compromise the feasibility of the econometric analysis. For these reasons, capital was discarded from the rolling stock maintenance analysis.

Another factor considered was labour absenteeism, measured as the total staff hours not worked due to illness, injury, industrial action or lateness. However, due to the differences in labour laws between the countries com-

prised in the analysis, the data on absenteeism shows significant outliers and some inconsistencies. For that reason it was discarded from the analysis.

The maintenance routines and when they take place, day or night and weekends or only weekdays, were also considered. However, due to the lack of observations, their inclusion in the analysis would have substantially reduced the observations included in the analysis and would have compromised the statistical significance of the estimates.

Rolling stock characteristics such as the average car weight of the rolling stock was assessed, although it is clearly linked to other variables already included in the analysis, such as the rolling stock age and equipment on board such as air conditioning. Furthermore, the main difference in average car weights is between the US metros and the other metros, so the variable would also account for the effect of being a north American metro and not only the effect of a heavier fleet. Fleet homogeneity in general has also been considered as having the same type of fleet may lead to economies of scale. These economies of scale may come from cost savings in larger maintenance procurement but also by labour and machinery specialization. However, fleet homogeneity is hard to measure as metros may have a mix of different rolling stock types which, despite being different, may share or not a smaller or larger share of their maintenance routines. Therefore, two rolling stock types may be very similar in their maintenance routines despite being from different manufacturers and, on the contrary, the same rolling stock from different generations or models may be substantially different. Overall, it is hardly possible to measure accurately the actual level of rolling stock homogeneity of a rolling stock fleet.

Network characteristics were also examined. Firstly, the percentage of driverless operation in the network was considered, although it was discarded due to data quality. The variable is defined as the percentage of the network kilometres operated driverless, which is not in strict relation to the percentage of the cars serving in a particular line or the percentage of the car km produced by driverless cars. Furthermore, there is not enough variability in the data as to be significant, given that most of the metros have zero or very low percentages of driverless network length.

Likewise, another network characteristic considered was the average distance between stations, which could illustrate how frequent the stops and door openings are. However, this effect is already accounted for by the inclusion of the average commercial speed, which is proportional to the average distance between stations.

Human factors of rolling stock maintenance

Rolling stock maintenance is one of the most labour intensive divisions of the urban rail system. Besides the use of abundant labour, it is also one of the riskier areas in the metro operations. Consequently, health and safety procedures can be fundamental to explain the costs and the productivity of labour in rolling stock maintenance.

In the CoMET-Nova consortia incidents are reported as events causing a delay of 5 or more minutes in the service. Therefore, labour incidents in maintenance are not recorded in the CoMET-Nova framework as they usually do not affect directly the reliability of the service.

Other consortia of the Rail and Transport Strategy Centre (RTSC) do

consider injuries of the staff. For instance, the International Suburban Rail Benchmarking Group (ISBeRG) collects data on minor and major injuries suffered by the metro workforce. However, this data is controversial as the definitions of major and minor vary between different countries and any attempt to standardise them (i.e. considering major injury when requiring an overnight stay in hospital) also depends on the local health and safety customs (i.e. different healthcare systems would affect the likeliness of hospital overnight stays).

Additionally, exhaustive research in the ISBeRG group has also found that the number of injuries may also be influenced by the reporting procedures. Those metros encouraging injuries reporting may appear as more dangerous despite having the same level of actual injuries. Therefore, due to all this data constraints the health and safety dimension of rolling stock maintenance remains elusive for quantitative analysis and this key area of operations is better suited for qualitative research.

8.4 Results

This section presents the results for the econometric models on rolling stock maintenance costs per car kilometre and per car models as shown in Table 8.3. The models are estimated by means of random effects as the Breusch-Pagan test discarded the null hypothesis of constant variance over entities (Table 11.17 on page 304). Most of the explanatory variables have been transformed into natural logarithms ($[Ln]$) in order to estimate cost elasticities. The rolling stock age and the age of the network would not be intelligible as

cost elasticities so their coefficients describe the change per additional year. The coefficients shown in Table 8.3 are estimated average cost elasticities and correspond to the beta (β) estimates of Equation 8.2. The beta (β) estimates refer to how the average CoMET-Nova metro would behave when there is a marginal change in the explanatory variables described.

$$LogC_{Rstock} = \beta_0 + \sum_{j=1}^{J} \beta_j log x_j + \alpha_i + \varepsilon_{it} \tag{8.2}$$

The output variable car kilometres in Table 8.3 shows statistically significant cost elasticities of -0.49 for rolling stock maintenance per car kilometre. This suggests that an increase of 10% in car kilometres leads to a decrease of approximately 4.9% in maintenance costs per car kilometre. In this case, the null hypothesis is the nonexistence of economies of density in the rolling stock maintenance costs per car km, $H_0 : \beta_{carkm} = 0$. This would imply that the production of car kilometres does not lower the maintenance costs per car kilometre, so $\beta_{carkm} = 0$. This is already discarded in the first model in Table 8.3 which provides evidence in favour of existing economies of density in the rolling stock maintenance costs per car km. Thus, total costs would increase, but less than proportionally to the increase of car kilometres. This shows that a more intensive usage of the assets leads to economies in rolling stock maintenance costs.

The results on economies of scale are surprising and contradict the general view that rail operators mainly have economies of density but not economies of scale. In the case of the economies of scale, the model on rolling stock maintenance suggests that an increase in the fleet size, holding all other vari-

Table 8.3: RS Maintenance Cost Model Results - Random Effects

	RS Maintenance cost per	
	car km	fleet
	(1)	(2)
Car km [Ln]	−0.490***	
	(0.107)	
Fleet size [Ln]		−0.465***
		(0.108)
Wages [Ln]	0.577***	0.536***
	(0.084)	(0.082)
Staff hours [Ln]	0.516***	0.536***
	(0.091)	(0.085)
Staff outsourcing [Ln]	−0.723***	−0.722***
	(0.230)	(0.228)
Fleet availability [Ln]	0.315	0.793**
	(0.317)	(0.317)
Speed [Ln]	−0.004	0.076
	(0.133)	(0.131)
MDBF-RS [Ln]	−0.009	−0.007
	(0.023)	(0.022)
RS with air con.	0.195	0.168
	(0.124)	(0.121)
Fleet age	−0.014	−0.010
	(0.012)	(0.012)
Fleet age squared	0.0003	0.0003
	(0.0003)	(0.0002)
Network age (years)	0.001	−0.001
	(0.002)	(0.002)
Time trend	−0.014**	−0.008
	(0.006)	(0.006)
Constant	28.114**	15.392
	(12.152)	(12.227)
Rho	0.7495	0.7399
Observations	142	142
R^2	0.527	0.715
Note:	*p<0.1; **p<0.05; ***p<0.01	

ables constant, decreases the rolling stock maintenance costs per car with an estimated elasticity of -0.465. Therefore, the model hints that larger fleets may tend to have lower maintenance costs due to the presence of substantial economies of scale in the rolling stock maintenance per car. These economies of scale may be due to several factors such as labour specialisation or introduction of machinery for certain maintenance routines. Likewise, these results also suggest that breaking down the maintenance of rolling stock into different units to promote competition may incur, at least initially, in lower cost efficiency. This has significant policy implications, as this research suggests that it is not advisable to increase the productivity and cost efficiency by means of breaking down existing rolling stock maintenance departments within the same metro operator.

The econometric models also analyse the effect of labour explanatory variables, particularly wages and how labour-intensive the rolling stock maintenance is at a particular metro. Overall metro wages in Table 8.3 show a cost elasticity of 0.577 and 0.536 for the models on rolling stock maintenance cost per car km and per car respectively. The estimated elasticities for wages are significant at a 99% confidence level, indicating that metros with 10% higher wages are expected to have approximately 5.5% higher rolling stock maintenance costs, in line with previous research on wages cost elasticity (i.e. 0.47 wage cost elasticity in McGeehan (1993)). This coefficient implies that the wage level of the metro does not translate to proportional increases in the rolling stock maintenance costs. This is expected, given that labour is not the only input employed in rolling stock maintenance. Furthermore, there can be substantial substitution effects of labour for capital (Wang and Liao,

2006). Likewise, metros with higher wages may also have more productive labour, and consequently, less than proportional increases in costs.

In order to assess how a wage increase can lead to higher labour productivity, a model without labour hours is estimated (see Annex, Table 11.16 on page 303) with its own wage coefficient (i.e. $\beta_{flexible_L}$). The increase in labour productivity is calculated by the difference between the wage price elasticity for the model where the total labour hours are kept fixed when the wages increase (β_{fixed_L} in Table 8.3) and the coefficient for the model when the total labour hours are allowed to vary when wages increase ($\beta_{flexible_L}$ in Table 11.16). If higher wages lead to higher productivity, there would be no impact on cost efficiency, as long as higher wages lead to lower labour hours hired. The null hypothesis is that there is no difference between the two coefficients as shown in Equation 8.3, or equally, that the difference is statistically insignificant.

$$H_0 : \beta_{fixed_L} - \beta_{flexible_L} \simeq 0 \tag{8.3}$$

The wage cost elasticities for the model allowing changes in the staff hours (Table 11.16 on page 303) is equal to 0.650 and 0.581 for the rolling stock maintenance cost per car km and per car respectively. The difference between the coefficients 0.073 (0.577 and 0.650) and 0.045 (0.536 and 0.581) implies that the effect of wage increases in cost efficiency is similar, regardless of whether metros can adjust labour hours hired or not. The corresponding t-value of the difference for the cost per car km model (i.e. -0.72), as shown in Equation 8.4, is not significant at a 95% confidence level; that is, θ is smaller

than the t_{n-k} threshold (i.e. 1.656) for a t distribution with 129 degrees of freedom. Therefore, metros do not seem to adjust the labour hours hired when wages change. However, these results may be inaccurate as they are based on the overall wages and not the rolling stock maintenance wages.

$$\theta = \frac{\beta_{fixed_L} - \beta_{flexible_L}}{(\sigma_{fixed_L} + \sigma_{flexible_L})/2} = \frac{0.577 - 0.65}{(0.084 + 0.095)/2} \simeq -0.72 \qquad (8.4)$$

The econometric models in Table 8.3 also include variables that are under the control of the metro in the short to medium term such as the fleet availability at the peak. Fleet availability at the peak presents positive coefficients which are statistically significant at a 90% confidence level for the model on rolling stock maintenance cost per car. For example, the estimated elasticities imply that increasing the fleet availability at the peak from 85% to 90% (i.e. a $\approx 5.8\%$ increase) would increase the rolling stock maintenance costs per car km by 1.8% and 4.6% in rolling stock maintenance costs per car (e.g. $5.8 * 0.31 \approx 1.8$).

Besides deciding on the rolling stock available at the peak hours, the management of the metro can also decide on the level of outsourcing in the rolling stock maintenance. The models presented in Table 8.3 control for the level of staff hours contracted out in rolling stock maintenance. An average CoMET-Nova metro moving from a 25% contracting out level in rolling stock maintenance staff hours to a 35% contracting out level (i.e. a 40% increase from the initial level) could lower its rolling stock maintenance costs by approximately 3%. This could be due to the fact that contracting can bring specialised labour not existing in-house and it can also increase competi-

tion between own staff and outsourced labour, leading to lower maintenance costs. However, future research could investigate whether savings resulting from contracting out are concentrated in certain maintenance tasks and also which contracting out schemes are more cost-efficient.

This research considered the average commercial speed which is also linked to the frequency of rolling stock stops, as more distance between stations enables higher speeds. In this case, and once all the other explanatory variables have been controlled, the average commercial speed seems to have no effect on the cost efficiency of rolling stock maintenance per car km and per car. The coefficients are barely zero and this can be due to the fact that the car km variable may be accounting for part of the effect due to higher speeds.

Another factor with no impact on rolling stock cost efficiency is rolling stock reliability. In this latter case, the mean distance between failure (i.e. MDBF) caused by rolling stock is not related to the cost efficiency of rolling stock maintenance. The coefficients on MDBF are negative, suggesting that higher reliability leads to lower rolling stock maintenance costs due to more proactive maintenance, but they are small and statistically insignificant. It is probable that a larger data set with more observations would help to clarify the effect of reliability and rolling stock maintenance cost efficiency. For the time being, it hints that higher reliability leads to lower maintenance costs, but the results are inconclusive.

With regard to the condition of the assets, the models presented in Table 8.3 also tested the impact of rolling stock with air conditioning and the impact of rolling stock age. Firstly, rolling stock with air conditioning seems

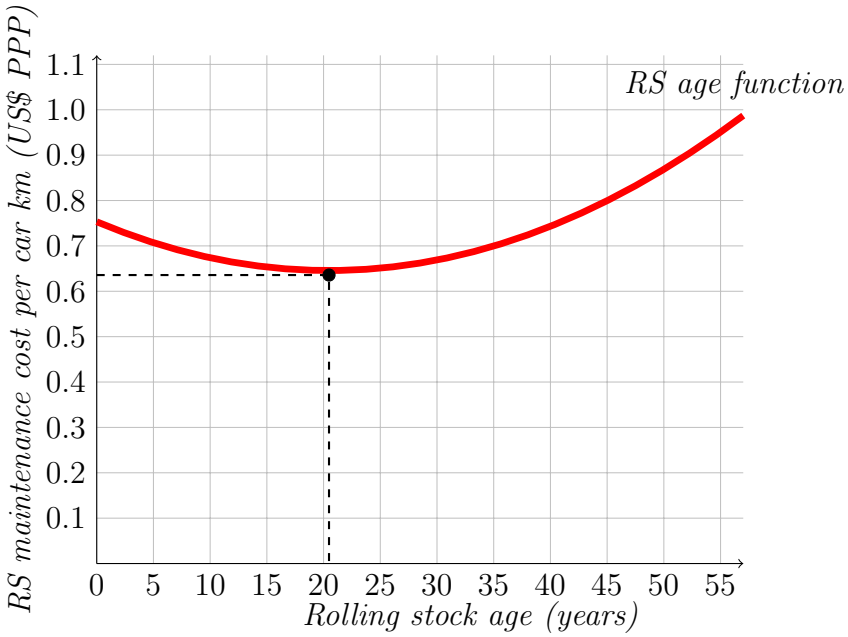
to increase the rolling stock maintenance costs, although these coefficients are not significant at a 90% confidence level. The models suggest that an increase in the percentage of rolling stock with air conditioning from 50% to 70% (i.e. a 40% increase from the initial level) could entail a rise of approximately 6.7% (i.e. $40 * 0.168 \approx 6.7$) in rolling stock maintenance costs per car, although the coefficient estimates are not statistically significant.

With respect to the rolling stock age and the effect of an additional year, the models in Table 8.3 tested how the effect evolves over time and whether the relation between rolling stock maintenance costs and rolling stock age is linear. In the models shown in Table 8.3, the linear rolling stock age variable equals -0.014 and -0.010 for the models on rolling stock maintenance per car km and per car respectively. However, the quadratic rolling stock age variable equals 0.00034 for both models as shown in Equation 8.5.

$$RSAgeFunction = RSMaint.Cost * (1 + (0.00034 * \text{year}^2 - 0.01393 * \text{year})) \quad (8.5)$$

Therefore, the age function shows a negative first degree variable with a positive quadratic variable, which would plot a convex function. This convex function is coherent with the theory explained in the data section (section 8.3) where the relation between rolling stock age and maintenance costs may not be linear. In particular, the maintenance costs could be higher at the beginning of the rolling stock life cycle, diminishing as the fleet matures and, later on, progressively increasing as the fleet becomes obsolete.

Figure 8.4: Average rolling stock maintenance cost per car km depending on age (in US$ PPP). All other cost determinants held constant.



The RS age function shown in Equation 8.5 is equal to the average rolling stock maintenance costs multiplied by the coefficients on rolling stock age. Figure 8.4 shows a representative fleet that has tended to decrease its maintenance costs since the introduction of the fleet until the twentieth year (i.e. 20.48). Afterwards, and with any other factors held equal, the costs would progressively increase. By the forty-first year, and holding all other factors constant, the rolling stock maintenance would exceed the initial maintenance costs. However, although the rolling stock age variables shown in Table 8.3 follow the expected theory on rolling stock maintenance costs (i.e. convex with a negative linear term and positive quadratic term as shown in Figure 8.4), these coefficients fail to be statistically significant at determining rail efficiency. This result may be due to the lack of observations, but it has also

been found in other research on rail efficiency (Sameni and Preston, 2012).

Besides the rolling stock age, the econometric models shown in Table 8.3 also assess the effect of the system age, measured in years since the beginning of operation. This is relevant, as certain characteristics of the metro (e.g. track gauge or tunnel dimensions among others) can influence the characteristics of the rolling stock. However, in the case of the two models presented in Table 8.3, it seems that the age of the metro network has no impact on the rolling stock maintenance costs once the other variables are held constant. This result was tested in the econometric models, with binary variables grouping metros into old, mature and new as shown in Table 11.12 on page 298. In this case, neither the binary variable for old metros, starting operations before 1945, nor the new ones, those inaugurated after 1980, show significant coefficients. Therefore, it seems that asset constraints set by the age of the metros do not affect rolling stock maintenance costs, probably due to the fact that rolling stock procurement already takes those constraints into account.

Finally, the model tested if there is a time trend in the rolling stock maintenance costs once all the other variables of the model are controlled. In this case, there seems to be a decreasing trend for the rolling stock maintenance costs per car km (-0.0142) and per car (-0.011). This implies that, discarding the effect of all the other variables included in the model, the metros in the CoMET and Nova consortia show a base line annual decrease of -1.1% to -1.4% in rolling stock maintenance costs per car and per car km respectively. This could be due to a general trend in productivity gains, the introduction of good practices shared in the CoMET and Nova consortia or

efficiency derived from general technological progress. In this case, it would be interesting to test whether this is a industry-wide trend or it is just the case of the CoMET-Nova metros.

The lower part of Table 8.3 reports general measures on the econometric model. Firstly, the *rho*-coefficient corresponds to the ratio between the standard deviation of the residuals within metros over the standard deviation of the residuals for the overall observations. In this case, the *rho*-coefficient is close to 0.75, which implies that around 75% of the variance in the model is due to the differences across metros. Therefore, most of the variation in the variables is due to the differences between the metros included in the sample and not the variation of these variables over time. Secondly, Table 8.3 also reports the goodness of fit of the two models by means of the R^2. The goodness of fit represents how well the variables included in the model explain the changes in the rolling stock maintenance costs. The most common measure of goodness of fit is the R-squared, which corresponds to the proportion of the total sample variation in the dependent variable (i.e. rolling stock maintenance costs per car km or per car) which is explained by the explanatory variables (Wooldridge, 2012). The R-squared in the Table 8.3 models range from the 0.527 for the rolling stock maintenance per car km model to the 0.715 for the model on rolling stock costs per car. This implies that the latter model explains a larger share of the variation of the data ($\approx 71\%$) while the first model explains around half of the data variation ($\approx 52\%$).

These values show that the econometric models have a moderate to high explanatory power. The remaining share of the rolling stock maintenance costs not explained in the model can be due to several factors. For in-

stance, the omission of explanatory variables due to data availability such as union power. Union power, regulatory framework or metro management competence, among many other variables, could play a role in rolling stock maintenance efficiency. However, data for variables such as union power may not be available and, for other variables such as regulatory framework, it may not be possible to measure it accurately. Likewise, more general factors such as work ethics, good practices or legal framework, among others, can be hardly included in the model, and consequently, there is always a share of the data variation that an econometric model cannot explain.

8.5 Conclusion

This chapter reviews the main efficiency determinants of the rolling stock maintenance cost per car kilometre and per car. The econometric models are based on a panel of 24 metros over a period of 10 years and are able to explain between 51% and 71% of the variation in rolling stock maintenance costs.

The research found significant returns to density and scale in the rolling stock maintenance costs, which could help to determine the optimal level of rolling stock utilisation to minimise maintenance. Likewise, the scale economies have been quantified which implies that breaking down fleet maintenance into different groups to spur on competition could, at least initially, lead to a decrease in efficiency. However, rolling stock staff hours contracted out seem to increase maintenance efficiency. Consequently, productivity gains could be obtained by contracting out staff rather than by outsourc-

ing entire maintenance tasks.

Other labour-related factors have been considered, such as wages and labour intensity. Wages seem to increase maintenance costs, but less than proportionally; thus, a 10% increase in wages is linked to a 5.7% increase in maintenance costs. This can be due to substitution effects between labour and capital, as well as labour productivity gains. The substitution effect has been tested with an additional model and there is no evidence that metros adjust labour hours when facing higher wages in rolling stock maintenance costs.

Operational variables have also been tested in the analysis. Average rolling stock speed and reliability are found to have a negligible impact on rolling stock maintenance cost efficiency. However, higher fleet availability at the peak is linked to significantly higher costs, probably due to the constraints on the timing and the maintenance routines (e.g. night maintenance works). With regard to the asset conditions, rolling stock with air conditioning tend to be more expensive, as air conditioning may be signalling more complex rolling stock and the coefficients for the variable fall short of the statistical significance threshold. The age of the assets has also been considered. The age of the overall network does not seem to impose a constraint in terms of cost efficiency, with almost zero coefficients for both models. Rolling stock age coefficients could corroborate the theory of higher maintenance costs for new rolling stock, with lower costs for mature assets and increasingly higher for those cars becoming obsolete. In this case, the rolling stock age coefficients are only significant at a 80% confidence level, so future research with more observations could clarify the effect of age on maintenance.

Lastly, the rolling stock maintenance costs show a decreasing trend in costs over time. This could be due to common industry-wide gains in productivity derived from technology or a particular behaviour in the CoMET-Nova members due to the introduction of good practices. In this case, further research would be needed to see if it is an industry-wide trend or just a common trend for the CoMET-Nova members.

9 — Conclusions

This chapter summarises the main contributions of this thesis. Firstly, section 9.1 reviews the main conclusions of the research, both for the general literature on rail cost efficiency but also for particular categories of operating costs. Secondly, Section 9.2 describes the main policy implications of improving urban rail cost efficiency, based on the contributions of this research. Lastly, Section 9.3 describes the limitations of this thesis, while section 9.4 suggests future research that could continue and complement this research.

9.1 Main Findings and Contributions to the Existing Literature

9.1.1 General Findings and Contributions

One of the research contributions outlined in section 1.5 is to understand to what extent economies of traffic density are relevant for metros. Returns to traffic density, a more intensive utilisation of the assets to increase output, have been found to be substantial at the aggregate operating costs. The effect seems to be significant both at the overall aggregate level of the op-

erating costs and also at the different constituents of operating costs. In particular, train and station service costs present large economies of density, as the marginal costs of additional passengers or car kilometres tend to be negligible. On the contrary, rolling stock maintenance shows moderate, although statistically significant, economies of traffic density. Overall, the study corroborates previous research on economies of density in the rail sector, although it also points out the significant differences between the different operational divisions of the metros. The academic research suggested the existence of economies of density. This research contributes to confirm the previous research, but it also points out the variation between different areas of railway operations. These differences between operational divisions is a contribution that has seldom been observed in the literature review.

These findings suggest that, most of the time, metros operate at a level significantly below their capacity, which is defined as the maximum capacity given its asset endowment. This is most likely due to the concentration of demand in acute peak periods. Metros must cope with the high demand levels during commuting times, although it entails an asset investment that will remain underused the rest of the day. This overcapacity appears in the analysis as the significant economies of density of a fixed infrastructure that is underused.

Nevertheless, this does not imply a social inefficiency per se. The metro's positive externalities, such as reduced congestion or increased land values of surrounding areas, can still offset the cost of underused investment. Likewise, although the economies of density suggest substantial spare capacity, the rigidity of commuting times and passenger demand restrict the ability of the

metro to match capacity to demand throughout the day. Likewise, although the marginal cost of additional passengers may be close to 0 for the metro, these additional passengers do represent a cost for the existing passengers in terms of crowding and its subsequent discomfort.

With regard to economies of scale, aggregate operating costs and service costs dimensions, such as train service and station service, show inconclusive and statistically insignificant results following the trail of academic research initiated by Keeler (1974). However, there is statistically significant evidence that larger fleets lead to lower rolling stock maintenance costs per car. Thus, there is evidence that metros, and rail operators in general, can benefit from economies of scale in areas of their operations related to maintenance. This may be due to the fact that larger fleets may lead to lower supply unit costs of spare parts and enable further labour specialisation. Additionally, larger fleets can also increase the opportunities for profitable automatisation of maintenance tasks, substituting labour for capital and leading to further efficiency gains.

These means of obtaining higher efficiency are usually not available in service operations. This implies that, as most of the academic research is based on aggregate operating costs which are mostly service costs, the negligible effect of scale on service cost efficiency tends to conceal the potential economies of scale in the maintenance divisions. Thus, this research provides evidence that aggregate cost research may have failed to note the substantial variation in economies of density and scale of the different rail operation divisions, and in particular, the differences between rail service and maintenance divisions in urban rail systems.

Labour-related cost factors have been shown to be key for cost efficiency as all models show highly significant wage coefficients. Overall, a 10% increase in wages tends to increase costs by between 2% and 6% depending on the cost category. However, the econometric models have also proven that this effect of wages can be corrected, at least partially, by means of staff outsourcing. Staff outsourcing increases cost efficiency in all cost categories, but the effect is particularly significant for train service and rolling stock maintenance divisions. This can be due to the increased competition between own and contracted labour, which makes a match between labour wages and labour productivity more likely. Likewise, outsourced labour can also bring new skills and know-how into the metro, increasing the overall productivity of the metro staff. On the contrary, the effect of contracting out in station service costs is more moderate, probably because metros facing expensive labour tend to already to have non-staffed stations and less station service labour, rather than outsourcing station service staff.

Likewise, higher wages may not increase costs if they also lead to higher labour productivity. The link between wages and productivity has been assessed by means of comparing the wage cost elasticities for cost functions with and without total labour hours hired. Those cost categories where the wage coefficients are similar for the two models, with and without controlling for total labour hours, show no productivity gains when wages increase. On the contrary, there is evidence that higher wages lead to lower increases in costs when the metros can adjust labour hours. The increase in labour productivity when wages increase has been found to be particularly significant for the overall aggregate operating costs. However, the evidence is weaker

for the subcategories of metro operating costs, probably due to the fact that wages for those categories were proxied with the overall wage rate and not the particular wage rate for the cost subcategories.

Other inputs like electricity prices have a statistically significant effect on increasing metro cost efficiency, particularly for those models that better explain the cost variations across metros and have a higher explanatory power, such as the aggregate operating cost and the train service cost models. The analysis of the link between energy prices and energy efficiency, as in the case of the wages and productivity, was not possible due to the exogenous interaction in energy efficiency with other entities (e.g. public authorities and regulators) and the lack of disaggregated data on energy consumption. The age of the urban rail system presented mixed results, showing that the effect of age is not homogenous across all urban rail divisions. Aggregate operating costs, and particularly station service costs, show that older metros tend to have higher costs, which is probably due to operational constraints set by the initial design of the system (e.g. old stations may not meet current safety standards unless they are heavily staffed). However, the cost efficiency of train service and rolling stock maintenance is not affected by the age of the metro. Therefore, train operations can probably adapt to the initial design without increasing operating costs (e.g. track curvature or tunnel dimensions). On the contrary, the efficiency of operating costs related to fixed assets, such as stations, are determined more by the initial design constraints. This age effect has also been tested by metro generations, breaking down the metros into old metros (i.e. inaugurated before 1945), mature metros (i.e. 1945-1988) and new metros. In the analysis, both old and new metros tend

to show higher costs, although these are only statistically significant for the aggregate operating and the station service costs. Therefore, based on this research, mature metros could anticipate that cost efficiency will deteriorate as assets become more obsolete.

Likewise, looking at the general time trend of the CoMET-Nova metros, there is an upward trend in costs for the aggregate costs and inconclusive results for the service divisions (i.e. train service and station service). However, particular divisions of the urban rail operations such as rolling stock maintenance show statistically significant downward time trends in costs. Thus, particular divisions of the urban rail such as rolling stock maintenance may be lowering costs, although the time trend for the overall aggregate cost may be increasing. This is plausible given that technological progress and productivity trends are not homogeneous across all the metro divisions. Thus, at least during the period 2005-2014 studied, the rolling stock maintenance showed a statistically significant increase in cost efficiency, all other cost factors held constant, despite the opposite trend in the aggregate operating costs. Consequently, this divergence confirms again that the cost factor effects are not homogeneous across the urban rail system, which reinforces one of the main contributions of this research, that is, to separately analyse the different operational divisions of the urban rail systems.

9.1.2 Cost-Specific Findings and Contributions

Although there is a set of cost factors tested in all models to enable comparisons between different operating divisions, the analysis also included partic-

ular cost factors that were not applied to all cost categories. In the aggregate operating costs category, subsidies are the most statistically significant cost factor. Every additional US$ of subsidies is associated to an increase of 0.5 US$ in aggregate operating costs. Thus, metros with higher subsidies tend to also be the ones with lower cost efficiency. A key question that requires a longer panel would be to analyse whether there is a causal relation between subsidies and lower cost efficiency.

Other variables, such as incidents per car kilometre, showed that more reliable metros tend to show higher cost efficiency. As cost efficiency is measured as output unit cost (e.g. train service costs per car km), more incidents would decrease output and consequently raise operating costs per output. Moreover, it is clear that incident delays are not the only cost of incidents, but rather than metros with higher incident rates also tend to have lower cost efficiency. On the contrary, other variables like the opening hours per week do not seem to impose an additional burden on the operating costs of the metro.

With regard to variables for the train service models, the most significant variable is staff productivity. Outsourcing in train service is rare due to the critical responsibility of the tasks carried out by the train service staff (i.e. mainly train drivers). However, labour productivity still determines train service cost efficiency. Driver productivity cost efficiency gains are not as significant as outsourcing; however, they can be obtained by means of better train movement planning, swift train turnovers and improved roster scheduling. Therefore, a substantial source of efficiency gains does not entail outsourcing labour or additional investment but rather better metro man-

agement and labour flexibility.

With regard to the characteristics of the rolling stock, these were not statistically significant at explaining the metro train service cost efficiency. However, this may be due to the lack of over time variation in the rolling stock characteristics, which is discussed further in the limitations of this study.

The variables on the characteristics of the station, such as the platform length proxied by the train length and the number of lifts, escalators and travelators, were found to be statistically significant. Overall, longer platforms and more station equipment (i.e. lifts and escalators) tends to increase costs as expected, due to increased station complexity and higher electricity demand.

Lastly, particular variables for the rolling stock maintenance costs also showed that variables like fleet availability at the peak have a significant effect on the cost efficiency of the metros. This is probably due to the constraint on maintenance routines when higher percentages of the fleet need to be available at the peak. However, and as mentioned above, other fleet rolling stock characteristics such as fleet age or rolling stock reliability were not significant. It could be either due to the fact that they are not significant for maintenance or that more variability in the data is needed, so the results on these two variables remain inconclusive.

9.2 Policy Implications

The importance of the research contributions presented in section 1.5 is determined by how the results can improve the actual cost efficiency of the

urban rail systems.

A key result refers to the estimation of the economies of density and size which can guide which operations of the urban rail systems are presumably more efficient at operating in one unit and which divisions can be franchised. Given that the economies of density are more important than the economies of size, this implies that small and medium sized metros can be as cost-efficient as large metros at producing output such as car kilometres or passenger journeys. Therefore, the results can suggest that regulators could break down the rail operations in order to introduce competition (e.g. operating franchises), as competition may foster improvements without causing diseconomies of scale. However, this research shows that not all the metro operating divisions are homogeneous in terms of returns to scale and density. There are some operating divisions where returns to scale are significant. For instance, rolling stock maintenance presents significant economies of scale, so breaking down the maintenance operations would probably lead to higher costs due to diseconomies of scale. This can be due to the fact that maintenance tasks can be more favorable for the use of specialised labour and machinery as explained above in section 9.1.1. Overall, the research casts light on how returns to density and returns to scale can vary within the same organisation and it also serves as a reference on how to foster competition without compromising existing efficiency gains.

The research has also found that age can have an impact on operating costs, although the impact is not linear. Mature metros, systems between 30 and 60 years old, tend to be at the highest level of cost efficiency, shortly before obsolescence and asset ageing increase the operating costs. This should

underpin the importance of continuous investment to avoid a sharp increase in costs during the transition from mature to old urban rail systems. Likewise, new metros tend to have higher costs than mature metros. Therefore, this research shows that new metros can be ambitious in their cost target-setting as cost efficiency tends to increase when metros become more mature, probably due to an increase in operative know-how and acquired expertise.

Time trends have also been found to be significant and positive in several models. For the overall operating costs, costs tend to increase over time across the CoMET and Nova members. Consequently, higher costs for a similar output could be expected in the future, despite keeping all other cost factors constant. This may be due to the ever-growing passenger expectations and the need to continuously improve service performance and passenger convenience. However, as with the case of the economies of scale and density, the increasing time trend is not homogeneous across different operation divisions. Aggregate operating costs show increasing costs but rolling stock maintenance shows a negative trend, probably related to the technological progress and automatisation of new rolling stock trains and signalling systems. Thus, rail operators could improve the accuracy of their target-setting by means of considering the corresponding time trend for each of the operational divisions, which in some cases can be divergent time trends, as shown in this research.

A key cost factor for aggregate operating costs is the level of subsidy, defined as the share of operating costs not funded by the operating commercial revenue. This research has found that, on average for the CoMET-Nova metros, every 1% of operating costs subsidised is associated to a 0.5% increase

in operating costs. Thus, although subsidies may be a potential solution for short-term financial deficits, this dissertation shows that short-term subsidies may actually be associated to a larger long-term problem of financial sustainability of the metro. However, this research has only found an association and not a causal relation as there was not enough data available for a causal analysis.

Short-run financial deficits can be caused by the decoupling of labour wages from labour productivity. This research demonstrates that wage increases are generally not followed by proportional increases in labour productivity. Therefore, most of the metros with higher wage levels tend to have higher costs, although also lower labour hours. This trend can be partially corrected by the outsourcing of labour hours, which has been found to improve cost efficiency for almost all the econometric models of this research. However, not all operational divisions seem to benefit equally from outsourcing. The most significant effects of labour outsourcing have been founds in the train service and the rolling stock maintenance operations. However, labour productivity can improve without outsourcing labour as in the case of train service, where higher driver productivity has been found to be a key factor to improve metro cost efficiency. Driver productivity can increase by means of labour flexibility and better labour hour scheduling. These good practices, among others, could also be transferred to other metro operations.

Lastly, other operational characteristics such as the percentage of fleet available at the peak can lower the cost efficiency of particular cost categories like rolling stock maintenance costs. The quantification of the average effect for the CoMET-Nova metros has been estimated in this research, so the urban

rail operators can use these estimates to consider whether the additional rolling stock units available at the peak can offset the increase in rolling stock maintenance costs.

9.3 Limitations of the Study

The main limitations of the research refer to data availability. For instance, this research considered the aggregate operating costs and all the operating cost constituents in the CoMET-Nova framework. However, some of them were not included:

Another limitation of the study is the lack of certain cost factors that were not able to be included in the analysis due to data unavailability. The main cost factors missing are as follows:

- Urban density is fundamental in optimising cost efficiency, as passenger journeys depend on the urban density surrounding the stations. However, there is no accurate measurement of urban density for all the cities included in the analysis. Furthermore, the key variable is not the metropolitan urban density but the actual density around the stations operated by the metro.

- Rolling stock characteristics would probably have an impact on most of the operating costs constituents. However, the data on rolling stock characteristics is scarce, less accurate and presents a low variance over time which compromises the feasibility of the econometric analysis.

- The time concentration of the passengers can restrict the optimisation

of costs. The demand profile data was considered, measured as the average number of passengers travelling in the network per hour. However, it does not specify where in the network and in what direction the passengers are travelling. Thus, although the demand profile is likely to be a key cost factor, it was never statistically significant and it was eventually discarded from the final cost models.

- Investment is probably linked to better metro performance, although the effect of investment may be lagged in time. However, including investment as a lagged variable would have substantially reduced the number of observations and compromised the statistical significance of the models.

Overall, the econometric analysis shows a substantial explanatory power concerning the cost efficiency of urban rail systems, as it is able to explain around half of the cost variation from the mean. A larger and longer panel would probably improve the statistical significance of some cost factors, although not necessarily the goodness of fit of the models. The other half of the operating costs variation is probably determined by the metro's best practices, organisational characteristics or working culture, among others factors. This cannot be easily quantified or included in the econometric analysis. Therefore, although the econometric analysis explains a large share of the operating cost variation, future research and non-econometric methods are required to explain the other half of the variation.

9.4 Directions for Future Research

Future research on urban rail cost efficiency can be based on both the findings of this research shown in section 9.1 and also on tackling the research limitations explained in section 9.3. In particular, future research could address the following topics:

- Replicating the cost efficiency analysis with urban rail operators outside the CoMET-Nova framework in order to assess to what extend the cost factor coefficient estimates are consistent. This would also be especially interesting for time trends as a way to analyse to what extent the CoMET-Nova consortia is improving the performance of the members. Moreover, given that the time span considered in this research is limited, time trends in the future could evolve in a different way to the estimates provided in this dissertation.

- Line level analysis has been discarded due to data availability, although it could facilitate the assessment of the effect of other cost factors. For instance, the effect of driverless lines on operating costs can be blurred at the network level, but it could be effectively assessed at the line level. Likewise, line level variables on signalling or rolling stock characteristics could introduce more variability and lead to higher statistical significance of the coefficient estimates. Additionally, variables such as age could be more relevant at the line level as different technologies may coexist in the same urban rail network at the same time (e.g. Northern line and Jubilee line in the London Underground).

- Reliability has been addressed at an aggregate level for incidents causing more than 5 minutes delay. New research could break down the effect of incidents depending on their duration and the cause of the incident. This breakdown could also distinguish incidents by the length of the interruption caused (e.g. 3, 5 and 10 minute delays).

- Productivity of labour has been assessed, but the analysis used average overall wages rather than the wages for each of the constituents of operating costs. Therefore, adding more years, more rail operators and also less aggregate wage data would certainly improve the assessment of the relation between labour wages and labour productivity.

- Outsourcing has been found to be a key cost factor improving cost efficiency. The results are strong and statistically significant for almost all operating cost constituents. However, it would be interesting to understand which particular tasks and divisions offer the biggest gains in efficiency when they are outsourced if more disaggregated data were available.

- Although some network characteristics have been assessed, and urban density has been also considered, it would be interesting to see how other metropolitan transport modes affect metro performance and cost efficiency. A large share of the success in a urban rail system depends on how the overall public transport system is designed (e.g. bus routes feeding metro stations with more passenger demand) and how other modes complement the rail operations.

- Although smart card data is not widely available, as more metros introduce smart cards, it might be possible to characterise passenger behaviour and assess how passenger behaviour affects urban rail cost efficiency.

10 — Bibliography

Ackerberg, D., Caves, K., Frazer, G. et al. (2006), 'Structural identification of production functions', *MPRA paper* **38349**.

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

Andersson, M. (2008), 'Marginal railway infrastructure costs in a dynamic context', *European Journal of Transport and Infrastructure Research* **8**(4), 268–286.

Andersson, M. (2011), 'Marginal cost of railway infrastructure wear and tear for freight and passenger trains in sweden', *European Transport Trasporti Europei* **48**, 3–23.

Arellano, M. and Bond, S. (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *The Review of Economic Studies* **58**(2), 277–297.

Arellano, M. and Bover, O. (1995), 'Another look at the instrumental vari-

able estimation of error-components models', *Journal of econometrics* **68**(1), 29–51.

Armstrong, J. and Preston, J. (2011), 'Alternative railway futures: growth and/or specialisation?', *Journal of Transport Geography* **19**(6), 1570–1579.

Babalik-Sutcliffe, E. (2002), 'Urban rail systems: analysis of the factors behind success', *Transport Reviews* **22**(4), 415–447.

Baltagi, B. (2005), *Econometric analysis of panel data*, Chippenham, Wiltshire.

Barbot, C., Costa, Á. and Sochirca, E. (2008), 'Airlines performance in the new market context: A comparative productivity and efficiency analysis', *Journal of Air Transport Management* **14**(5), 270–274.

Barke, D. and Chiu, W. (2005), 'Structural health monitoring in the railway industry: a review', *Structural Health Monitoring* **4**(1), 81–93.

Barros, C. P. (2008), 'Airports in argentina: Technical efficiency in the context of an economic crisis', *Journal of Air Transport Management* **14**(6), 315–319.

BBC (2007), 'Rail 'safer' after privatisation'.

Berger, A. N. and Mester, L. J. (1997), 'Inside the black box: What explains differences in the efficiencies of financial institutions?', *Journal of Banking & Finance* **21**(7), 895–947.

Bladon, K., Rennison, D., Izbinsky, G., Tracy, R. and Bladon, T. (2004), 'Predictive condition monitoring of railway rolling stock'.

Blundell, R. and Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of econometrics* **87**(1), 115–143.

Blundell, R. and Bond, S. (2000), 'Gmm estimation with persistent panel data: an application to production functions', *Econometric reviews* **19**(3), 321–340.

Bly, P., Webster, F. and Pounds, S. (1980), 'Effects of subsidies on urban public transport', *Transportation* **9**(4), 311–331.

Bogetoft, P. and Otto, L. (2010), *Benchmarking with DEA, SFA, and R*, Vol. 157, Springer Science & Business Media.

Braeutigam, R. R., Daughety, A. F. and Turnquist, M. A. (1982), 'The estimation of a hybrid cost function for a railroad firm', *The Review of Economics and Statistics* **64**(3), 394–404.

Brage-Ardao, R., Graham, D. J. and Anderson, R. J. (2015), 'Determinants of train service costs in metro operations', *Transportation Research Record: Journal of the Transportation Research Board* (2534), 31–37.

Breithaupt, M. (2010), 'Latest developments in BRT from around the world', *World Urban Transport Leaders Summit* .

Breusch, T. S. and Pagan, A. R. (1980), 'The lagrange multiplier test and its applications to model specification in econometrics', *The Review of Economic Studies* **47**(1), 239–253.

Brown, R. S., Caves, D. W. and Christensen, L. R. (1979), 'Modelling the structure of cost and production for multiproduct firms', *Southern Economic Journal* **46**(1), 256–273.

Canavan, S., Graham, D. J., Melo, P. C., Anderson, R. J., Barron, A. and Cohen, J. M. (2015), The impacts of moving block signalling on technical efficiency: An application of propensity score matching on urban metro rail systems, *in* 'Transportation Research Board 94th Annual Meeting', number 15-2694.

Cantos, P. (2001), 'Vertical relationships for the european railway industry', *Transport Policy* **8**(2), 77–83.

Cantos, P. and Maudos, J. (2000), 'Efficiency, technical change and productivity in the european rail sector: A stochastic frontier approach', *International Journal of Transport Economics* **27**(1), 55–76.

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. (2002), 'Cost and revenue inefficiencies in the european railways', *International Journal of Transport Economics* **29**, 279–308.

Caves, D. W., Christensen, L. R. and Swanson, J. A. (1980), 'Productivity in US railroads, 1951-1974', *The Bell Journal of Economics* pp. 166–181.

Caves, D. W., Christensen, L. R. and Swanson, J. A. (1981*a*), 'Economic performance in regulated and unregulated environments: a comparison of us and canadian railroads', *The Quarterly Journal of Economics* pp. 560–581.

Caves, D. W., Christensen, L. R. and Swanson, J. A. (1981*b*), 'Productivity growth, scale economies, and capacity utilization in us railroads, 1955-74', *The American Economic Review* pp. 994–1002.

Caves, D. W., Christensen, L. R. and Tretheway, M. W. (1980), 'Flexible cost functions for multiproduct firms', *The Review of Economics and Statistics* pp. 477–481.

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, *in* A. F. Daughety, ed., 'Analytical Studies in Transport Economics', Cambridge University Press, pp. 97–120.

Charnes, A. (1994), *Data Envelopment Analysis: Theory, Methodology, and Applications*, Springer Science & Business Media.

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.

Charnes, A., Cooper, W. W. and Rhodes, E. (1979), 'Measuring the efficiency of decision-making units', *European journal of operational research* **3**(4), 339.

Cheng, Y. and Tsao, H. (2000), 'Rolling stock maintenance strategy selection, spares parts' estimation, and replacements' interval calculation', *International Journal of Production Economics* **128**(1), 404–412.

CIESIN, I. (2010), 'WRI, 2010. Gridded Population of the World (GPW), version 4', *Center for International Earth Science Information Network (CIESIN) Columbia University, International Food Policy Research Institute (IFPRI) and World Resources Institute (WRI), Palisades, NY* .

Clark, X., Dollar, D. and Micco, A. (2004), 'Port efficiency, maritime transport costs, and bilateral trade', *Journal of development economics* **75**(2), 417–450.

Cobb, C. W. and Douglas, P. H. (1928), 'A theory of production', *The American Economic Review* pp. 139–165.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J. and Battese, G. E. (2005), *An introduction to efficiency and productivity analysis*, Springer Science & Business Media.

Coelli, T. and Perelman, S. (1996), 'Efficiency measurement, multiple-output technology and distance functions: With application to european railways'.

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.

Colombi, R., Kumbhakar, S. C., Martini, G. and Vittadini, G. (2014),

'Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency', *Journal of Productivity Analysis* **42**(2), 123–136.

Cooper, W. W., Seiford, L. M. and Tone, K. (2007), *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, Springer Science & Business Media.

Couto, A. and Graham, D. J. (2008), 'The contributions of technical and allocative efficiency to the economic performance of european railways', *Portuguese Economic Journal* **7**(2), 125–153.

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* pp. 241–251.

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

Cullinane, K., Wang, T.-F., Song, D.-W. and Ji, P. (2006), 'The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis', *Transportation Research Part A: Policy and Practice* **40**(4), 354–374.

De Borger, B. (1992*a*), 'Estimating a multiple-output generalized box-cox cost function: Cost structure and productivity growth in belgian railroad operations, 1950–1986', *European Economic Review* **36**(7), 1379–1398.

De Borger, B. (1992*b*), 'Estimating a multiple-output generalized box-cox

cost function: Cost structure and productivity growth in belgian railroad operations, 1950–1986', *European Economic Review* **36**(7), 1379–1398.

De Palma, A., Lindsey, R., Quinet, E. and Vickerman, R. (2011), *A handbook of transport economics*, Edward Elgar Publishing.

Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica: Journal of the Econometric Society* pp. 273–292.

Denny, M., Fuss, M., Everson, C. and Waverman, L. (1981), 'Estimating the effects of diffusion of technological innovations in telecommunications: The production structure of bell canada', *Canadian Journal of Economics* pp. 24–43.

Deprins, D. and Simar, L. (1989), 'Estimating technical inefficiencies with correction for environmental conditions', *Annals of Public and Cooperative Economics* **60**(1), 81–102.

Diewert, W. E. (1971), 'An application of the shephard duality theorem: A generalized leontief production function', *The Journal of Political Economy* pp. 481–507.

Farrell, M. J. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society. Series A (General)* pp. 253–290.

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.

Farsi, M., Filippini, M. and Kuenzle, M. (2006), 'Cost efficiency in regional bus companies: An application of alternative stochastic frontier models', *Journal of Transport Economics and Policy* pp. 95–118.

Filippini, M. and Maggi, R. (1992), 'The cost structure of swiss private railways', *International Journal of Transport Economics* **19**(3), 307–327.

Filippini, M. and Maggi, R. (1993), 'Efficiency and regulation in the case of the swiss private railways', *Journal of Regulatory Economics* **5**(2), 199–216.

Friedlaender, A. F., Berndt, E. R., Chiang, J. S. W., Showalter, M. H. and Vellturo, C. A. (1993), 'Rail costs and capital adjustments in a quasi-regulated environment', *Journal of Transport Economics and Policy* **27**(2), 131–152.

Gathon, H.-J. and Perelman, S. (1992), Measuring technical efficiency in european railways: a panel data approach, *in* 'International Applications of Productivity and Efficiency Analysis', Springer, pp. 131–147.

Gathon, H.-J. and Pestieau, P. (1995), 'Decomposing efficiency into its managerial and its regulatory components: The case of european railways', *European Journal of Operational Research* **80**(3), 500–507.

Graham, D., Couto, A., Adeney, W. and Glaister, R. (2003*a*), 'Economies of scale and density in urban rail transport: effects on productivity', *Transportation Research Part E* **39**, 443–458.

Graham, D., Crotte, A. and Anderson, A. (2009), 'A dynamic panel analysis

of urban metro demand', *Transportation Research Part E: Logistics and Transportation Review* **45**(5), 787–794.

Graham, D. J. (2008), 'Productivity and efficiency in urban railways: Parametric and non-parametric estimates', *Transportation Research Part E: Logistics and Transportation Review* **44**(1), 84–99.

Graham, D. J., Couto, A., Adeney, W. E. and Glaister, S. (2003*b*), 'Economies of scale and density in urban rail transport: effects on productivity', *Transportation Research Part E: Logistics and Transportation Review* **39**(6), 443–458.

Graham, D. J., Couto, A., Adeney, W. E. and Glaister, S. (2003*c*), 'Economies of scale and density in urban rail transport: effects on productivity', *Transportation Research Part E: Logistics and Transportation Review* **39**(6), 443–458.

Greene, W. (2002), *Econometric Analysis*, New Jersey.

Greene, W. (2004), 'Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the world health organization's panel data on national health care systems', *Health economics* **13**(10), 959–980.

Greene, W. H. (2008), *The Econometric Approach to Efficiency Analysis*, Oxford University Press, chapter 2, pp. 92–250.

Griliches, Z. and Mairesse, J. (1995), Production functions: the search for identification, Technical report, National Bureau of Economic Research.

Harmatuck, D. (1979), 'A policy sensitive railroad cost function', *The Logistics and Transportation Review* **15**, 277–315.

Hausman, J. A. (1978), 'Specification tests in econometrics', *Econometrica: Journal of the Econometric Society* pp. 1251–1271.

Hayashi, F. (2000), 'Econometrics. 2000', *Princeton University Press. Section* **1**, 60–69.

Hensher, D. A., Daniels, R. and Demellow, I. (1995), 'A comparative assessment of the productivity of australia's public rail systems 1971/72–1991/92', *Journal of Productivity Analysis* **6**(3), 201–223.

Hensher, D. A. and Golob, T. F. (2008), 'Bus rapid transit systems: a comparative assessment', *Transportation* **35**(4), 501–518.

Holmgren, J. (2013), 'The efficiency of public transport operations–an evaluation using stochastic frontier analysis', *Research in Transportation Economics* **39**(1), 50–57.

Iseki, H. (2010), 'Effects of contracting on cost efficiency in us fixed-route bus transit service', *Transportation Research Part A: Policy and Practice* **44**(7), 457–472.

Jain, P., Cullinane, S. and Cullinane, K. (2008), 'The impact of governance development models on urban rail efficiency', *Transportation Research Part A: Policy and Practice* **42**(9), 1238–1250.

Jan, A. and Phillips, D. (2011), Rail value for money study. Rolling stock whole life costs, Technical report, ARUP.

Jowit, J., Milmo, D. and Watt, N. (2012), 'Rail network to see "biggest investment" since the victorians'.

Karlaftis, M. G. and Tsamboulas, D. (2012), 'Efficiency measurement in public transport: Are findings specification sensitive?', *Transportation Research Part A: Policy and Practice* **46**(2), 392–402.

Keeler, T. E. (1974), 'Railroad costs, returns to scale, and excess capacity', *The Review of Economics and Statistics* **56**(2), 201–208.

Kim, H. Y. (1987), 'Economics of scale and scope in multiproduct firms: evidence from us railroads', *Applied Economics* **19**(6), 733–741.

Koopmans, T. C., Alchian, A. A., Dantzig, G. B., Georgescu-Roegen, N., Samuelson, P. and Tucker, A. (1951), *Activity analysis of production and allocation: proceedings of a conference*, Wiley New York, NY.

Kumar, S., Espling, U. and Kumar, U. (2008), 'A holistic procedure for rail maintenance in sweden', *Proc. ImechE - PartF: Journal of Rail and Rapid Transit* **222**, 331.

Kumbhakar, S. C. (1988), 'Estimation of input-specific technical and allocative inefficiency in stochastic frontier models', *Oxford economic papers* pp. 535–549.

Kumbhakar, S. C., Lien, G. and Hardaker, J. B. (2014), 'Technical efficiency in competing panel data models: a study of norwegian grain farming', *Journal of Productivity Analysis* **41**(2), 321–337.

Kumbhakar, S. C. and Lovell, C. K. (2003), *Stochastic frontier analysis*, Cambridge University Press.

Laurin, C. and Brozec, Y. (2001), 'Privatization and productivity improvement: the case of canadian national', *Transportation Research Part E: Logistics and Transportation Review* **37**(5), 355 – 374.

Levinsohn, J. and Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *The Review of Economic Studies* **70**(2), 317–341.

Liebert, V. and Niemeier, H.-M. (2013), 'A survey of empirical research on the productivity and efficiency measurement of airports', *Journal of Transport Economics and Policy (JTEP)* **47**(2), 157–189.

Lovell, C. K. and Sickles, R. C. (1983), 'Testing efficiency hypotheses in joint production: a parametric approach', *The Review of Economics and Statistics* pp. 51–58.

Mackett, R. and Sutcliffe, E. B. (2003), 'New urban rail systems: a policy-based technique to make them more successful', *Journal of Transport Geography* **11**(2), 151–164.

Mancuso, P. and Reverberi, P. (2003), 'Operating costs and market organization in railway services. the case of italy, 1980–1995', *Transportation Research Part B: Methodological* **37**(1), 43.

Marschak, J. and Andrews, W. H. (1944), 'Random simultaneous equations

and the theory of production', *Econometrica, Journal of the Econometric Society* pp. 143–205.

Marti, M., Neuenschwander, R. and Walker, P. (2009), Annex 1b - track maintenance and renewals costs in switzerland. deliverable 8, eu sixth framework project catrin (cost allocation of transport infrastructure cost), Technical report.

McGeehan, H. (1993), 'Railway costs and productivity growth: The case of the republic of ireland, 1973-1983', *Journal of Transport Economics and Policy* **27**, 19–32.

Merkert, R. and Hensher, D. A. (2011), 'The impact of strategic management and fleet planning on airline efficiency–a random effects tobit model based on dea efficiency scores', *Transportation Research Part A: Policy and Practice* **45**(7), 686–695.

Murillo-Zamorano, L. R. (2004), 'Economic efficiency and frontier techniques', *Journal of Economic Surveys* **18**(1), 33–77.

Nash, C. and Smith, A. (2007), 'Modelling performance: rail', *Handbook of transport modelling, Pergamon* pp. 565–575.

Nash, C. and Smith, A. (2014), 'Rail efficiency: cost research and its implications for policy'.

Olley, G. S. and Pakes, A. (1992), The dynamics of productivity in the telecommunications equipment industry, Technical report, National Bureau of Economic Research.

Ouellette, P., Petit, P., Tessier-Parent, L.-P. and Vigeant, S. (2010), 'Introducing regulation in the measurement of efficiency, with an application to the canadian air carriers industry', *European Journal of Operational Research* **200**(1), 216–226.

Oum, T. H. and Waters II, W. G. (1996), 'A survey on recent developments in transportation cost function research', *Logistics and Transportation Review* **32**(4), 423–464.

Oum, T. H., Waters, W. and Yu, C. (1999), 'A survey of productivity and efficiency measurement in rail transport', *Journal of transport economics and policy* pp. 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* pp. 121–138.

Oum, T. H., Yu, C. and Fu, X. (2003), 'A comparative analysis of productivity performance of the world's major airports: summary report of the atrs global airport benchmarking research report—2002', *Journal of Air Transport Management* **9**(5), 285–297.

Parisio, L. (1999), 'A comparative analysis of european railroads efficiency: a cost frontier approach', *Applied Economics* **31**(7), 815–823.

Pels, E., Daniel, V. E. and Rietveld, P. (2004), Returns to density in operations of the netherlands railways, Technical report, Tinbergen Institute, Tinbergen Institute Discussion Papers TI 2004-004/3.

Perelman, S. and Pestieau, P. (1988), 'Technical performance in public enterprises: A comparative study of railways and postal services', *European Economic Review* **32**(2), 432–441.

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. (2008), 'Public transport subsidisation', *The Implementation and Effectiveness of Transport Demand Management measures: An International Perspective* pp. 189–209.

Pucher, J., Markstedt, A. and Hirschman, I. (1983), 'Impacts of subsidies on the costs of urban public transport', *Journal of Transport Economics and Policy* pp. 155–176.

Quddus, M. A., Harris, N. and Graham, D. J. (2007), 'Metro station operating costs: an econometric analysis'.

Railway Gazette (2014), 'Chinese cities open 10 more metro lines', **12**.
URL: *http://www.railwaygazette.com/news/urban/single-view/view/chinese-cities-open-10-more-metro-lines.html*

Rao, D. P., O'Donnell, C. J., Battese, G. E. and Coelli, T. (1998), 'An introduction to efficiency and productivity analysis'.

Roodman, D. (2009), 'How to do xtabond2: An introduction to difference and system gmm in stata', *Center for Global Development working paper* (103).

Sameni, M. and Preston, J. (2012), 'Value for railway capacity: Assessing efficiency of operators in great britain', *Transportation Research Record: Journal of the Transportation Research Board* (2289), 134–144.

Sampaio, B. R., Neto, O. L. and Sampaio, Y. (2008), 'Efficiency analysis of public transport systems: Lessons for institutional planning', *Transportation research part A: policy and practice* **42**(3), 445–454.

Sanchez, P. C. and Villaroya, J. M. (2000), 'Efficiency, technical change and productivity in the european rail sector: a stochastic frontier approach', *International Journal of Transport Economics/Rivista internazionale di economia dei trasporti* pp. 55–76.

Santos, J., Simoes, P., Costa, A. and Marques, R. C. (2013), 'Efficiency of european metros: the portuguese case', *Proceedings of the ICE-Transport* **167**(3), 143–155.

Savage, I. (1997), 'Scale economies in united states rail transit systems', *Transportation Research Part A: Policy and Practice* **31**(6), 459–473.

Seiford, L. M. (1996), 'Data envelopment analysis: the evolution of the state of the art (1978–1995)', *Journal of Productivity Analysis* **7**(2-3), 99–137.

Simar, L. and Wilson, P. W. (2000), 'A general methodology for boot-

strapping in non-parametric frontier models', *Journal of applied statistics* **27**(6), 779–802.

Simar, L. and Wilson, P. W. (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of econometrics* **136**(1), 31–64.

Smith, A., Wheat, P. and Smith, G. (2010), 'The role of international benchmarking in developing rail infrastructure efficiency estimates', *Utilities policy* **18**(2), 86–93.

Soto, M. et al. (2009), 'System gmm estimation with a small sample'.

Tavares, G. (2002), 'A bibliography of data envelopment analysis (1978-2001)', *RUTCOR, Rutgers University* **11**, 14.

Tervonen, J. and Idstrom, T. (2004), Marginal rail infrastructure costs in finland 1997-2002, Technical report, Report from the Finnish Rail Administration.

The Economist (2011), 'Privatising amtrak. a modest proposal'.

The Economist (2012), 'Metro economies'.

The Economist (2013), 'Going underground'.

Tovar, B. and Martín-Cejas, R. R. (2010), 'Technical efficiency and productivity changes in spanish airports: a parametric distance functions approach', *Transportation Research Part E: Logistics and Transportation Review* **46**(2), 249–260.

Tsai, C.-H. P., Mulley, C. and Merkert, R. (2015), 'Measuring the cost efficiency of urban rail systems: An international comparison using dea and tobit models', *Journal of Transport Economics and Policy (JTEP)* **49**(1), 17–34.

UITP (2014), *World Metro Figures 2013*, International Association of Public Transport.

Union Bank of Switzerland (2012), *Union Bank of Switzerland Prices and Earnings Survey*, UBS.

Varian, H. R. and Norton, W. (1984), *Microeconomic analysis*, Vol. 2, Norton New York.

Venables, A. J. (2007), 'Evaluating urban transport improvements: cost-benefit analysis in the presence of agglomeration and income taxation', *Journal of transport economics and policy* pp. 173–188.

Viton, P. A. (1980), 'On the economics of rapid-transit operations', *Transportation Research Part A: General* **14**(4), 247–253.

Viton, P. A. (1993), 'Once again, the costs of urban rapid transit', *Transportation Research Part B: Methodological* **27**(5), 401–412.

Wang, L., Chu, J. and Wu, J. (2007), 'Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process', *International Journal of Production Economics* **107**(1), 161–163.

Wang, S.-E. and Liao, C.-H. (2006), 'Cost structure and productivity growth

of the taiwan railway', *Transportation Research Part E: Logistics and Transportation Review* **42**(4), 317–339.

Wanke, P. F., Barbastefano, R. G. and Hijjar, M. F. (2011), 'Determinants of efficiency at major brazilian port terminals', *Transport Reviews* **31**(5), 653–677.

Wooldridge, J. (2012), *Introductory econometrics: A modern approach*, Cengage Learning.

Wooldridge, J. M. (2010), *Econometric analysis of cross section and panel data*, MIT press.

World Bank (2014), *World Development Indicators 2014*, World Bank Publications.

Zhang, M. (2009), 'Bus versus rail: Meta-analysis of cost characteristics, carrying capacities, and land use impacts', *Transportation Research Record: Journal of the Transportation Research Board* (2110), 87–95.

11 — Annex

11.1 Total and average cost elasticities

The cost elasticities estimated for the average cost function (ϵ_{AC}) can easily be transformed into cost elasticities for a total cost function (ϵ_T).

The cost function of a metro producing x units of output (e.g. car kilometres) can be formulated as $c = f(x)$, so the total cost elasticities are defined in Equation 11.1:

$$\epsilon_T = \frac{\partial c}{\partial x}\frac{x}{c} = \frac{\partial c/\partial x}{c/x} = \frac{MC}{AC} \tag{11.1}$$

Given that the average cost is defined as $AC = \frac{c}{x}$, the elasticity of average cost is shown in Equation 11.2:

$$\epsilon_{AC} = \frac{\partial AC}{\partial x}\frac{x}{AC} = \frac{\partial}{\partial x}(c/x)\frac{x}{c/x} = \frac{x^2}{c}\left[\frac{x\frac{\partial c}{\partial x} - c}{x^2}\right] = \frac{x}{c}\frac{\partial c}{\partial x} - 1 = \epsilon_T - 1 \tag{11.2}$$

Consequently, the cost elasticity for the average cost as estimated in the results models for Chapters 5 to 8 can be transformed into total cost elasticity be means of adding 1 to the coefficients as in Equation 11.3:

$$\epsilon_T = \epsilon_{AC} - 1 \tag{11.3}$$

11.2 Reduced Operating Costs Functions - GMM

Table 11.1: Operating Cost Reduced Model Results - GMM

	Operating Costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	0.783***		
	(0.095)		
Pax. journeys [Ln]		0.811***	
		(0.109)	
Train hours [Ln]			0.847***
			(0.082)
Wages [Ln]	0.152**	0.095	0.194**
	(0.067)	(0.073)	(0.084)
Elect. price [Ln]	0.097***	−0.007	0.114*
	(0.029)	(0.038)	(0.059)
Staff hours [Ln]	0.157***	0.061	0.051
	(0.054)	(0.061)	(0.047)
Total elect. [Ln]	−0.061**	−0.124	0.085
	(0.027)	(0.079)	(0.076)
Network length [Ln]	−0.057	0.270***	0.174*
	(0.077)	(0.101)	(0.102)
Number of stations [Ln]	−0.075	−0.198**	−0.359***
	(0.069)	(0.092)	(0.096)
Staff outsourcing [Ln]	−0.335**	−0.147	−0.243***
	(0.143)	(0.163)	(0.076)
AR(1)	−1.51 (0.13)	−2.09 (0.037)	−2.2 (0.028)
AR(2)	−1.08 (0.281)	−1.8 (0.072)	−0.96 (0.336)
Sargan	12.06 (0.914)	10.3 (0.962)	8.98 (0.983)
Observations	141	141	141

Note: *p<0.1; **p<0.05; ***p<0.01

11.3 Production function - Comparison estimations

Table 11.2: Car kilometres - Production functions

	Car kilometres produced		
	Pooled OLS	Within groups	Diff-GMM
	(1)	(2)	(3)
lag(CKM, 1)	0.621***	0.114***	0.065
	(0.033)	(0.025)	(0.082)
L	0.057***	0.026	0.012
	(0.015)	(0.020)	(0.018)
E	0.107***	0.497***	0.478***
	(0.025)	(0.042)	(0.132)
NL	0.184***	0.463***	0.388**
	(0.029)	(0.076)	(0.184)
ST	−0.119***	−0.319***	−0.053
	(0.031)	(0.102)	(0.243)
FL	0.186***	0.309***	0.275***
	(0.032)	(0.039)	(0.092)
Constant	−0.613***		
	(0.100)		
AR(1)			−1.95 (0.052)
AR(2)			−1.64 (0.102)
Sargan			16.81 (0.157)
Observations	211	211	27
R^2	0.994	0.972	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 11.3: Passenger journeys - Production functions

	Passenger journeys produced		
	Pooled OLS	Within groups	Diff-GMM
	(1)	(2)	(3)
lag(PJN, 1)	0.893***	0.507***	0.455***
	(0.028)	(0.038)	(0.090)
L	0.108***	0.018	0.039
	(0.018)	(0.031)	(0.025)
E	0.041*	0.401***	0.451***
	(0.023)	(0.066)	(0.131)
NL	−0.011	0.007	−0.044
	(0.030)	(0.114)	(0.130)
ST	0.006	−0.146	0.075
	(0.029)	(0.155)	(0.139)
FL	−0.025	0.119*	0.046
	(0.029)	(0.061)	(0.053)
Constant	0.376***		
	(0.071)		
AR(1)			−1.8 (0.071)
AR(2)			−0.41 (0.681)
Sargan			14.83 (0.251)
Observations	212	212	27
R^2	0.994	0.922	
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11.4: Train hours - Production functions

	Train hours produced		
	Pooled OLS	Within groups	Diff-GMM
	(1)	(2)	(3)
lag(THR, 1)	0.891***	0.124**	0.014
	(0.028)	(0.048)	(0.199)
L	0.067***	0.052	0.054**
	(0.020)	(0.034)	(0.023)
E	−0.023	0.419***	0.500**
	(0.028)	(0.083)	(0.207)
NL	0.075**	0.434***	0.103
	(0.037)	(0.153)	(0.344)
ST	−0.015	−0.167	0.222
	(0.040)	(0.205)	(0.350)
FL	0.011	0.124*	0.184*
	(0.030)	(0.068)	(0.099)
Constant	−0.422**		
	(0.170)		
AR(1)			−0.77 (0.439)
AR(2)			0.66 (0.511)
Sargan			11.18 (0.513)
Observations	178	178	27
R^2	0.990	0.895	
Note:			*p<0.1; **p<0.05; ***p<0.01

11.4 Detailed definition of explanatory variables

Table 11.5: Operating Cost - Explanatory variables definition

Variable	Definition
Car km (m)	Car kilometres actually operated in revenue service excluding empty stock movements, movements from depots, engineering trains, driver training runs and cancellations of scheduled runs.
Passenger journeys (m)	Annual total of all passengers recorded or estimated as travelling from origin to destination irrespective of the fare paid.
Train hours (m)	Total actual car hours for all trains operating in revenue (passenger carrying) service.
Wages (PPP US$ per hour)	Total metro-related staff cost including salaries and all other labour costs (e.g. social security fees, charges or taxes borne by the employer, pension contributions and any other fringe benefits) over the total hours scheduled to be worked by all full-time, part time and temporary employees for work related to existing metro operations.
Electricity price	Cost of energy over total amount of energy consumed for all purposes (e.g. traction, infrastructure, stations, etc.).
Total labour hours (m)	Hours scheduled to be worked by own staff and contracted-out staff.
Total electricity (kWh)	Amount of energy consumed for both traction and non-traction.
Network length (km)	Total length of network used by trains operating in passenger service (adjusted for parallel lines), as a weighted average for the whole calendar year.
Number of stations	Total number of stations normally served by the metro organisations train services during the weekday peak hour. Multiline stations count as one.
% staff contracted out	Percentage of contracted out staff hours over all the total staff hours.
Incidents per car km	Incidents, of any cause, resulting in a delay to the train service of 5 minutes or more over the total revenue car kilometres.
Opening hours / week	Number of hours opened to the public per week.
Subsidy	Share of the total operating costs not covered by the total operating revenue.
Age of the network	Years since the start of metro operations.
Year	Time trend.

Table 11.6: Train Service - Explanatory variables definition

Variable	Definition
Car km (m)	Car kilometres actually operated in revenue service excluding empty stock movements, movements from depots, engineering trains, driver training runs and cancellations of scheduled runs.
Passenger journeys (m)	Annual total passengers recorded or estimated as travelling from origin to destination irrespective of the fare paid.
Train hours (m)	Total actual car hours for all trains operating in revenue (passenger carrying) service.
Wages (PPP US$ per hour)	Total metro-related staff cost including salaries and all other labour costs (e.g. social security fees, charges or taxes borne by the employer, pension contributions and any other fringe benefits) over the total hours scheduled to be worked by all full-time, part time and temporary employees for work related to existing metro operations.
Electricity price	Cost of energy over total energy consumed for all purposes (e.g. traction, infrastructure, stations, etc.).
Train service labour hours (m)	Hours scheduled to be worked for train drivers, other train-borne staff and non-train-borne staff that supports train operations (e.g. central control room).
Traction electricity	Amount of energy consumed for powering the trains measured in kWh.
% of train service staff contracted out	Percentage of contracted out train service staff hours over all the total train service hours.
Driver productivity	Percentage of the total driver time, both own staff and contracted, that has been transformed into driving (i.e. actual revenue train hours).
Standardized train capacity (pax/ train)	Sum of all seats available plus an average of four standees per standing square metre for each car. All multiplied by the average number of cars per train.
Network length (km)	Total length of network used by trains operating in passenger service (adjusted for parallel lines), as a weighted average for the whole calendar year.
Number of stations	Total number of stations normally served by the metro organisations train services during the weekday peak hour. Multiline stations count as one.
% of AC rolling stock	Percentage of fleet (car units) with air conditioning over the total fleet. Average over the year.
Age of the network	Years since the start of metro operations.
Year	Time trend.

Table 11.7: Station Service - Explanatory variables definition

Variable	Definition
Passenger journeys (m)	Annual total of all passengers recorded or estimated as travelling from origin to destination irrespective of the fare paid.
Number of stations	Total number of stations normally served by the metro organisations train services during the weekday peak hour. Multiline stations count as one.
Wages (PPP US$ per hour)	Total metro-related staff cost including salaries and all other labour costs (e.g. social security fees, charges or taxes borne by the employer, pension contributions and any other fringe benefits) over the total hours scheduled to be worked by all full-time, part time and temporary employees for work related to existing metro operations.
Electricity price	Cost of energy over total energy consumed for all purposes (e.g. traction, infrastructure, stations, etc.).
Station service labour hours (m)	Hours scheduled to be worked for station management, supervision, platform supervision, part time platform assistants, multi-functional duties and operational control of subcontractors used for station operations duties
Non-traction electricity (kWh)	Amount of energy consumed for powering the metro's stations and other facilities and infrastructure.
Train hours (m)	Total actual car hours for all trains operating in revenue (passenger carrying) service.
% of station service staff contracted out	Percentage of contracted out station service staff hours over all the station service staff hours.
Peak ratio	Ratio between the highest demand level in the peak hours over the lowest demand level in the inter-peak hours.
Number of escalators, lifts and travelators	Number of lifts, escalators and travelators.
Train length (meters)	Average car length multiplied by the average number of cars per train.
Age of the network	Years since the start of metro operations.
Year	Time trend.

Table 11.8: Rolling Stock Maintenance - Explanatory variables definition

Variable	Definition
Car km (m)	Car kilometres actually operated in revenue service excluding empty stock movements, movements from depots, engineering trains, driver training runs and cancellations of scheduled runs.
Fleet (number of cars)	Total number of cars owned/leased by the metro which are suitable for operation in normal passenger service. Average over the year.
Wages (PPP US$ per hour)	Total metro-related staff cost including salaries and all other labour costs (e.g. social security fees, charges or taxes borne by the employer, pension contributions and any other fringe benefits) over the total hours scheduled to be worked by all full-time, part time and temporary employees for work related to existing metro operations.
RS maintenance staff hours (m)	Hours scheduled to be worked in maintaining passenger and engineering trains, train cleaning and maintenance of buildings (i.e. depots) for maintaining equipment for rolling stock maintenance. This variable includes both own and contracted out staff hours.
Fleet availability at peak (%)	Total number of rail cars available for use in the weekday peak hour. Average over the year.
% RS maintenance staff contracted out	Percentage of contracted out rolling stock maintenance hours over all the rolling stock maintenance hours.
Average speed (km/h)	Average scheduled speed at which trains operate in passenger service.
RS Mean distance between failure (m of km)	Incidents caused by rolling stock failure resulting in a delay to the train service of 5 minutes or more over the total actual operated revenue car kilometres.
% of rolling stock with AC	Percentage of fleet (car units) with air conditioning over the total fleet. Average over the year.
Rolling stock age	Average age of the fleet (car units) in commercial operation. Average over the year.
Age of the network	Years since the start of metro operations.
Year	Time trend.

11.5 Econometric models with age categorisation

Table 11.9: Operating Cost Model Results - Age groups

	Operating costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.800***		
	(0.083)		
Pax. journeys [Ln]		−0.810***	
		(0.051)	
Train hours [Ln]			−0.370***
			(0.072)
Wages [Ln]	0.360***	0.330***	0.210***
	(0.052)	(0.052)	(0.062)
Elect. price [Ln]	0.160***	0.160***	0.120***
	(0.022)	(0.021)	(0.024)
Staff hours [Ln]	0.530***	0.470***	0.340***
	(0.059)	(0.061)	(0.078)
Total elect. [Ln]	0.380***	0.380***	0.017
	(0.077)	(0.063)	(0.074)
Network length [Ln]	0.120	0.250***	−0.160
	(0.096)	(0.091)	(0.120)
Number of stations [Ln]	−0.440***	−0.480***	−0.160
	(0.100)	(0.100)	(0.140)
Staff outsourcing [Ln]	−0.520***	−0.510***	−0.510**
	(0.180)	(0.180)	(0.250)
Incidents per car km [Ln]	0.041***	0.044***	0.047***
	(0.010)	(0.010)	(0.011)
Opening hours per week [Ln]	0.094	0.240	−0.170
	(0.280)	(0.270)	(0.330)
Subsidsy (% of opex) [Ln]	0.500***	0.560***	0.530***
	(0.035)	(0.035)	(0.039)
Old metros (before 1945)	0.260***	0.270***	0.360**
	(0.087)	(0.089)	(0.160)
New metros (after 1988)	0.190**	0.170*	0.490***
	(0.092)	(0.088)	(0.170)
Time trend	0.009***	0.009***	0.011***
	(0.002)	(0.002)	(0.003)
Constant	−18.000***	−18.000***	−21.000***
	(4.500)	(4.400)	(5.200)
Rho	0.8501	0.7882	0.8522
Observations	136	136	136
R^2	0.499	0.660	0.841

Note: *p<0.1; **p<0.05; ***p<0.01

Table 11.10: Train Service Cost Model Results - Age groups

	Train service costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.990***		
	(0.240)		
Pax. journeys [Ln]		−0.940***	
		(0.110)	
Train hours [Ln]			−0.800***
			(0.210)
Wages [Ln]	0.620***	0.580***	0.610***
	(0.095)	(0.091)	(0.096)
Elect. price [Ln]	0.200***	0.210***	0.190***
	(0.069)	(0.068)	(0.069)
Staff hours [Ln]	0.530***	0.460***	0.460***
	(0.130)	(0.130)	(0.150)
Traction elect. [Ln]	0.490**	0.420***	0.430***
	(0.190)	(0.130)	(0.140)
Staff outsourcing [Ln]	−0.520	−0.490	−0.480
	(0.330)	(0.330)	(0.330)
Driver productivity [Ln]	−0.250**	−0.270**	−0.330**
	(0.120)	(0.120)	(0.140)
Train capacity [Ln]		0.140	
		(0.180)	
Network length [Ln]	0.099	0.120	0.098
	(0.220)	(0.170)	(0.190)
Number of stations [Ln]	−0.250	−0.160	−0.300
	(0.230)	(0.210)	(0.230)
RS with air con. (%)	−0.084	−0.150	−0.082
	(0.140)	(0.130)	(0.140)
Old metros (before 1945)	0.140	0.180	0.094
	(0.160)	(0.150)	(0.170)
New metros (after 1988)	0.260	0.240	0.230
	(0.180)	(0.160)	(0.190)
Time trend	−0.007	−0.007	−0.008
	(0.005)	(0.005)	(0.005)
Constant	14.000	12.000	16.000
	(11.000)	(11.000)	(11.000)
Rho	0.8327	0.8562	0.8742
Observations	138296	138	138
R^2	0.494	0.650	0.823
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11.11: Station Service Cost Model Results - Age groups

	Station service costs per	
	Pax. journey	Station
	(1)	(2)
Pax. journeys [Ln]	−1.310***	
	(0.163)	
No. of stations [Ln]		−1.153***
		(0.147)
Wages [Ln]	0.684***	0.719***
	(0.195)	(0.189)
Elect. price [Ln]	−0.089	−0.085
	(0.107)	(0.106)
Station staff hr [Ln]	0.637***	0.531***
	(0.127)	(0.107)
Non-traction elect. [Ln]	0.183	0.151
	(0.164)	(0.159)
Staff outsourcing [Ln]	−0.791**	−0.817**
	(0.359)	(0.354)
Lifts and esc. [Ln]	0.356**	0.287*
	(0.164)	(0.152)
Train length	0.013***	0.011***
	(0.003)	(0.003)
Old metros (before 1945)	0.672***	0.731***
	(0.226)	(0.234)
New metros (after 1988)	0.652**	0.794***
	(0.283)	(0.255)
Time trend	0.012	0.008
	(0.011)	(0.011)
Constant	−26.630	−17.507
	(22.022)	(21.775)
Rho	0.6793	0.6321
Observations	122	122
R^2	0.577	0.643
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 11.12: RS Maintenance Cost Model Results - Age groups

	RS Maintenance cost per	
	car km	fleet
	(1)	(2)
Car km [Ln]	−0.483***	
	(0.108)	
Fleet size [Ln]		−0.463***
		(0.110)
Wages [Ln]	0.612***	0.557***
	(0.084)	(0.083)
Staff hours [Ln]	0.503***	0.529***
	(0.092)	(0.086)
Staff outsourcing [Ln]	−0.750***	−0.759***
	(0.235)	(0.233)
Fleet availability [Ln]	0.349	0.820**
	(0.317)	(0.319)
Speed [Ln]	0.002	0.077
	(0.132)	(0.130)
MDBF-RS [Ln]	0.0003	−0.001
	(0.025)	(0.025)
RS with air con.	0.157	0.157
	(0.134)	(0.132)
Fleet age	−0.008	−0.007
	(0.014)	(0.014)
Fleet age squared	0.0002	0.0002
	(0.0003)	(0.0003)
Old metros (before 1945)	0.100	−0.003
	(0.133)	(0.140)
New metros (after 1988)	0.152	0.112
	(0.186)	(0.184)
Time trend	−0.014**	−0.010
	(0.006)	(0.006)
Constant	28.402**	17.781
	(12.635)	(12.532)
Rho	0.7847	0.7786
Observations	142	142
R^2	0.527	0.702
Note:	*p<0.1; **p<0.05; ***p<0.01	

11.6 Econometric models without input quantities

Table 11.13: Operating Cost Model Results - No input quantities

	Operating costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.329***		
	(0.056)		
Pax. journeys [Ln]		−0.482***	
		(0.045)	
Train hours [Ln]			−0.284***
			(0.058)
Wages [Ln]	0.152***	0.107*	0.077
	(0.057)	(0.062)	(0.059)
Elect. price [Ln]	0.124***	0.135***	0.100***
	(0.024)	(0.027)	(0.026)
Network length [Ln]	0.217*	0.577***	−0.003
	(0.128)	(0.124)	(0.123)
Number of stations [Ln]	−0.325**	−0.378**	−0.135
	(0.144)	(0.149)	(0.157)
Staff outsourcing [Ln]	0.007	0.167	0.090
	(0.221)	(0.230)	(0.231)
Incidents per car km [Ln]	0.038***	0.038***	0.041***
	(0.011)	(0.012)	(0.012)
Opening hours per week [Ln]	−0.050	0.372	−0.336
	(0.331)	(0.357)	(0.347)
Subsidy (% of opex) [Ln]	0.575***	0.700***	0.549***
	(0.041)	(0.044)	(0.042)
Network age (years) [Ln]	0.002	0.003	0.003
	(0.002)	(0.002)	(0.002)
Time trend	0.013***	0.009***	0.012***
	(0.003)	(0.003)	(0.003)
Constant	−22.787***	−18.605***	−21.701***
	(6.652)	(6.442)	(6.748)
Rho	0.987	0.9738	0.9851
Observations	141	141	141
R^2	0.704	0.789	0.680
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11.14: Train Service Costs Model Results - No input quantities

	Train service costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−0.060		
	(0.100)		
Pax. journeys [Ln]		−0.490***	
		(0.094)	
Train hours [Ln]			0.137
			(0.124)
Wages [Ln]	0.511***	0.523***	0.613***
	(0.101)	(0.105)	(0.105)
Elect. price [Ln]	0.247***	0.188**	0.154**
	(0.074)	(0.078)	(0.074)
Staff outsourcing [Ln]	0.073	−0.120	−0.262
	(0.331)	(0.338)	(0.307)
Driver productivity [Ln]	−0.563***	−0.527***	−0.749***
	(0.105)	(0.110)	(0.106)
Train capacity [Ln]		0.208	
		(0.199)	
Network length [Ln]	−0.345	0.430**	0.210
	(0.208)	(0.182)	(0.200)
Number of stations [Ln]	0.368*	0.177	−0.330
	(0.212)	(0.237)	(0.256)
RS with air con. (%)	0.071	−0.198	−0.141
	(0.116)	(0.131)	(0.139)
Network age (years)	−0.0004	0.001	−0.004
	(0.002)	(0.002)	(0.002)
Time trend	−0.007	−0.014**	−0.009*
	(0.006)	(0.006)	(0.005)
Constant	10.117	23.133*	21.552**
	(10.950)	(11.690)	(10.742)
Rho	0.777	0.7945	0.8651
Observations	136	136	136
R^2	0.359	0.535	0.797
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 11.15: Station Service Cost Model Results - No input quantities

	Station service costs per	
	Pax. journey	Station
	(1)	(2)
Pax. journeys [Ln]	−1.052***	
	(0.171)	
No. of stations [Ln]		−0.812***
		(0.263)
Wages [Ln]	0.064	0.079
	(0.268)	(0.263)
Elect. price [Ln]	−0.170	−0.194
	(0.122)	(0.125)
Staff outsourcing [Ln]	−0.186	−0.206
	(0.438)	(0.438)
Lifts and esc. [Ln]	0.741***	0.639***
	(0.201)	(0.187)
Train length	0.014***	0.013***
	(0.004)	(0.003)
Network age (years)	0.012**	0.009
	(0.005)	(0.006)
Time trend	0.004	0.006
	(0.013)	(0.013)
Constant	−9.824	−14.632
	(24.692)	(25.759)
Rho	0.8884	0.8848
Observations	122	122
R^2	0.330	0.290
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 11.16: RS Maintenance Cost Model Results - No input quantities

	RS Maintenance cost per	
	car km	fleet
	(1)	(2)
Car km [Ln]	0.042	
	(0.064)	
Fleet size [Ln]		0.141*
		(0.073)
Wages [Ln]	0.650***	0.581***
	(0.095)	(0.096)
Staff outsourcing [Ln]	−0.417	−0.418
	(0.252)	(0.256)
Fleet availability [Ln]	0.283	1.222***
	(0.346)	(0.344)
Speed [Ln]	−0.092	0.040
	(0.140)	(0.141)
MDBF-RS [Ln]	−0.003	0.0001
	(0.026)	(0.026)
RS with air con.	0.256	0.225
	(0.165)	(0.171)
Fleet age	−0.036***	−0.034**
	(0.013)	(0.013)
Fleet age squared	0.001***	0.001***
	(0.0003)	(0.0003)
Network age (years)	−0.001	−0.004*
	(0.002)	(0.002)
Time trend	−0.011*	−0.0004
	(0.007)	(0.007)
Constant	20.766	−4.264
	(13.202)	(13.014)
Rho	0.851	0.8571
Observations	142	142
R^2	0.426	0.588
Note:	*p<0.1; **p<0.05; ***p<0.01	

11.7 Breusch-Pagan Test Results

Table 11.17: Breusch-Pagan Lagrange Multiplier Test

model	chi_sq	p_value
Operating cost per car km	1034.44	0.00
Operating costs per passenger journey	1338.69	0.00
Operating costs per train hour	2073.53	0.00
Train service costs per car km	2063.07	0.00
Train service costs per passenger journey	1746.18	0.00
Train service costs per train hour	2676.04	0.00
Station service costs per passenger journey	1602.42	0.00
Station service costs per station	1766.50	0.00
Rolling stock maintenance costs per car km	3153.14	0.00
Rolling stock maintenance costs per car	2950.10	0.00

Note: H_0 is no variances across entities.

11.8 SFA Efficiency Index

The stochastic frontier analysis (SFA) approach has been described in section 3.2.2 on page 83. That section reviews the SFA methodology and explains why SFA is not be a suitable approach for this research with the CoMET-Nova data set.

This annex presents the estimates using the SFA approach for the operating costs category. Table 11.18 shows the results for the three models (i.e. operating costs per car km, per passenger journey and per train hour). As it can be seen, most of the variables are not significant with p-values between brackets far above the usual threshold of 0.1. This is due to the small size of the CoMET-Nova data set both in terms of cross sectional units and time series. The lack of data can lead to biased results when using maximum likelihood to estimate the SFA models as reported in the the methodology chapter (section 3.2.2 on page 83). Consequently, the efficiency scores estimated in the model are also unreliable.

The efficiency scores estimated with the SFA approach are clearly inaccurate. The efficiency scores for the three models are almost identical between the different metros as shown in the descriptive statistics (Table 11.19) and the anonymised graphs for operating costs per car km (Figure 11.1), per passenger journey (Figure 11.2) and per train hour (Figure 11.3). These results have been considered unfeasible following the expertise of RTSC professionals but also the evident contrast with the variation of operating costs across metros shown in Figure 5.2, 5.3 and 5.4 in pages 121 to 123.

Overall, it can be concluded that due to the data characteristics of the CoMET-Nova data set, the SFA approach is not appropriate for the study of the urban rail cost efficiency. However, the CoMET-Nova data set will increase over time, with larger time series, and also in cross sectional units, with new metro members. This may enable other researchers to revisit this work and apply the SFA in the future.

Figure 11.1: SFA Efficiency indeces for operating costs per car kilometre

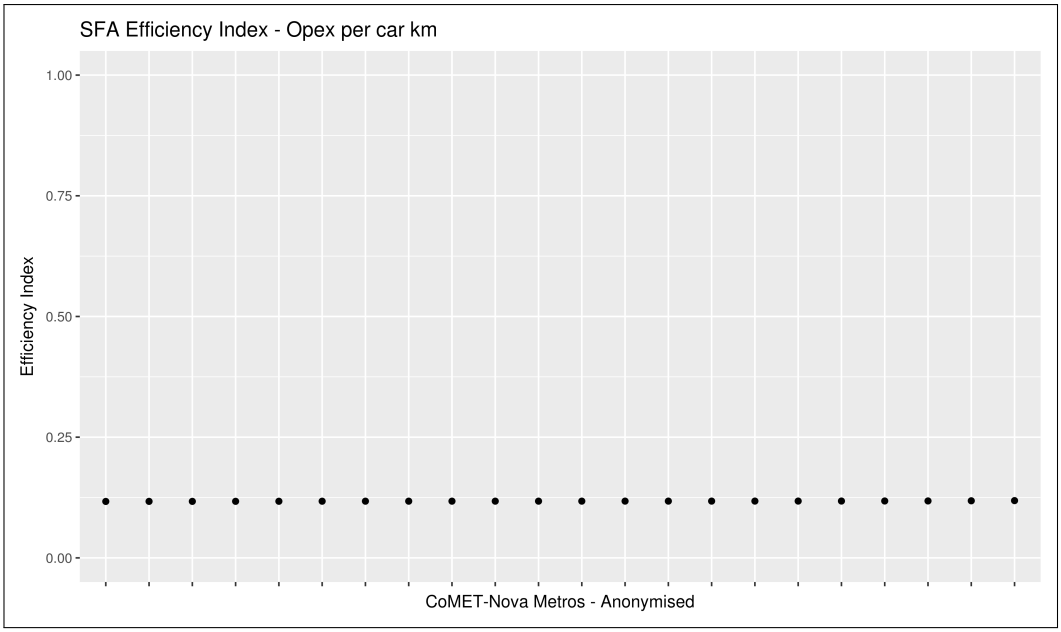


Table 11.18: Operating Cost Model Results - SFA

	Operating Costs per		
	car km	pax. journeys	train hour
	(1)	(2)	(3)
Car km [Ln]	−1.111		
	(0.909)		
Pax. Journeys [Ln]		−0.994	
		(0.321)	
Train hours [Ln]			−1.119
			(0.265)
Wages [Ln]	0.487	0.436	0.469
	(0.626)	(0.663)	(0.639)
Elect. Price [Ln]	0.189	0.185	0.206
	(0.850)	(0.852)	(0.836)
Staff hours [Ln]	0.565	0.529	0.561
	(0.572)	(0.597)	(0.575)
Total elect. [Ln]	0.609	0.550	0.596
	(0.543)	(0.583)	(0.552)
Network length [Ln]	0.273	0.233	0.218
	(0.785)	(0.815)	(0.827)
Number of stations [Ln]	−0.392	−0.390	−0.319
	(0.695)	(0.697)	(0.749)
Staff outsourcing [Ln]	−0.118	0.054	−0.007
	(0.906)	(0.956)	(0.993)
Incidents per car km [Ln]	0.029	0.033	0.029
	(0.976)	(0.973)	(0.976)
Opening hours per week [Ln]	−0.580	−0.297	−0.595
	(0.562)	(0.766)	(0.552)
Subsidy (% of opex) [Ln]	0.286	0.303	0.296
	(0.774)	(0.762)	(0.767)
Network Age (years)	0.003	0.003	0.003
	(0.997)	(0.997)	(0.996)
Time trend	0.001	0.001	0.001
	(0.999)	(0.998)	(0.999)
Constant	0.113	−2.795	−0.455
	(0.909)	(0.006)	(0.649)
Sigma2	0.00565	0.00602	0.00573
Sigma2v	0.00282	0.00301	0.00286

**Values in brackets are p-values.*

Table 11.19: SFA Efficiency Index Descriptive Statistics

Variable	Mean	Median	Std.Dev	Min	Max
Opex per car km	0.1175	0.1175	0.004	0.1169	0.1186
Opex per pax journey	0.2347	0.2352	0.0028	0.228	0.2387
Opex per train hour	0.2307	0.2302	0.003	0.2257	0.2387

Figure 11.2: SFA Efficiency indeces for operating costs per passenger journey

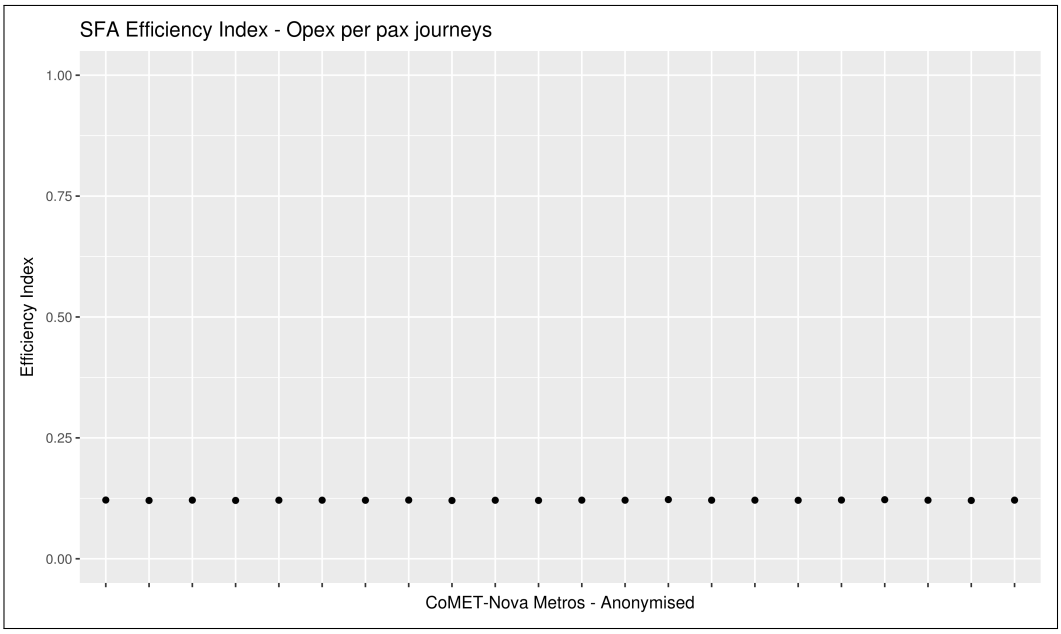


Figure 11.3: SFA Efficiency indeces for operating costs per train hour

