Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

by

Maite (Maria Teresa) Pena-Alcaraz

M.S., Engineering Systems Modeling, Comillas Pontifical University, Madrid (Spain), 2010
B.S., Industrial Engineering, Comillas Pontifical University, Madrid (Spain), 2009
B.S., Mathematics, Complutense University of Madrid (Spain), 2012

Submitted to the Engineering Systems Division
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature of Author ……………………………………………………………….........................
Engineering Systems Division
May 4, 2015

Certified by ……………………………………………………….......
Joseph M. Sussman
JR East Professor of Civil and Environmental Engineering and Engineering Systems
Thesis Committee Chair

Certified by ……………………………………………………….......
Mort D. Webster
Associate Professor of Energy Engineering, Pennsylvania State University
Thesis Committee Member

Certified by ……………………………………………………….......
Andres Ramos
Professor, Comillas Pontifical University
Thesis Committee Member

Certified by ……………………………………………………….......
Jose Antonio Gomez-Ibanez
Derek C. Bok Professor of Urban Planning and Public Policy, GSD and HKS, Harvard University
Thesis Committee Member

Certified by ……………………………………………………….......
Ignacio Pérez-Arriaga
Visiting Professor at the Center for Energy and Environmental Policy Research, MIT
Professor and Director of the BP Chair on Energy and Sustainability, Comillas Pontifical University
Thesis Committee Member

Accepted by ……………………………………………………….......
Munther A. Dahleh
William A. Coolidge Professor of Electrical Engineering and Computer Science
Director, Institute for Data, Systems and Society

1

To my parents, for teaching us the value of effort and education

To my siblings, Ali and Manuel, for reminding me the importance of challenging assumptions

To my husband, Noel, for supporting me on this journey

Analysis of Capacity Pricing and Allocation Mechanisms in Shared Railway Systems

by

Maite Pena-Alcaraz

Submitted to the Engineering Systems Division
on May 4, 2015 in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Engineering Systems

Abstract

In the last 15 years, the use of rail infrastructure by different train operating companies (shared railway system) has been proposed as a way to improve infrastructure utilization and to increase efficiency in the railway industry. Shared use requires coordination between the infrastructure manager and multiple train operators. Such coordination requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, *capacity allocation*, and the access charges they have to pay, *capacity pricing*.

The objective of this thesis is to contribute to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in representative shared railway systems.

There are strong interactions between capacity planning and infrastructure operations in the railway industry; the operations on the infrastructure determine the available capacity in the system. As a consequence, the framework developed in this thesis to evaluate the performance of shared railway systems under alternative capacity pricing and allocation consists of two models: 1) a train operator model and 2) an infrastructure manager model. The train operator model is a financial model that anticipates how train operators would respond to the capacity pricing and allocation mechanisms and determine their demand for infrastructure use. The infrastructure manager model is a network optimization model that determines the optimal train timetable (infrastructure manager's decisions) that accommodates the train operators' demands for scheduling trains, considering the topology of the system, safety constraints, and other technical aspects of the infrastructure for shared railway systems. To be able to solve the train timetabling optimization problem in meaningful instances, this thesis develops a novel approximate dynamic programming algorithm based on linear programming that extends previous algorithms proposed in the literature to effectively solve large network optimization problems.

This thesis then uses the train operator model to compare the operational decisions of train operators in shared railway systems with the operational decisions of even-handed

integrated railway companies. We show that train operators in shared railway systems make the same operational decisions as an integrated railway company when variable access charges reflect variable infrastructure manager's costs to operate trains on the infrastructure. We also identify two cases in which the train operators may have incentives to deviate from the integrated railway systems' operational decisions: 1) when the infrastructure manager needs to recover part of the infrastructure management fixed costs, or 2) when the railway system is congested. This motivates the choice of the two case studies of this thesis, one based on the Central Corridor in Tanzania, and the other one based on the Northeast Corridor in the US.

We then show how to use the framework proposed in this thesis to analyze the trade-offs associated with the use of alternative mechanisms in these two cases. To our knowledge, this is the first effort to compare alternative mechanisms to price and allocate capacity in the same shared railway system. The results of this thesis show that there are important trade-offs associated with each mechanism and none of them is superior to the other on all dimensions. We thus recommend that system stakeholders carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking the system into one of them. This is particularly important today since several countries are currently restructuring their railway sector to allow shared use. We claim that the improved understanding of the system performance gained with the framework proposed in this thesis is important to be able to design adequate capacity pricing and allocation mechanisms that can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure in the railway industry.

Thesis Chair: Joseph M. Sussman
Title: JR East Professor of Civil and Environmental Engineering and Engineering Systems, Massachusetts Institute of Technology

Thesis Co-supervisor: Mort D. Webster
Title: Associate Professor of Energy Engineering, John and Willie Leone Family Department of Energy and Mineral Engineering, College of Earth and Mineral Sciences, Penn-State University

Acknowledgements

This PhD has been a life-changing journey. I want to briefly thank many people and institutions that have supported me during this journey and contributed to make it unforgettable, as well as those who helped me to "catch this train".

I want to start thanking my thesis committee. I have been lucky to have five extraordinary supervisors and mentors who have guided me during these years: professors Joseph Sussman, Mort Webster, Andres Ramos, Jose Antonio Gomez-Ibanez, and Ignacio Perez-Arriaga. Joe and Mort, thank you for your guidance, patience, example, and support. Andres, Tony, and Ignacio, thank you for sharing your knowledge and experience, and for always being happy to have a discussion about this research. I am indebted for all I have learned from the five of you, not only from a professional standpoint, but also from a personal standpoint.

I also want to acknowledge the support from two excellent scholarships. The research of this thesis was generously supported by La Caixa Foundation and Rafael del Pino Foundation. I want to thank Emilia, Pilar, Amadeo, and Vicente, who made sure that everything was going well during these years. I also want to thank other sources of funding that have supported this PhD. I want to acknowledge the support from the National University Rail Center (NURail), the Japan East Railway Company (JR East), the Japan International Transportation Institute (JITI), and the National Science Foundation (NSF). I was also fortunate to work with professors Richard Larson and Mort Webster as a teaching assistant in two quantitative courses they taught. Dick and Mort, thank you, not only for the funding support, but also for the opportunity to teach and learn from you.

I am very grateful to have been able to carry out my research at MIT. My work has tremendously benefited from many of the classes and excellent professors: John Tsitsiklis, David Autor, Nancy Rose, Richard de Neufville, Chris Magee, Fred Salvucci, Mikel Murga, and Eva Ponce, among others. This research has also benefited from many discussions with students in Professors Sussman and Webster's research groups. I want to thank my JITI project collaborators: Andres, Joel and Naomi, my JR East team, Joanna and Patton, as well as Sam, Alex, Bruno, Iori, Ryusuke, Tatsu, and Tolu for the very fruitful discussions about the NEC and about the role of rail infrastructure in the society. The nurturing research environment provided by my department, the Engineering Systems Division, and by my office micro-cosmos has also contributed immensely to this research. Thank you Abby, Alex, Berk, Fernando, Hector, Tommy, Vivek, and Claudia for all the research (and non-research) conversations. I have learned a lot form all of you. Thank you Beth and all the ESS leaders for making sure that we always had everything we needed to organize student seminars or any other useful (professional or social) activity.

This research has also benefited from internships and collaborations with the World Bank. Thank you Henry, Martin, Martha, Bernard, Sevara, and Jose Luis for

giving me the opportunity to work with you and learn about the railway system in Tanzania, India, Kenya, Rwanda, and Burundi. I also want to thank the Institute for Research in Technology (IIT), my "home" before MIT, where I learned so much about railways and optimization. Thank you Carlos Batlle, Alexis, and many others for being unconditional friends and advisers. I also want to thank Comillas Pontifical University and Portaceli for all they taught me over the years.

I also want to thank Enrique, Marta, Joel, Carles, Toni, Carlos, Xabi, Paula, Ferran, David, Carmen, Helena, Anna, Quique, Roberto, Paula, Rafa, Claire, David, Santiago, Chema, Lucinio, Carlos, Patty, Adam, Asun, and many other Spanish and non-Spanish friends who have made these years memorable. Thank you for all the barbeques, dinners, games, hikes, trips to the beach, and your support at many times. Thanks also to Coque, Nacho, Bea, Alberto, Javi, Mario, Chus, Diego, Ale, Richi, Luis, Elisa, Ceci, Gonzalo, Rocio and David, my undergraduate, mathematical competitions, and high-school friends, to keep calling me and writing to me even when I was busy and took ages to answer.

Last but not least, I want to thank my family for all their support. My grandparents, my aunts and uncles, my cousins, my parents, my siblings and my husband have been supporting me all this time. Mom and Dad, thank you for convincing me that I could take any train I want if I work hard enough. Ali and Manuel, thank you for being unconditional friends and allies on every journey. Noel, thank you for your love, for your support, for your company, for preparing the best meals and the best tuppers, and for smiling and encouraging me to continue when I was tired. This would have not been possible without you.

I know that trying to keep this brief, there are many of you that I did not named explicitly, although you have greatly contributed to this experience. Thank you.

With the desire of keep riding many new trains with you,
Maite Pena-Alcaraz

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADP	Approximate Dynamic Programming
ALP	Approximate Linear Programming
ARLP	Adaptive Relaxed Linear Programming
DP	Dynamic Programming
FRA	Federal Railroad Administration, US
HSR	High Speed Rail
IM	Infrastructure Manager
IRR	Internal Rate of Return
LP	Linear Programming
NEC	Northeast Corridor, US
NPV	Net Present Value
MILP	Mixed-Integer Linear Programming
OD	Origin-Destination (pair)
QARLP	Q-factor Adaptive Relaxed Linear Programming (algorithm)
RLP	Relaxed Linear Programming
TO	Train Operator

Chapter 1 - Introduction

"The more we share, the more we have" – L. Nimoy (1995)

Infrastructures create the opportunity for essential services that should be put in place to enable economic activity, economic growth, and development (Munnell, 1992; World Bank, 2008). However, investment in infrastructure is extremely expensive, so high demand and strong business cases are typically necessary to justify such investments. In the past, large infrastructure systems were vertically integrated, i.e., the same entity was in charge of both managing and operating the system. The lack of competition and efficiency incentives of those entities erodes the performance of the system and has motivated over the years the introduction of new institutional structures. Shared infrastructure systems are widely proposed today as to 1) minimize the amount of infrastructure required to serve societal needs, and 2) achieve high levels of utilization and help recover the costs of the very high capital-intensive infrastructure in service systems such as the electric power sector, telecommunications, or transportation.

The greatest challenge for shared infrastructures is managing and coordinating access of competitive agents to the infrastructure (Gomez-Ibanez, 2003). The key question is determining an appropriate mechanism for deciding which agents can access the infrastructure at each time (*capacity allocation*) and the access price that each agent should pay (*capacity pricing*). The benefits of sharing infrastructure are particularly substantial in railway systems, where the infrastructure represents around 40-60% of the total final service cost (Gomez-Ibanez, 2003), creating the potential for sizable savings.

However, the analysis and comparison of alternative capacity pricing and allocation mechanisms is particularly complex in railway systems, because there are strong interactions between capacity planning and infrastructure operations (Krueger et al., 1999; Pouryousef and Lautala, 2015).

The objective of this thesis is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems according to several performance metrics. The main contribution of this thesis is the development of a framework to systematically analyze and compare alternative capacity pricing and allocation mechanisms. The results of this research are expected to be valuable for railway regulators, but also to infrastructure managers (IMs) and train operators (TOs), allowing them to better understand the implications of these mechanisms at the system level and to better plan shared-use railway systems by the implementation of appropriate capacity pricing and allocation mechanisms. This understanding will be valuable for other shared infrastructure systems as well.

1.1 Shared Railway Systems: Promises and Challenges

Recently, governments have started promoting the use of shared railway systems. Up until 1988, all major railways both managed the infrastructure and operated the trains, i.e., they were vertically integrated (Drew, 2006). In contrast, in shared railway systems, multiple TOs utilize the same infrastructure, i.e., there is some level of vertical separation between infrastructure management and train operations. Examples of shared railway systems are the Northeast Corridor (NEC) in the US and the railway system in most European countries. Several countries in Asia and Africa are also opening access to their railway systems.

Proponents of shared railway systems stress that the use of shared railway systems allows 1) a more efficient use of expensive railway infrastructure and 2) the introduction of competition. Achieving a more efficient use of current infrastructure is positive not only in cases when resources to invest in infrastructure are limited, but also in cases where additional deployment of infrastructure is simply not possible. Recovery of infrastructure investment is one of the main reasons behind the implementation of open-access in Tanzania (Pena-Alcaraz et al., 2014; World Bank, 2014). The difficulties in adding additional capacity, especially near the densely populated area of Penn Station in New York City, are one of the main reasons why multiple TOs share existing railway infrastructure in the NEC (Gardner, 2013). According to (Gomez-Ibanez, 2003), rail infrastructure is a natural monopoly but the train operations business is not. As a result, with shared use and open access, new competitors would be able to enter the train operations business with its consequent benefits for the end users. This is the main rationale behind the European Union railway packages (Perennes, 2014).

As mentioned above, however, shared railway systems can only provide these benefits when there is a strong coordination between the IM and the TOs (Gomez-Ibanez, 2003). Such coordination, in turn, requires capacity planning mechanisms that determines which trains can access the infrastructure at each time, capacity allocation, and the access price they need to pay, capacity pricing (Pena-Alcaraz, 2015). It is important to maintain transparency when the IM is also one of the TOs.

There is a broad literature that has explored various capacity pricing and allocation mechanisms for railways (Affuso, 2003; Crozet, 2004; Gibson, 2003; Nash, 2005; Perennes, 2014). In general, different countries have promoted different

mechanisms for capacity pricing and allocation, with differing objectives. Studies have tended to focus on one mechanism and evaluated it according to performance metrics unique to that mechanism, making the comparison across different mechanisms to price and allocate railway capacity quite difficult and the implications for other systems ambiguous.

The objective of this research is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems according to several performance metrics. We consider multiple criteria to analyze the performance of capacity pricing and allocation mechanisms from the perspective of the IM (cost recovery, capacity utilization), the TOs (access charges, trains scheduled), and the end users (number of services, passenger fares or freight shipping rates). This thesis hypothesizes that alternative capacity pricing and allocation mechanisms would perform well for some metrics, but there is no silver-bullet mechanism that would perform well in all the metrics for every shared railway system. A better understanding of these trade-offs in performance is of particular importance today (Drew and Nash, 2011; Nash, 2010), since several countries are restructuring their railway sector to allow shared use. This understanding would allow the regulators of each country to design the most appropriate capacity pricing and allocation mechanisms to unlock the benefits of shared use in their railway system.

1.2 Literature Review

As noted earlier, the objective of this thesis is to analyze and compare alternative mechanisms for capacity pricing and allocation in shared railway systems. This section presents an overview of the main capacity pricing and allocation mechanisms proposed in

the literature, and discusses the experiences in shared railway systems for different countries. This section also summarizes some lessons from other network industries that are then used as guiding principles in this research.

1.2.1   Capacity Pricing and Allocation Mechanisms

There are three main types of mechanisms to price and allocate capacity: negotiation-based, administrative-based, and market-based mechanisms. Under negotiation-based mechanisms, the TOs and the IM negotiate to determine which trains can access the infrastructure and at what price. The main drawback of negotiations is that they can be very complex and time consuming (Nash, 2003). In addition, they often result in non-transparent bilateral contracts that prevent adaptation to future needs or create barriers to new operators. Under administrative-based mechanisms, the regulator establishes access rules and oversees the capacity pricing and allocation process. The regulator punishes (e.g. fines) any deviation from the rules. The use of these types of mechanisms relies on the ability of the regulator to gather information from the TOs and the IM to eliminate information asymmetries. These mechanisms are also slow to adapt to new system needs.

The shortcomings of negotiation-based and administrative-based mechanisms, together with the need for transparency and non-discriminatory access have motivated the introduction of market-based mechanisms for capacity pricing and allocation. In the NEC, for instance, with current bilateral infrastructure access contracts, 1) the price that each TO pays to access the infrastructure depends on the time at which the contract was signed (companies who signed their agreements when there was still plenty of excess capacity are paying much less than other companies to operate the same type of services),

2) access charges and slots are rigid (none of the companies want to lose their current slots and it is difficult to make room for new trains because multiple contracts would need to be renegotiated), and 3) the IM is not able to raise enough revenues to afford basic maintenance of the lines (this has contributed to the current backlog in maintenance in the NEC) (Gardner, 2013). As a result, the Federal Railroad Administration (FRA) requires all the railroads to agree on a market-based mechanism for pricing and allocating capacity to substitute for the current negotiation-based mechanism (PRIIA, 2008). According to (Gibson, 2003), there are two main types of market-based mechanisms for capacity pricing and allocation: 1) price-based and 2) capacity-based.

Price-based mechanisms are those that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not. Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which train to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges). An example of a price-based mechanism would be a cost-allocation mechanism that assigns infrastructure-related cost proportionally to the volume of infrastructure use (Crozet, 2004; Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014). The access charge could also be adjusted considering the TOs' demand for scheduling trains (e.g. introducing congestion prices). These charges could also be adjusted with a base tariff that allows the IM to recover infrastructure costs that are fixed in nature.

Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity, e.g. an auction (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes,

2014; Stern and Turvey, 2003). There are multiple types of auctions: simple auctions in which TOs bid to get some predefined slots (either in a segment of the infrastructure or for the full path) or submit their desired timetable when they bid, and combinatorial auctions where the TOs' bid depends on the result of the auction. Capacity-based mechanisms have been widely studied in the literature but have not yet been implemented on the railway system in any country.

1.2.2   *International Context*

As Table 1-1 shows, shared railway systems are not an isolated phenomenon in the NEC. Starting in 1991 several countries have started opening access to their railway systems. However, different countries have adopted very different mechanisms to price and allocate railway capacity. The US uses negotiation-based, Australia and India administrative-based and European Union countries market-based mechanisms to allocate capacity.

Capacity pricing also varies from country to country. Although most IMs charge TOs the marginal cost of operating the train on the infrastructure (Nash, 2003), the calculation of the marginal cost of operating one more train on the infrastructure is based on several assumptions. As a result, (Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014) conclude that charging mechanisms in shared railway corridors are getting more heterogeneous.

Furthermore, different countries design capacity planning mechanisms with different objectives and evaluate those using different metrics. As a result, the comparative performance of different mechanisms is still unclear (Drew and Nash, 2011; Nash, 2010). According to Nash (2003), "it is important to recognize that the concept of

multiple operators may be relatively new for railroads: This means that the institutional framework has not been developed, and the intellectual understanding may not be in place, to facilitate planning and operating the shared-use system." The authors warn against moving ahead quickly with the design of pricing and allocation mechanisms before understanding the implications of such mechanisms for all stakeholders.

Table 1-1 Railway systems international organization around the world (Source: author, based on (ADB, 2014; Agosta, 2015; Gomez-Ibanez and de Rus, 2006; Levy, 2015; Olievschi, 2013; Pozzo di Borgo, 2005; Sakamoto, 2012; Texeira and Prodan, 2014; The Economist, 2015))

Area	Railway System Characteristics
Africa	Seven African countries present vertically-integrated railway systems (Algeria, Botswana, Egypt, Morocco, Namibia, South Africa, and Tunisia). In the rest of Africa, most railway systems were concessioned between mid-1990s and 2010 (e.g. Burkina Faso, Cameroon, Ivory Coast, Mozambique, Senegal, Tanzania, and Togo). Starting in 2010, there has been a promotion of open-access and shared corridors, especially in those countries were concessions failed (e.g. Guinea and Tanzania)
North and South America	Canada and the US present vertical integration in their freight railway system. In both countries, private freight operators have to accommodate passenger operators on the tracks. Other countries like Cuba and Honduras also have a vertically-integrated railway system. Argentina has recently announced that the railway system will be nationalized and vertically integrated by the end of 2015. The railway system is currently vertically separated and concessioned in most countries in Latin America (e.g. Argentina, Brazil, Chile, Colombia, and Mexico). In the US there is an important shared railway system, the NEC. A similar shared system has been proposed now in California (blended system) to accommodate high-speed rail (HSR) and commuter services.
Asia	The railway system is vertically integrated in countries like China, India, Indonesia, Japan, Malaysia, etc. Starting in 2013, Indian railways changed the regulation to allow for open access and shared use in the new dedicated freight corridor. There is also open-access in Russia and Kazakhstan (freight cars only). Mongolia and Uzbekistan are also implementing open-access railway policies.
Europe	Most European countries have implemented open-access and shared railway use for freight following the EU first railway package of 1991. The passenger railway system in Europe is also vertically separated in most countries, also moving to open-access and shared use. By 2014 only Italy already had competition on the tracks, with two competing HSR companies offering services, although other countries like Romania, Spain, and Ukraine are also moving towards the introduction of competition.
Oceania	The railway system in New Zealand was renationalized and vertically integrated in 2008. The railway system in Australia is vertically separated and operates with open-access and shared use policies.

The power sector, the telecommunication sector, and the aviation industry have made significant progress sharing infrastructure as compared to the railway industry. The experience in these networks show that: 1) marginal access pricing have significant advantages over other capacity pricing mechanisms in most circumstances, 2) capacity pricing and allocation are complementary problems, and 3) price-based and capacity-based mechanisms often yield equivalent results. First, marginal access pricing ensures that vertically-separated agents' operational decisions match the decisions of an integrated company. In addition, (Perez-Arriaga and Olmos, 2009; Perez-Arriaga, 2013; Rubio, 1999) show that marginal access pricing allows the IM to recover fixed infrastructure costs when there is no lumpiness in infrastructure investment, no uncertainty, no information asymmetry, and when the operators do not have market power. Note that those conditions never occur in reality. However, the economic results are still useful when there are small deviations from these conditions (e.g. operators do not have strong market power). Second, capacity pricing and capacity allocation are two sides of the same coin in shared corridors. For instance, if access charges are low, the demand may exceed the available capacity, as occurred in US airports after the Airline Deregulation Act of 1978 (Vaze, 2011). In these cases, a capacity allocation mechanism is needed. Conversely, the number and type of slots available affect the operators' willingness to pay to access the infrastructure (Laffont and Tirole, 1993; Laffont and Tirole, 2000; Vazquez, 2003). Third, there is no difference between price-based and capacity-based mechanisms to price and allocate capacity if there is perfect information and no uncertainty according to (Weitzman, 1974). However, under imperfect

information and uncertainty the specific design influences the performance of the system (Czerny, 2010).

As mentioned before, however, there is an important difference between the railway industry and other network industries. In shared railway systems, capacity planning at the strategic level is tightly coupled with infrastructure operations at the tactical level (Krueger et al., 1999; Pouryousef and Lautala, 2015). The understanding of the implications of pricing and allocation mechanisms thus requires 1) determining the TOs' demand for scheduling trains and 2) designing the optimal train timetable to determine whether the TOs' demand for scheduling trains can be accommodated on the infrastructure. Consequently, it is difficult to predict whether the conclusions from other network industries will apply in this case. This thesis addresses these questions and uses the lessons from other network industries as guiding principle for the research, when appropriate.

1.3 Thesis Contributions

The objective of this thesis is to contribute to the field of shared railway systems coordination by 1) developing a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms, and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in representative shared railway systems.

There are strong interactions between capacity planning and infrastructure operations in the railway industry; the operations on the infrastructure determine the available capacity in the system. The framework developed in this thesis to evaluate the performance of shared railway systems under alternative capacity pricing and allocation

consists of two models: 1) a financial model that anticipates how TOs would respond to the capacity pricing and allocation mechanisms and determine their demand for infrastructure use, and 2) a network optimization model that determines the optimal train timetable (IM's decisions) that accommodates the TO demand for scheduling trains, considering the topology of the line, safety constraints, and other technical aspects of the infrastructure. We use this framework to analyze the trade-offs associated with the use of alternative mechanisms in the context of the national rail system for Tanzania, which is implementing a new open-access model, and in the context of the NEC in the US, where the FRA is now trying to develop a new capacity pricing and allocation mechanism.

The main contributions of this thesis are:

1. Formulation of a TO Model to anticipate the response of TOs to capacity pricing.

2. Formulation of the train timetabling problem (IM Model) for capacity allocation in shared railway systems to model interactions between capacity planning and infrastructure operations.

3. Development of an approximate dynamic programming algorithm based on linear programming to be able to solve the train timetabling problem in relevant instances.

4. Identification of cases in which traditional capacity pricing mechanisms result in sub-optimal operational decisions, with an illustration of this problem in case studies based on the Central Corridor in Tanzania and the Northeast Corridor in the US.

5. Comparison of two representative capacity pricing and allocation mechanisms in those systems, concluding that there are important trade-offs between them.

The first contribution of this thesis is the formulation of a financial model to capture how rational TOs would behave under a capacity pricing and allocation mechanism, given the institutional and regulatory framework and the technical characteristics of the specific type of railway service that they provide (see Chapter 2). The model analyzes three main operational decisions: 1) the TO's demand for scheduling trains, 2) the TO's willingness to pay to access the infrastructure, and 3) the passenger fare or freight shipping rate they would charge to end users. The first two operational decisions are then used as inputs for the IM Model. The model proposed is simple by design. The main objective is to allow the regulator and the IM to anticipate the response of the TOs to alternative capacity pricing and allocation mechanisms. More detailed models would rely on extensive information about the TOs that is not readily available to the regulator or the IM (Levy et al., 2015).

The results obtained in Chapter 2 show that while the estimates of the fares charged to end users are sensitive to the demand curve and elasticity assumptions; the TOs' willingness to pay for access estimates are robust to model assumptions. This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. This model analyzes each TO independently of other TOs. Once we know all of the TOs' demands for scheduling trains, we need to determine if there is capacity available to schedule all the services.

The second contribution of this thesis is the formulation of a train timetabling model for shared railway systems. The train timetabling problem has been widely studied

in the literature (Cacchiani et al., 2010; Caimi et al., 2009; Caimi et al., 2011; Caprara et al., 2002; Caprara et al., 2011; Castillo et al., 2009; Cordeau et al., 1998; Ghoseiri et al., 2004; Liebchen, 2008; Liebchen and Peeters, 2009; Pena-Alcaraz et al., 2011; Zhou and Zhong, 2005). However, we argue in Chapter 3 that traditional train timetabling models cannot be used to analyze capacity planning mechanisms in shared railway systems for three reasons. First, most models assume a fixed number of trains to be scheduled on the infrastructure. However, the number of trains to schedule is the main decision variable of the capacity allocation problem in shared railway systems. Second, most approaches also assume a single TO that tries to schedule trains. This TO could iteratively solve the train timetabling problem, introducing small modifications in each train's desired timetable until the resulting timetable meets its needs. In shared railway systems, in contrast, multiple TOs request access to the infrastructure and the TO is informed afterwards whether its train can be scheduled or not. We argue that consequently, TOs have incentives to be flexible in shared railway systems to ensure that most of their trains get scheduled in the first iteration. Third, most of these models assume that all trains follow the same path. Again, this assumption does not hold when the nature of the services operated in the shared railway system is different. For example, commuter services are typically scheduled around metropolitan areas, whereas intercity and freight operators offer services between cities.

Chapter 3 of this thesis presents a train timetabling problem formulation for shared railway systems that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. This approach 1) introduces a discrete variable that indicates whether a train can be scheduled or not, 2) uses flexibility margins

31

to ease conflicts, making fast trains travel slowly when there are slow trains ahead and making slow trains wait at sidetracks when fast trains overtake them, and 3) specifies safety constraints (spacing of the trains) for each train path.

These three additional considerations make the train timetabling problem for shared railway corridors very difficult to solve. From a computational standpoint, the size of the model increases rapidly (more than linear) with the number of stations and the number of trains to schedule. As a result, commercial solvers are only able to solve the problem for a relatively small number of trains. Furthermore, most of the techniques developed in the train timetabling literature are designed for traditional single-operator train timetabling problems and cannot be used in this case. Most classical decomposition approaches do not work because of the large number of discrete variables needed to specify which trains are scheduled and the order in which trains go through each station. Any technique that exogenously fixes train order cannot be used here because of the large flexibility margins and because train spacing constraints are specific to each individual train.

As a result, this thesis proposes an alternative class of solution algorithms using approximate dynamic programming techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007) to be able to solve the problem in meaningful instances. The third contribution of this thesis is the development of a novel Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends previous algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). This algorithm allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap (Pena-

Alcaraz et al., 2015a). This is the main contribution of this thesis because this algorithm makes it possible to analyze price-based and capacity-based mechanisms considering interactions between infrastructure operations and capacity available.

The fourth contribution of this thesis spans Chapters 2, 4, and 5. Chapter 2 compares the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs under different institutional and regulatory environments. It shows that both vertically-separated perfectly-competitive and vertically-separated regulated monopolistic TOs would make the same operational decisions as an integrated railway company when variable access charges reflect variable IM costs to operate trains on the infrastructure. Not surprisingly, the capacity pricing literature (Lopez-Pita, 2014; Nash, 2003; Texeira and Prodan, 2014) recommends the use of access charges equal to the infrastructure marginal costs of operating the train. However, Chapter 2 also shows that this approach cannot be used in two cases: 1) when the IM needs to recover part of the infrastructure management fixed costs, or 2) when the railway system is congested (overcrowded or overloaded with traffic). Most railway systems fall into at least one of these two categories.

We illustrate the first case in Chapter 4 in the case of the Central Corridor in Tanzania. The Central Corridor goes from the port (Dar es Salaam) to an inland container terminal (Isaka) that serves as a dry port for Rwanda, Burundi, Uganda, and the Eastern portion of Democratic Republic of Congo. The infrastructure is owned by RAHCO, a publicly owned company. TRL is the only current TO; it operates around six trains per week. Although the corridor is single track, there is plenty of spare capacity that could be used by multiple private companies that have expressed interest in starting operating new

33

services between Dar es Salaam and Isaka (Pena-Alcaraz et al., 2014; World Bank, 2014). Due to the low number of trains that operate in the system, the infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. As a result, maintenances costs are assumed fixed.

If access charges are set following the traditional approach (access charges equal to the variable costs of managing and maintaining infrastructure), TOs would pay zero (0) access charges to access the infrastructure and the IM would not collect any revenues. However, it is critical to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational. As a result, the IM has to allocate infrastructure fixed costs among TOs through the access charges. Chapter 4 shows that the introduction of (non-zero, in this case) variable access charges distorts the operational decision of TOs. Chapter 4 then discusses how to avoid this problem with other price-based mechanisms such as the introduction of fixed access charges and how to allocate infrastructure costs among different types of TOs. Chapter 4 also analyzes the potentials of capacity-based mechanisms and shows that they would not allow the IM to recover infrastructure costs. Because the Central Corridor is currently not congested, capacity pricing and allocation can be solved independently. In other words, it is easy to analyze each TO independently because there is enough spare capacity to schedule all the trains.

Chapter 5 discusses how to analyze alternative capacity pricing and allocation mechanisms in the context of the NEC, a very congested shared railway system. In this case, we need both the TO Model to anticipate how each TO will respond to the mechanism and the IM Model to determine the final allocation of infrastructure capacity. The main spine of the NEC stretches from Boston, MA to Washington, DC. This segment

34

is shared by several passenger and freight TOs: an intercity passenger TO that operates around 150 trains per day, eight commuter TOs that operate over 2,000 trains per day, and several freight TOs that operate around 70 trains per day. Until now, capacity pricing and allocation in the corridor has been managed via bi-lateral contracts negotiated between the IM and the TOs. However, the limitations of this negotiation-based mechanism motivated the FRA's request to Amtrak and the rest of the commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism by the end of 2015 (PRIIA, 2008). Chapter 5 analyzes the performance of the system under two proposed capacity pricing and allocation mechanisms: a price-based (cost-allocation and priority-rule) mechanism proposed by Amtrak (Gardner, 2013; NEC Commission, 2014) and a capacity-based (auction) mechanism (Affuso, 2003; Perennes, 2014).

The results of Chapter 5 show that the capacity-based mechanism considered could result in almost 20% more IM cost recovery and 20% more trains scheduled as compared to the price-based mechanism considered in the NEC. The price-based mechanism, on the other hand, ensures higher profits for the TOs, making them more resilient to uncertainty in end users' transportation demand. However, this mechanism is not very resilient to uncertainty in infrastructure capacity availability. Under the capacity-based mechanism, NEC intercity TOs are in a better position than commuter TOs to access the tracks with current levels of service. The priority level of each TO is a design choice in price-based mechanisms. This choice, however, has important implications for NEC commuter and intercity passengers and TOs, especially if the IM does not have access to sophisticated methods to solve the train timetabling problem. To our

knowledge, this is the first effort to compare a price-based mechanism and a capacity-based mechanism in the same shared railway system. The results show that there are very important trade-offs among them and that none of them is superior to the other in all dimensions.

1.4   Thesis Organization

The main body of this thesis consists of five chapters: Chapter 2 presents the TO Model used 1) to compare the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs considering the institutional and regulatory environments and the technical characteristics of the service, and 2) to anticipate the TO response to alternative capacity pricing and allocation mechanisms.

Chapter 3 describes the IM Model. It presents first the formulation used to determine the optimal train timetable in shared railway systems. This model considers TOs' demand for scheduling trains and determines the optimal train timetable; i.e., the optimal IM decision regarding which trains to schedule and their timetable. Chapter 3 then describes the algorithm proposed to solve the train timetabling problem in large-scale systems and analyzes the results obtained in different cases.

Chapter 4 uses the TO Model to analyze alternative capacity pricing mechanisms in the context of the Central Corridor in Tanzania. This chapter discusses the main policy implications of this research for shared railway systems where the IM has to assign fixed costs among different types of railway services.

Chapter 5 uses the TO and the IM Models to analyze and compare alternative capacity pricing and allocation mechanisms in the context of the NEC. This chapter

discusses the main policy implications of this research for railway systems in congested shared railway systems.

Chapter 6 concludes this work by summarizing our finding and the main conclusions of our research. It also discusses the main policy implications and future directions of this research.

We now begin with a discussion of the TO Model and the TO response to capacity pricing and allocation mechanisms within different institutional and regulatory environments.

Chapter 2 -   The Train Operator Problem: Determining Train Operator Response to Alternative Capacity Pricing and Allocation Mechanisms

"All models are wrong, but some are useful" – G.E.P. Box (1987)

The introduction of shared railway systems requires the design and implementation of capacity pricing and allocation mechanisms. Chapter 1 shows that there are several alternative mechanisms to price and allocate railway capacity. The objective of this thesis is to analyze and compare such mechanisms. The improved understanding of the implications of different capacity pricing and allocation mechanisms would allow the stakeholders of shared railway systems to design and choose the mechanism that best fits their system needs and overarching goals. This chapter argues that while there is an extensive literature focused on the relation between capacity pricing and infrastructure costs, the response of train operators (TOs) to capacity pricing and allocation mechanisms is still unclear. The first step to analyze capacity pricing and allocation mechanisms is thus to anticipate how TOs respond to capacity pricing.

This chapter proposes the use of a TO financial model (TO Model) to capture the interrelation between different TO's operational decisions. The two main contributions of this chapter are 1) to demonstrate that the TO Model responses are robust to a broad set of model inputs; and 2) to compare the operational decisions of an integrated railway company with the operational decisions of vertically-separated TOs considering the institutional and regulatory environments and the technical characteristics of the service. The operational decisions of integrated railway companies are then used as a benchmark

to identify necessary conditions under which profit maximizing TOs would make operational decisions that also maximizes social welfare. An early version of this work is accepted for publication in Transportation Research Record (Levy, Pena-Alcaraz, Prodan, and Sussman, 2015).

The rest of the chapter is structured as follows: Section 2.1 reviews the main studies that analyze how capacity pricing and allocation affects the performance of shared railway systems, summarizes the contributions of this chapter, and discusses the modeling assumptions. Section 2.2 presents the TO Model and introduces the cases in which the model will be discussed. Section 2.3 analyzes how the TOs' estimated response changes with any changes in the TO Model inputs. Section 2.4 then uses the TO Model to compare the responses of TOs in different cases. All the results of this chapter are illustrated with examples based on the Northeast Corridor (NEC) in the US. Section 2.5 concludes with some highlights and recommendations.

2.1   Capacity Pricing and Allocation in Shared Railway Systems

This section summarizes the literature on capacity pricing and allocation in shared railway systems, presents the main contributions of this chapter, and finishes presenting a discussion of the main modeling assumption.

2.1.1   Literature Review

This section summarizes the main findings of two different bodies of literature: one that studies capacity pricing in the railway industry from the perspective of the infrastructure manager (IM) and other that studies the financial performance of TOs.

The capacity pricing literature in the railway industry focuses on the potentials for infrastructure cost recovery (Nash, 2005; Texeira and Lopez-Pita, 2012; Lopez-Pita,

2014; Texeira and Prodan, 2014). However, there are two main challenges to being able to relate infrastructure costs with capacity utilization in the railway industry: the nature of railway infrastructure costs and the nature of railway capacity. First, (NEC Commission, 2014) shows that different types of infrastructure costs vary with different operational variables (trains, trains-miles, gross-ton-miles, frequency of service, etc.). Second, the available capacity in the railway industry depends on infrastructure operations (Krueger et al., 2009; Pouryousef and Lautala, 2015). To understand infrastructure operations we need to be able to anticipate TOs operational decisions. According to (Nash et al., 2004; Drew and Nash, 2011), the impact of capacity pricing and allocation mechanisms on the TO operational decisions are still inconclusive (Nash et al., 2004; Drew and Nash, 2011).

There is also a broad literature that describes the TO revenues and costs for different operational decisions (Belli et al., 2001; Martland, 2011; PPIAF et al., 2011). These financial models pay little attention to access charges, since the need to establish capacity pricing and allocation mechanisms on the railway systems is still relatively new (Nash, 2003). Moreover, these models are usually descriptive, and they are rarely used to estimate TO operational decisions. For these reasons, TO financial models have not yet been used to anticipate how TOs respond to different pricing mechanisms. This literature gap prevents us from understanding the implications of capacity pricing and allocation mechanisms for shared railway systems.

2.1.2 Chapter Contributions

The objective of this chapter is to help fill the identified literature gap by 1) developing a TO financial model (TO Model) that explicitly models the relation between TO operational decisions and access charges; 2) analyzing the sensitivity of the results of

the model to changes in the model inputs; and 3) proposing the use of the model to compare the behavior of TOs within different institutional and regulatory environments.

First of all, this chapter proposes a TO Model that discusses the TOs' response to alternative capacity pricing and allocation mechanisms as a function of the regulatory environment, the competitive landscape, and the characteristics of the type of railway service that each TO offers. The TOs' response is characterized by: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure. We call these variables *operational decisions*.

Second, the objective of the TO Model is to allow regulators and IMs to anticipate the TO demand for scheduling trains and their ability to pay to access the infrastructure. In that sense, the TO Model is designed to rely only on public information about the TOs that is already available to the regulators and IMs. We need to be sure, though, that the estimates obtained without detailed TO information are accurate. This chapter carries out sensitivity analyses and studies the results obtained with different model inputs. We show that the TO demand for scheduling trains and the TO ability to pay to access the infrastructure are very robust to model inputs. In other words, the TO demand for infrastructure use does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM.

Third, we use the TO Model to analyze the operational decisions of integrated railway companies. We use these results to compare those obtained within different institutional and regulatory environments. The results show that the operational decisions

of a profit-maximizing TO match the decisions of an even-handed integrated railway company when variable access charges reflect variable IM costs to operate trains on the infrastructure. This finding justifies the use of traditional mechanisms adopted in most countries to price and allocate capacity. However, the results also show that there are two cases in which these mechanisms cannot be used: 1) when the IM needs to recover part of the infrastructure management fixed costs; and 2) when the railway system is congested. Unfortunately, most railway systems fall into at least one of these two cases. This motivates the choice of the two cases studied later in this thesis: Chapter 4 studies the case of Tanzania where the IM needs to recover infrastructure costs and Chapter 5 studies the NEC, the most congested railway system in the US.

2.1.3 Model Assumptions

The TO Model proposed in this chapter assumes that: 1) TOs make operational decisions with the objective of maximizing profits; 2) each train serves a single origin-destination (OD) pair; and 3) different types of services are not substitutable. This section explains why we make these assumptions, how we use them, and how we expect these assumptions to affect the results.

The first assumption is necessary to determine TO operational decisions given the main revenue and cost streams. This is a standard assumption to replicate the decision process followed by private companies and is consistent with the privatization of TOs that has followed the implementation of many shared railway systems. This assumption allows us to determine the most likely TO operational decisions given revenues and costs.

The second assumption is used to compute the demand transported, relating the end users' demand for transport with the train capacity. We assume a single OD pair

because the data that TOs publish in their annual financial reports typically aggregates all the revenues obtained in the same corridor. As a result, we only have information about the fares and distance of an average trip. The impact of this assumption in the results depends on the nature of the services. In the NEC, where most trips occur between Boston and New York City, and New York City and Washington DC, this assumption leads us to underestimate the capacity of the trains. If most trips share a segment (as it does with trips from New York City to Philadelphia and New York City to Washington DC) this assumption would lead to overestimating train capacity. This bias could be corrected by (in order of increased complexity): adjusting the final number of trains in cases in which the capacity of the train is binding, adjusting the train capacity used as an input of the model, or including information of all OD pairs.

The third assumption is used to be able to solve the TO Model independently for different types of services. If the TOs provides partially substitutable services (e.g. the case Amtrak's high-speed service Acela and Amtrak's regional service in the NEC that serve the same OD pairs with different speeds) then the access charges, number of services, and fares of these services should be determined at the same time considering all the costs and revenues related to these services in the TO Model. Otherwise, the end user's demand may be overestimated.

Note again that we use the TO Model to anticipate the TOs' response to capacity pricing and allocation. In that sense, we are mostly interested in estimating the TO demand for scheduling trains and the TO ability to pay for access that we use as inputs of the IM Model in Chapter 3. Section 2.3 shows that these variables are very robust to model inputs. This finding suggests that the level of detail of the TO Model is adequate

for the purpose of this thesis. More sophisticated TO Models are necessary to address other research questions that require higher level of detail in the understanding of the TO operations or the TO–end user interactions.

2.2 Train Operator Model

The objective of this section is to discuss the operational decisions of rational TOs under a capacity pricing and allocation mechanism, given the institutional and regulatory framework and the technical characteristics of the specific type of railway service that they provide. As discussed above, there are three main operational decisions that TOs initially control: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure.

As was mentioned before, we assume that rational TOs are profit maximizing. In other words, they make operational decisions with the objective of maximizing profits. We also analyze whether TOs are medium term sustainable agents, i.e., whether they are able to ensure positive cash flows in the medium term. TO profits and cash flow can be determined by analyzing TO revenues and costs for a given number of trains. In the rest of the chapter we use capital letters to denote operational decisions (decision variables) and lower-case letters to denote model inputs (parameters):

N number of trains services that the TO would like to schedule.

There are three main types of costs that TOs face:

AC the cost of accessing the tracks or access charges if the TO schedules any trains. This cost often has a fixed and a variable component: $AC(N \neq 0) = AC_f +$

$AC_v \cdot N$.

fc fixed costs, such as the cost of buildings and the cost of purchasing cars and locomotives in the medium term.

vc variable costs of operating a train, such as fuel, personnel, train maintenance, and train lease, if trains are being leased.

TOs face fixed costs independently of any operational decision. These costs do not vary over the medium term. Variable costs depend on the number of trains operated. TOs know variable cost per train before they make operational decisions. Finally, TOs face access charges if they decide to schedule trains. The exact value of the access charges depend on the level of service. In general, TOs also know how much they will be charge as a function of their demand. This does not happen in the Netherlands, where the IM determines access charges once it knows all the trains scheduled (Texeira and Lopez-Pita, 2012). Although determining access charges is a good way to ensure that all infrastructure costs are recovered; this practice is not recommended neither implemented in most countries because it increases the uncertainty faced by TOs.

The two main sources of revenue come from transporting users (cargo or passenger) and from the government (subsidies). The revenues obtained from transporting users can be determined by multiplying the passenger fare or freight shipping rate by the demand transported. The demand transported is limited by either the capacity (reduced by a reasonable average loading factor) of the trains or by end users' demand. According to literature, end users' transportation demand depends fundamentally on the fare, the frequency of the service, and the travel time (Bebiano et al., 2014). While intercity passengers are typically more sensitive to the fare and the

travel time, commuter passengers are typically more sensitive to the fare and the frequency, and freight users tend to be sensitive to the fare. The frequency of the service is inversely proportional to the number of services when we assume a single OD pair and uniform services. Government subsidies depend in general on the demand transported too.

F passenger fare or freight shipping rate.

s government subsidies.

c capacity of the trains (maximum number of passengers or net tons).

d end users' demand for transportation.

tt travel time.

Summarizing, the costs and revenues of a TO can be determined using the following formulas:

$$cost(N, AC) = fc + vc \cdot N + AC(N) \tag{2.1}$$

$$revenues(F, N) = s(F, N) + F \cdot min(d(F, N, tt), N \cdot c) \tag{2.2}$$

Note that some of these variables may be pre-determined or conditioned by regulations. For instance, the fare of commuter services is typically set by the government. Likewise, access charges under price-based mechanisms are fixed inputs for TOs.

In order to characterize how TOs operate, we will use equations (2.1) and (2.2) to relate the TO operational decisions. For instance, given the access charges (AC), the TO demand for scheduling trains and the fares charged to the end users can be determined maximizing profits, $\max_{N,F}[revenues(F, N) - costs(N)]$:

$$\max_{N,F}[s(F, N) + F \cdot min(d(F, N, tt), c \cdot N) - fc - vc \cdot N - ac(N)] \tag{2.3}$$

Fixed costs do not depend directly on the number of trains operated or the fare. If the subsidy does not depend directly on the number of trains then equation (2.3) is equivalent to: $\max\limits_{N,F}[F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N - ac(N)]$.

Since the number of trains and the fares depend on the access charges, the access charges can be determined implicitly using sensitivity analysis. In general, the TO maximum willingness to pay to access the infrastructure can be determined considering that all variable costs should not exceed variable revenues (to ensure that the TO is interested in operating trains). The function that equation (2.3) maximizes depends on the number of trains. We know that if the TO decides not to operate any trains, the TO will not have any variable costs and it will not have to pay to access the infrastructure. Similarly, it will not receive any revenues from operations. However, in the medium term, the TO faces fixed costs fc independently of the decision of how many trains N to operate and may sometimes receive a subsidy $s(0)$ (typically $s(0) = 0$). As a result, the TO faces a profit of $s(0) - fc$ when it does not operate any train. As a result, the TO would never operate a number of trains that results on smaller profits than $s(0) - fc$, because it would be better off simply not operating any trains. In other words, the maximum of equation (2.3) can never be lower than $s(0) - fc$. As a result, the TO maximum willingness to pay to access the infrastructure as a function of F, N is presented in equation (2.4).

$$
\begin{aligned}
AC_v \cdot N &\le \partial s(F,N)\big/\partial N \cdot N + F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N \\
AC_f &\le \max(0, s(F,N) + F \cdot \min(d(F,N,tt), c \cdot N) - vc \cdot N - AC_v \cdot N)
\end{aligned}
\tag{2.4}
$$

The maximum willingness to pay to access the infrastructure for which the TO is sustainable in the medium term can be determined ensuring that the resulting cash flow is

positive, $revenues - costs(AC) \geq 0$. Although capital expenditures (CAPEX) and financing costs are also required to compute cash flows, we will initially assume that TOs have almost no CAPEX and negligible financing costs. As a result, the TOs willingness to pay to access the infrastructure as a function of the number of trains and the fares can be calculated using equation (2.5):

$$AC(n) \leq s(f,n) + f \cdot \min(d(f,n,tt), c \cdot n) - fc - vc \cdot n \qquad (2.5)$$

The implications of equations (2.3), (2.4), and (2.5) depend on the context in which TOs operate. The context is determined by both the institutional and regulatory environment, and the technical characteristics of the type of railway service that the TO provides.

2.2.1 Institutional and Regulatory Environment

There are three main institutional factors that we have to consider to study equations (2.3), (2.4), and (2.5): the vertical structure of the system, the regulation of the TOs, and the competitive landscape.

The vertical structure of the system determines whether the railway system is vertically integrated or vertically separated. Although capacity pricing and allocation mechanisms are critical when there is some level of vertical separation between the TO and the IM, we will also study the behavior of vertically-integrated railway systems as a basis of comparison for vertically-separated railway systems.

The regulatory environment determines how much control the TOs have over the three operational decisions: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated; and 3) the access charges paid to the infrastructure manager to access the infrastructure. This chapter distinguishes unregulated

and regulated systems. Within unregulated markets, the TO fully controls the first two decisions. The TO level of control over the access charges is affected by the capacity pricing and allocation mechanism in place. In regulated environments, a central planner controls the fares that end users pay or the rents that the TO extracts.

Since the capacity pricing and allocation mechanism also impacts the interactions between the TOs and the IM, the mechanism itself informs us about the most relevant interactions to study. For instance, determining the TO willingness to pay to access infrastructure as a function of the number of trains scheduled is particularly relevant when capacity-based mechanisms are used to allocate and price capacity. Likewise, determining the number of trains that the TO is willing to provide given the access charges is particularly relevant to design price-based mechanisms.

Finally, the competitive landscape also determines the response of the TOs to capacity pricing and allocation mechanisms. TOs will behave substantially different if they operate in monopolistic, oligopolistic, or a perfectly competitive market.

2.2.2   *Technical Characteristics*

The technical characteristics of the railway system determine the parameters of the model (cost, capacity of the trains). They also determine the nature of the demand. As was mentioned before, the end users' transportation demand depends fundamentally on the fare, the frequency of the service (inversely proportional to the number of trains), and the travel time (Bebiano et al., 2014). While intercity passengers are typically more sensitive to the fare and the travel time, commuter passengers are typically more sensitive to the fare and the frequency, and freight users tend to be sensitive to the fare. The literature proposes three functional forms to capture the dependency of the demand on

these factors: linear demand function, isoelastic demand function, and bounded isoelastic demand function. In the three cases, the elasticity determines the relation between changes in demand as other factor (x) changes: $e = \dfrac{\Delta d/d_0}{\Delta x/x_0}, e = \dfrac{\partial d}{\partial x}$.

Figure 2-1 shows a comparison of these functions in a case in which the demand depends only on the fare, for an elasticity value $e = -0.67$, and an initial demand $d_0 = 31{,}250$ for an initial fare $f_0 = \$96.5$. An isoelastic curve with these parameters indicates implies that there is an unlimited demand for the transportation service as the fares decrease. Similarly, it assumes that there are few passengers willing to pay extremely large fares. With these parameters, the isoelastic demand function indicates that there is at least one passenger willing to pay up to \$481 million to travel by train. In practice, isoelastic demand functions approximate the end users' demand well for intermediate values of fares; but it is unrealistic to assume that there are no end users' demand or fare willingness to pay bounds. This thesis thus analyzes only linear and bounded isoelastic curves to capture the nature of end users' transportation demand.

Considering the combinations of institutional and regulatory, and technical factors we propose four main cases of study:

1. Vertically-separated unregulated monopolistic train operator
2. Vertically-separated regulated monopolistic train operator
3. Vertically-separated perfectly competitive train operator
4. Vertically-integrated railway company
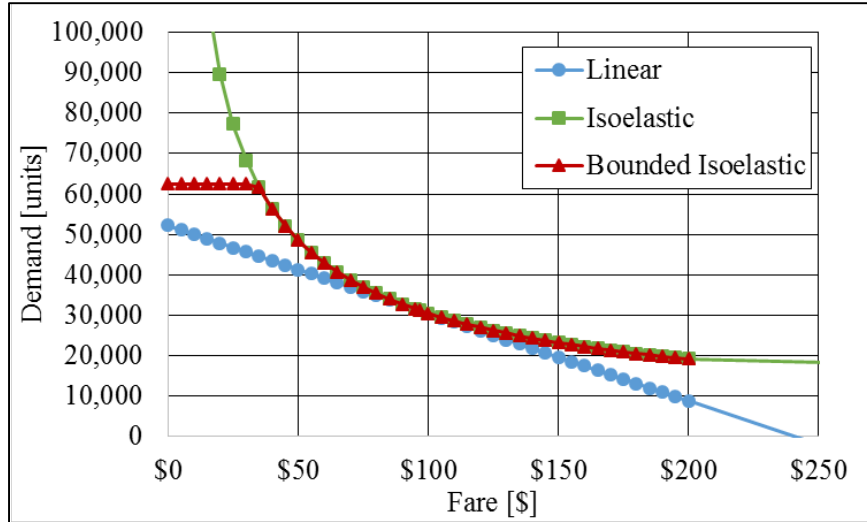
Figure 2-1 Comparison of linear, isoelastic, and bounded isoelastic demand functions (assuming demand only depends on fare) (Source: author)

2.3 Model Sensitivity to Inputs

This section analyzes the behavior of vertically-separated unregulated monopolistic intercity passenger TOs in the context of the NEC for different model inputs to understand the robustness of the results. We first compare the results obtained for different end users' demand functional forms and for different values of the elasticity. We then analyze the results obtained for different cost values.

We use data published by the TOs in their annual financial plan to determine the inputs of the model. According to (Amtrak, 2014) a TO like Amtrak faces fixed operational (direct) costs of $fc = \$102.5m$ per year ($fc = \$281k$ per day) and variable operational costs of $vc = \$1.25m$ per train and per year ($vc = \$3,425$ per train and per day). In 2013, Amtrak's average fare were equal to $f_0 = \$96.5$, the number of trains averaged $n = 150$ trains per day, with a realized demand of $d_0 = 11.4m$ passengers per year ($d_0 = 31,250$ passengers per day), and with an average train capacity of $c = 210$ passengers assuming a physical capacity of 250 seats with 85% load factor (Amtrak,

51

2014). No subsidies are required for the operations of intercity services in the NEC (Amtrak, 2010; Amtrak, 2012). We assume that end users' demand for traveling on a specific type of intercity service depends mostly on the fare (and not that much on the frequency or small variations on travel time). According to (Morrison, 1990) the elasticity of the demand of intercity passengers in the NEC to the fare is equal to $e = -0.67$. We also assume that the access charges depend linearly on the number of trains: $ac(n) = ac_f + ac_v \cdot n$.

2.3.1 Bounded Isoelastic Demand Function

In this case we assume that the end users' demand for rail services is a bounded: isoelastic curve that depends on the fare charged by the TO: $d(f) = \min(\bar{d}, kf^e), f \leq \bar{f}$. The value of k is equal to $k = d_0 f_0^{-e}$ for $f < \bar{f}$. $\bar{f}, \bar{d}$ are, respectively, the maximum fare and the maximum expected demand.

Equation (2.3) can be used to determine the number of trains that the TOs would like to operate and the fare charged to end users given the access charges. From an analytic standpoint, it is easier to determine those numbers assuming that the number of trains that the TOs would like to operate is continuous (equal to $n(f) = \frac{d(f)}{c}$) and that the demand is isoelastic. In that case, if $-1 < e \leq 0$, then the optimal fare f^{*c} is unbounded (the optimal solution is to charge end users as much as possible), and if $e \leq -1$, then the optimal fare is $f^{*c} = \frac{vc+ac_v}{c} \frac{e}{e+1}$.

In practice, though, the TOs can only operate an integer number of trains and the demand is bounded by $\bar{d}$. Considering this, the optimal number of trains that the TOs could operate is:

$$n^* = min\left(\left\lceil\frac{d(f^{*c})}{c}\right\rceil, \left\lceil\frac{\bar{d}}{c}\right\rceil\right) \quad if \quad \frac{vc+ac_v}{fc} \leq \frac{d(f^{*c})}{c} - \left\lfloor\frac{d(f^{*c})}{c}\right\rfloor$$

$$n^* = min\left(\left\lfloor\frac{d(f^{*c})}{c}\right\rfloor, \left\lfloor\frac{\bar{d}}{c}\right\rfloor\right) \quad if \quad \frac{d(f^{*c})}{c} - \left\lfloor\frac{d(f^{*c})}{c}\right\rfloor < \frac{vc+ac_v}{fc} \leq 1 \tag{2.6}$$

$$n^* = 0 \qquad\qquad if \qquad\qquad 1 < \frac{vc+ac_v}{fc}$$

Note that $\lceil x \rceil$ and $\lfloor x \rfloor$ represent the closest integer number over or under x respectively. The TO would decide whether to have some excess capacity or some unmet demand depending on the amount of exceeding capacity or unmet demand and the relation between variable costs and revenues. The TO will not operate any trains if variable costs are larger than the revenues. Considering this, the optimal fare (f^*) would be:

$$f^* = min(d_0^{-\frac{1}{e}}f_0, \bar{f}) \quad if \quad -1 < e \leq 0$$

$$f^* = min\left(\left(\frac{n^*c}{k}\right)^{\frac{1}{e}}, \bar{f}\right) \quad if \quad e \leq -1 \tag{2.7}$$

Equation (2.7) shows that for elasticity values between -1 and 0 the TOs are better-off increasing the fare as much as they can, i.e. up to $\bar{f}$ in this case.

Figure 2-2 shows the fare and number of trains that a TO would schedule as a function of the elasticity assuming $\bar{f} = \$200, \bar{d} = 62,500$ (double the current demand). In this case, the fares range from \$97 to \$200, and the number of trains from 68 to 129 trains per day when the IM does not charge any variable access charge per train. Note that both the fares and the number of trains to schedule are fairly robust when elasticity changes around $e = -0.67$: the fares change less than 1% and the number of trains change less than 10% for changes in elasticity of $\pm 20\%$.

Figure 2-2 Fare and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges, maximum fares of $200, and a cost and revenue structure similar to Amtrak (Source: author)
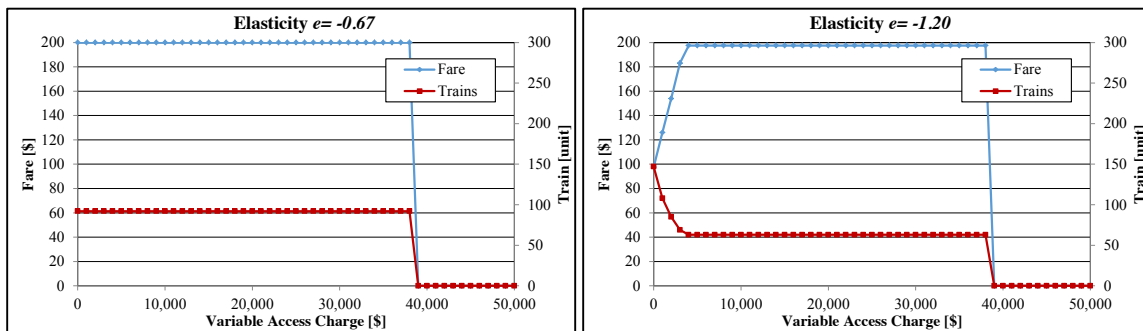
Figure 2-3 Fare and number of trains to be scheduled by a TO as a function of the variable access charges for different elasticity values assuming a cost and revenue structure similar to Amtrak (Source: author)

As Figure 2-3 shows, the number of trains that the TO operates is robust for a large range of access-charge values too. When the elasticity values $e \in (-1,0)$, the number of trains that the TO would operate (92 trains in this case for $e = -0.67$) does not change when access charges increase (unless the TO has no incentive to operate trains and would then operate 0 trains). For elasticity values $e < -1$, the number of trains would decrease as the variable access charges increase from $0 to $4,000 and does not

change afterwards (again, unless the TO has no incentive to operate trains and would then operate 0 trains).

We can use equations (2.4) and (2.5) to determine the maximum variable charge that a TO would be able to pay to access the infrastructure. This maximum access charge is equal to $ac_v(n = 0) = \$39,000$ independently of the elasticity. If access charges increase over that value the TO would not operate any trains (see Figures 2-3, 2-4, and 2-5). The maximum variable charges that TOs can sustainably pay in the medium term, i.e., the maximum variable access charges for which the TO has no profits neither losses after paying for capital at an adequate rate of return are $ac_v(\pi = 0) = ac_v(n = 0) - {fc}/{n^*}$, i.e., up to $\$35,000$ for an elasticity value of $e = -0.67$ (Figure 2-4) and up to $\$34,000$ for an elasticity value of $e = -1.2$ (Figure 2-5).
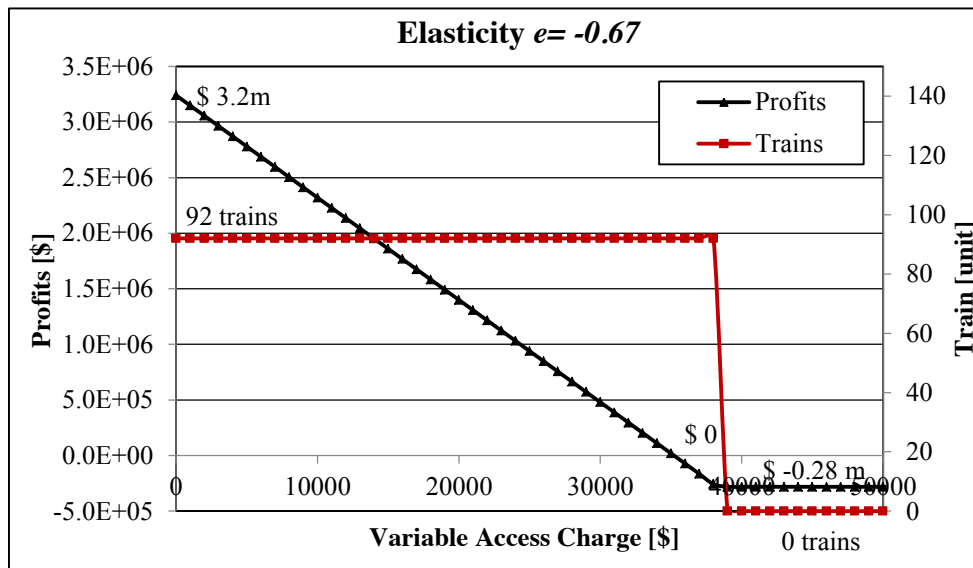
Figure 2-4 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 assuming a cost and revenue structure similar to Amtrak
(Source: author)
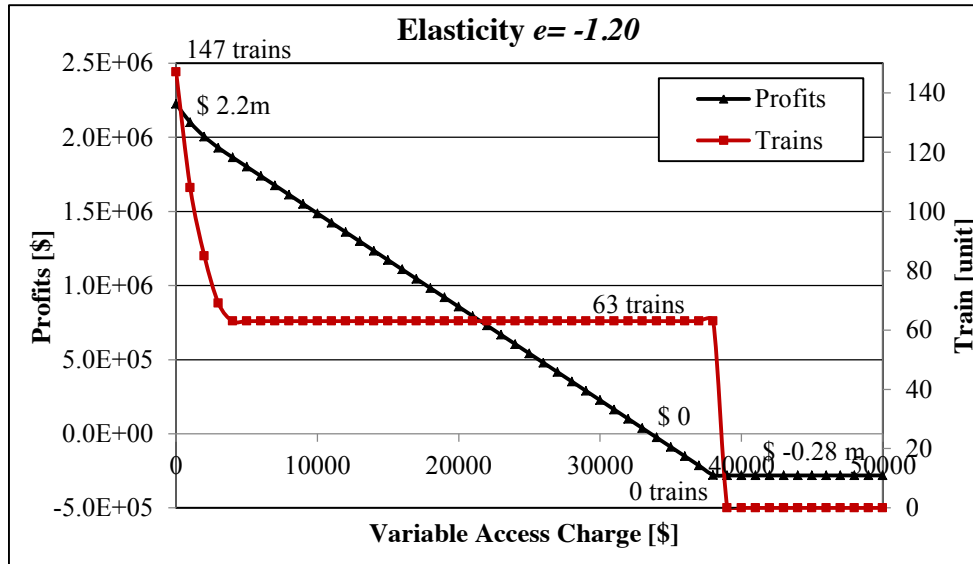
Figure 2-5 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -1.2 assuming a cost and revenue structure similar to Amtrak
(Source: author)

The first point labeled in Figures 2-4 and 2-5, represents the total profits obtained when the TO does not pay any access charges. This point represents the maximum possible profits that the TO could ever achieve in the market: $3.2m$ or $2.2m$ (depending on the elasticity value). This point also allows us to determine the maximum fixed access charges that the TO can afford. We know that a TO would never accept losses over $281,000 \approx \$0.3m$ (its fixed costs). As a result, the maximum fixed access charges that the TO can pay are respectively $3.5m$ ($3.2m + \$0.3m$) or $2.5m$ ($2.2m + \$0.3m$). Note that the TO can only be sustainable in the medium term if total fixed access charges are smaller than $3.2m$ or $2.2m$ respectively (so it can have $0 profits).

The second point labeled corresponds to the maximum variable access charges that the TO can sustainably afford: $35,000 per train per day or $34,000 per train per day respectively. At that point, the TO would operate 92 or 63 trains (depending on the

elasticity value assumed). The final point corresponds to the variable access charges ($39,000 per train per day in both cases) for which the TO would not have any incentives to operate trains (since operating a train would increase the burden of the debt).

If the demand increases year by year, the number of trains that operators want to schedule would increase by the same rate. However, the fares would not change if the elasticity of the demand to the fare is lower than -1.

2.3.2   Linear Demand Function

This case analyzes how the results presented in the previous cases would change when we assume that the demand is a linear function of the fare: $d(f) = e \cdot \frac{d_0}{f_0} \cdot f +$

$(1 - e) \cdot d_0$ (to ensure that elasticity is $e = \frac{\Delta d / d_0}{\Delta f / f_0}$).

The optimal number of trains and fare (n^*, f^*) to maximize profits can also be determined using equation (2.3). The results show that the number of trains and the fares to maximize profits are either:

$$n^* = \left\lceil \frac{(1-e)d_0}{2c} \right\rceil, f^* = \frac{(e-1)f_0}{2e},$$

$$n^* = \frac{(1-e)d_0}{2c} + \frac{ed_0(vc+ac_v)}{2c^2 f_0}, f^* = \frac{vc+ac_v}{2c} + \frac{(e-1)f_0}{2e}, \text{ or} \qquad (2.8)$$

$$n^* = 0, f^* = 0$$

Equation (2.8) implies that when variable costs are small with respect to the fares that end users can afford, the optimal solution is to maximize revenues and offer the minimum number of trains that allow serving all the demand for the optimal fare. However, when variable costs are comparable to the fares that end users can afford, the optimal solution is a trade-off between maximizing revenues and covering variable costs. In this case, the capacity should be optimized in such a way that most demand is served

without providing excess train capacity. Finally, in those cases in which the end users cannot viably accept a fare level that allows TOs to cover at least the variable costs, the TO should not operate any train.

We also illustrate these results with an example inspired in the Amtrak intercity services of the NEC. Figure 2-6 shows the fare and number of trains that a TO would schedule as a function of the elasticity. In this case, the fares range from $97 to $298, and the number of trains from 87 to 149 trains per day when the IM does not charge any variable access charge per train. Again, for changes in elasticity of $\pm15\%$, the number of trains changes less than 5%, and the fares change less than 10%. When we compare these numbers with the ones obtained in Section 2.3.1 we see that the numbers of trains in this case differ less than 20% with respect to the bounded isoelastic demand. The fares however vary up to 50% with respect to the ones obtained in the bounded isoelastic demand case.

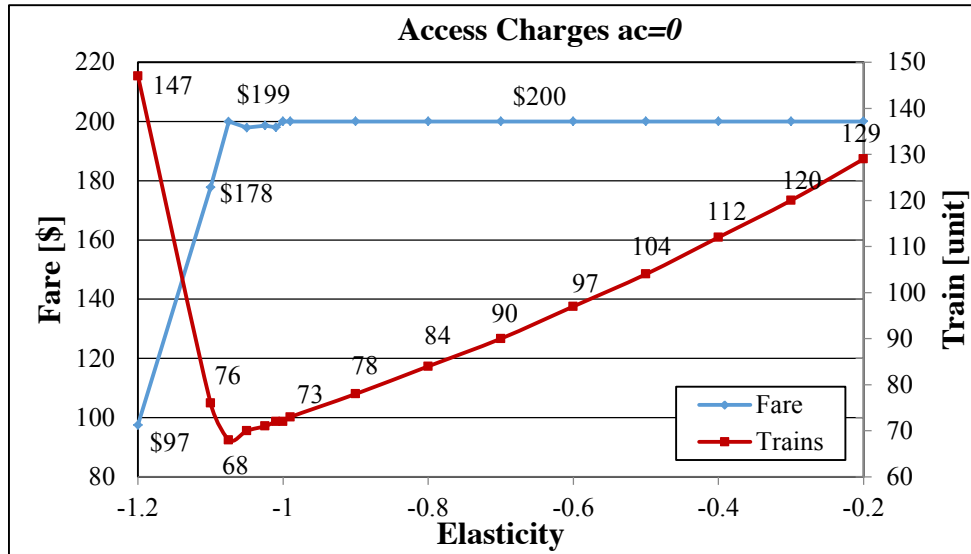**Figure 2-6 Fare and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges and a cost and revenue structure similar to Amtrak** (Source: author)

Figure 2-7 represents the expected profits of the TO and optimal number of trains for different values of the variable access charges. In this case, the optimal operational decision is to operate 116 trains per day with fares on the order of $128 if access charges are low. If the access charges increase over $32,000 per train then the profit maximizing strategy suggest operating between 37 and 0 trains. The TO would not operate any trains if the access charges are over $47,000 per train. The access charges willingness to pay is within 20% of the value obtained in Section 2.3.1. The number of trains is also within 20% of the values obtained in Section 2.3.1 for most access charges. The fares, however, vary considerably (over 35%) with respect to the bounded isoelastic demand case.

Figure 2-7 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 assuming a cost and revenue structure similar to Amtrak
(Source: author)

2.3.3   Cost Sensitivity Analysis

This case analyzes how the results presented in the previous cases would change for different fixed-cost and variable-cost values. The first finding in this section is that the number of trains and the fares do not change when fixed costs change. This result makes sense for two reasons. First, once the TO incurs in the fixed cost, fixed costs are

perceived as sunk costs and although they affect TO profit, they do not affect TO operational decisions. Second, and related to the previous point, fixed costs do not appear in equations (2.6), (2.7), and (2.8).

Our second finding is that the results are also very robust to changes in the variable access charges. As mentioned above, current Amtrak's variable costs are equal to $vc = \$3,425$ per train and per day (Amtrak, 2014). Figure 2-8 shows that the fares and the number of trains do not vary for different variable access charges unless access charges increase 1,000%. In that case, variable costs would be so high that the TO would prefer not to operate trains.
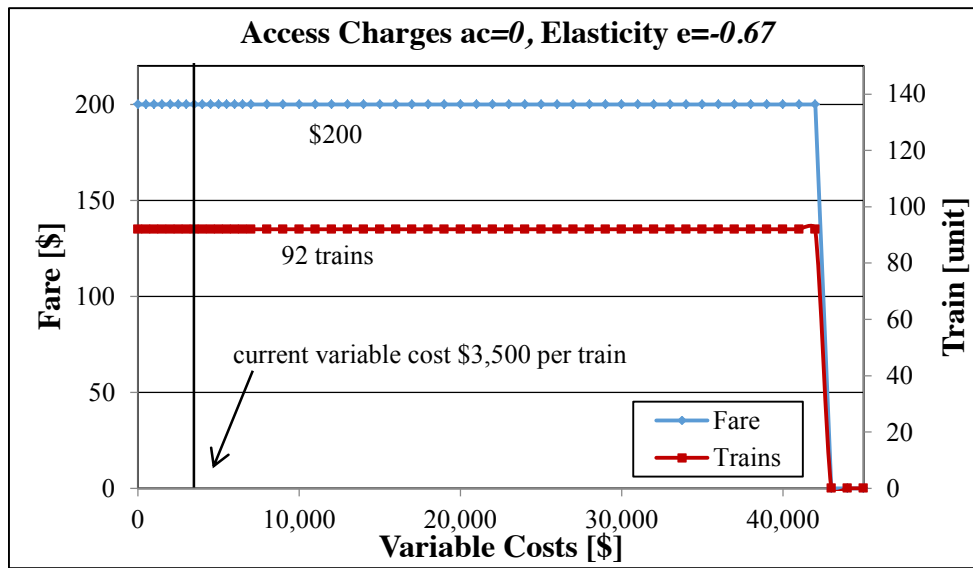
Figure 2-8 Fare and number of trains to be scheduled by a TO as a function of the variable costs assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

Figures 2-9 shows that the fares vary less than 8% and the number of trains vary less than 7% when variable costs change $\pm100\%$.
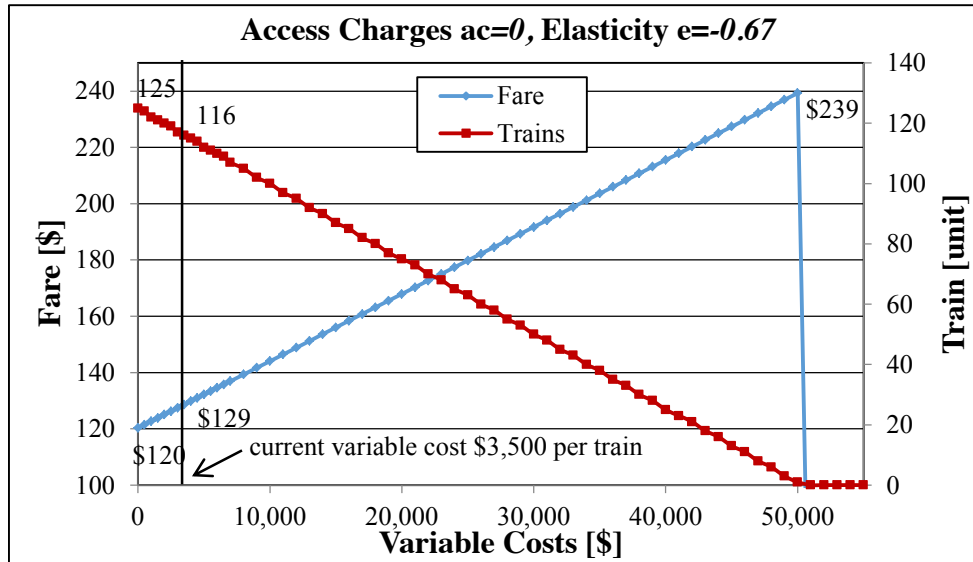
Figure 2-9 Fare and number of trains to be scheduled by a TO as a function of the variable costs assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

In summary, these results show that the TO demand for scheduling trains and the TO ability to pay to access the infrastructure are very robust to model inputs. In other words, the TO demand for infrastructure use does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. The results in Chapter 5 predicts differences in the number of services scheduled under alternative mechanisms that are over 20%. The robustness results shown in this chapter demonstrate that the differences obtained are significant (higher than the variances obtained in the sensitivity analysis).

The passenger fares or freight shipping rates charged to the end users are not as robust. As a result, more detailed information may be required to study in detail how TOs set fares and shipping rates. Although we use the fares to compute TO profits, this thesis does not focus on the interactions between TOs and end users.

2.4   Model Results and Implications

Section 2.3 shows that the TO Model estimates are robust to a broad set of model inputs. In this section, we use the TO Model to the behavior of TOs in different cases. We propose the study of TOs in both vertically-separated and vertically-integrated railway systems, within unregulated and regulated markets. Although each TO competes with other TOs to access the infrastructure, we distinguish whether each TO competes or not with other TOs to offer the same railway services to the end users (e.g., high-speed rail). Note that within the same vertically-separated railway system, different TOs providing different services may face different regulatory and competitive schemes. For instance, in shared systems like the NEC, one could argue that a private intercity passenger TO is best modeled as an unregulated monopoly, the commuter TOs are best modeled as regulated monopolies, and the freight TOs are best modeled as competitive TOs.

2.4.1   Vertically-Separated Unregulated Monopolistic Train Operator

This case assumes vertical separation between TO and the IM. It also assumes that although the TO may compete with other types of TOs to access the infrastructure, it is the only TO who provides a specific type of railway service to the end users. That is, the TO is a monopoly in its railway service market and there are no substitute services offered by other TOs. Unregulated refers to the fact that except for the constraints imposed by the capacity pricing and allocation mechanism in place, the TO has full control over the fares and the number of trains. Following the discussion that leads to equation (2.3), TOs in this case would determine its operational decisions with the objective of maximizing its profits given the access charges. Their willingness to pay to access the infrastructure is given by equations (2.4) and (2.5).

The two cases in this section analyze how an intercity passenger TO would respond to capacity pricing and allocation mechanisms. We assume here that end users' demand for traveling on a specific type of intercity service depends mostly on the fare (and not that much on the frequency or small variations on travel time). We refer the reader to Sections 2.3.1 and 2.3.2 for a discussion of the TO behavior when the demand is modeled as a bounded isoelastic function and a linear function respectively. Section 2.4.1.1 discusses the results obtained when the TO implement a revenue management mechanism. This case assumes that that the end users' demand is a bounded isoelastic function. Section 2.4.1.2 compares the results obtained in the three vertically-separated unregulated monopolistic intercity TO. We assume again that there are no operations subsidies and that the access charges depend linearly on the number of trains: $ac(n) = ac_f + ac_v \cdot n$.

2.4.1.1 Intercity Passenger TO using Revenue Management Mechanisms with Bounded Isoelastic Demand Function

This case proposes to study the behavior of TO assuming that the demand as a function of a fare behaves as a bounded isoelastic curve, that is, $d(f) = \min(\bar{d}, kf^e)$, for $f < \bar{f}$. This time we assume that the TO has a perfect revenue management mechanism in place that allows it to charge each end user the maximum fare that they are willing to pay.

In this case, the optimal number of trains (n^*) to maximize profits can be determined using equation (2.3). The results obtained show the number of trains to maximize profits is (depending on how the maximum fare compares to the variable costs):

$$n^* = k \cdot (vc + ac_v)^e \cdot c^{-1-e} \quad if \; \overline{f} \cdot c > vc + ac_v$$
$$n^* = k \cdot \overline{f}^e c^{-1} \; or \; n^* = 0 \qquad otherwise \tag{2.9}$$

If the result is not an integer, it should be rounded to the immediate lower or upper integer (the one that maximizes profits). Figures 2-10 and 2-11 show the number of trains and the TO profits for different elasticity values with no access charges, and the number of trains and the TO profits for different variable access charges and elasticity value equal to -0.67 using the same Amtrak-inspired example used in the previous cases.
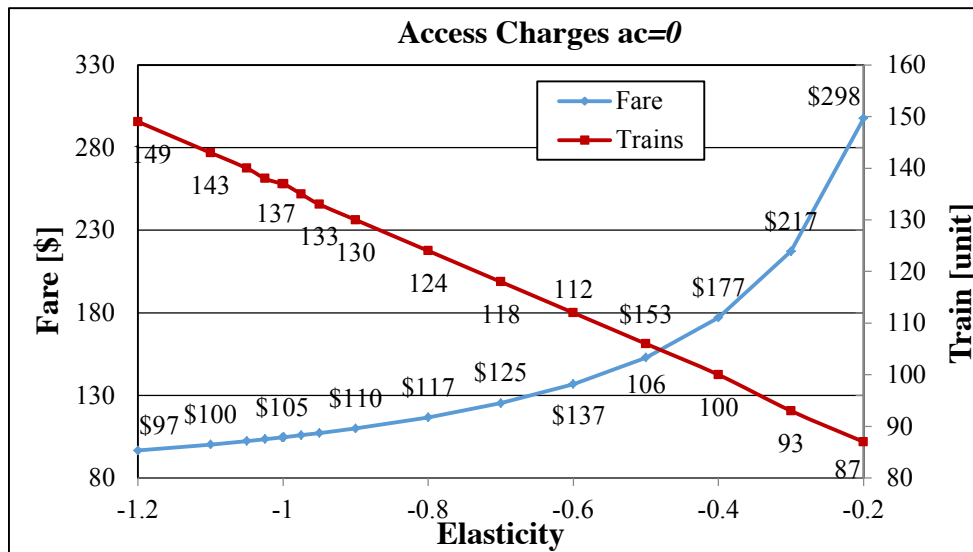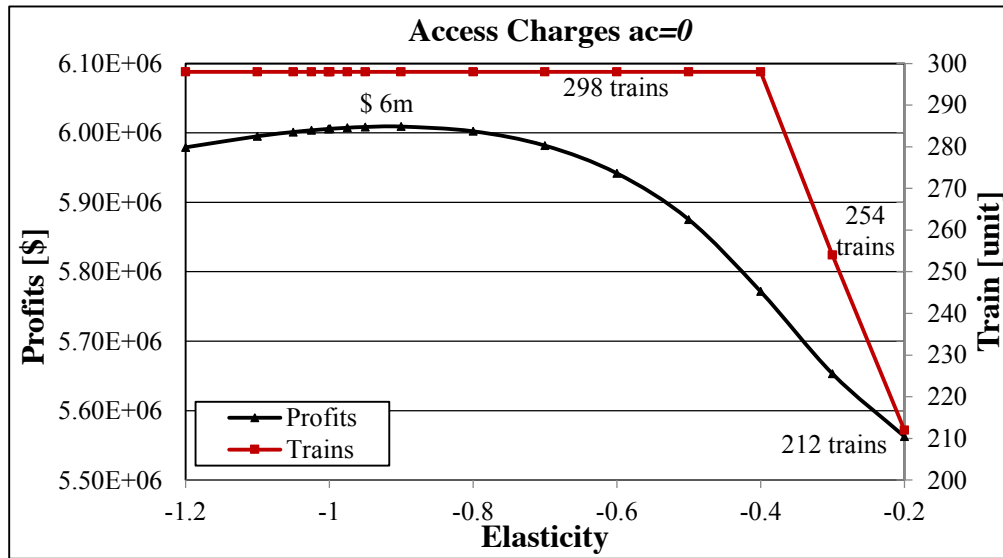

Figure 2-10 TO profits and number of trains to be scheduled by a TO as a function of the elasticity assuming 0 access charges and a cost and revenue structure similar to Amtrak (Source: author)

In this case, for elasticity values over -0.4, the maximum number of trains that the TOs are willing to operate when there are no access charges is driven by the maximum demand that TOs expect.

When the access charges increase, the operators have incentives to operate fewer trains. This case shows two important results: the maximum variable access charges that a TO like Amtrak (with elasticity value -0.67) would be able to pay are below $\$39k$ per

train. Variable access charges should not be above $35k$ to ensure sustainable TO operations in the medium term. With these level of variable access charges TOs would still operate 92 to 95 trains.
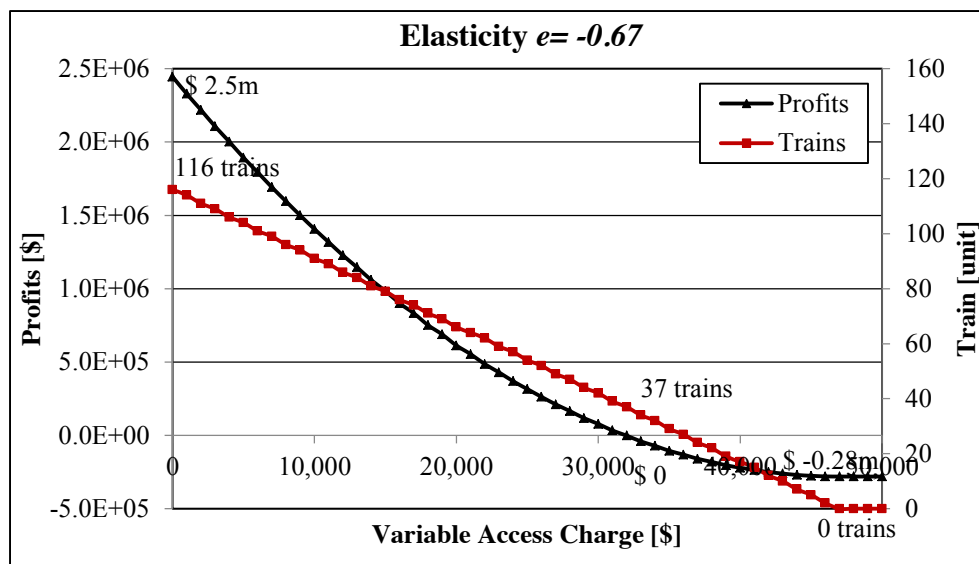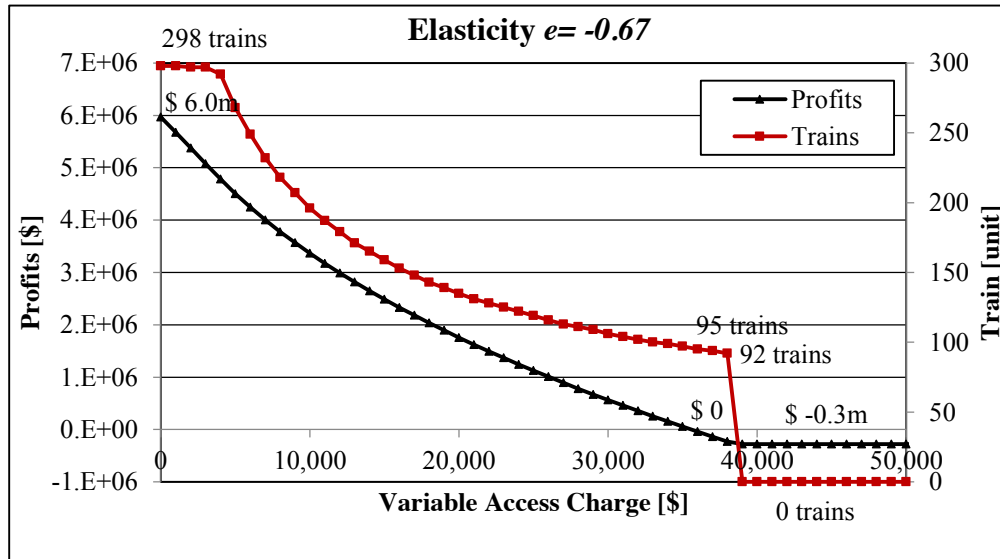

Figure 2-11 Profits and number of trains to be scheduled by a TO as a function of the variable access charges for elasticity equal to -0.67 assuming a cost and revenue structure similar to Amtrak
(Source: author)

2.4.1.2 Result Comparison

In this section we summarize and compare the main results obtained for the different cases of vertically-separated, unregulated, monopolistic intercity passenger TOs.

Figures 2-12 and 2-13 compare the number of trains that a TO like Amtrak would operate as a function of the variable access charges for elasticity values of -0.67 (very inelastic demand; as Morrison, 1990 characterized the NEC intercity passenger demand), and -1.2 (elastic demand). We also compare the optimal number of trains obtained with this TO model with the current number of trains operated by Amtrak in the corridor.

First, the results suggest that even if the TOs have a perfect revenue management mechanism in place, the number of trains operated would decrease with respect to the

current number of trains operated if variable access charges exceed $10,000 to $15,000 per train.

Second, the results show that for variable access charges above $35,000 per train, the number of trains would importantly decrease over 85% with respect to current number of trains independently of the demand functional form assumed and fare mechanism in place.

Third, the TO always operates more trains than in the case with revenue management as compared to the cases without revenue management. This makes sense since the utilization of revenue management mechanisms allows the TOs to charge different fares to end users. As a result, the TO has no need to reduce the number of trains to be able to increase fares.

Finally, note again that the TO ability to pay to access the infrastructure is very robust to different model assumptions. The value of the maximum access charges over which the TO would not operate any trains varies less than 20% across these cases. This is very important, because the operations of a TO that implements revenue management mechanisms is very different that the operations of a TO that does not.
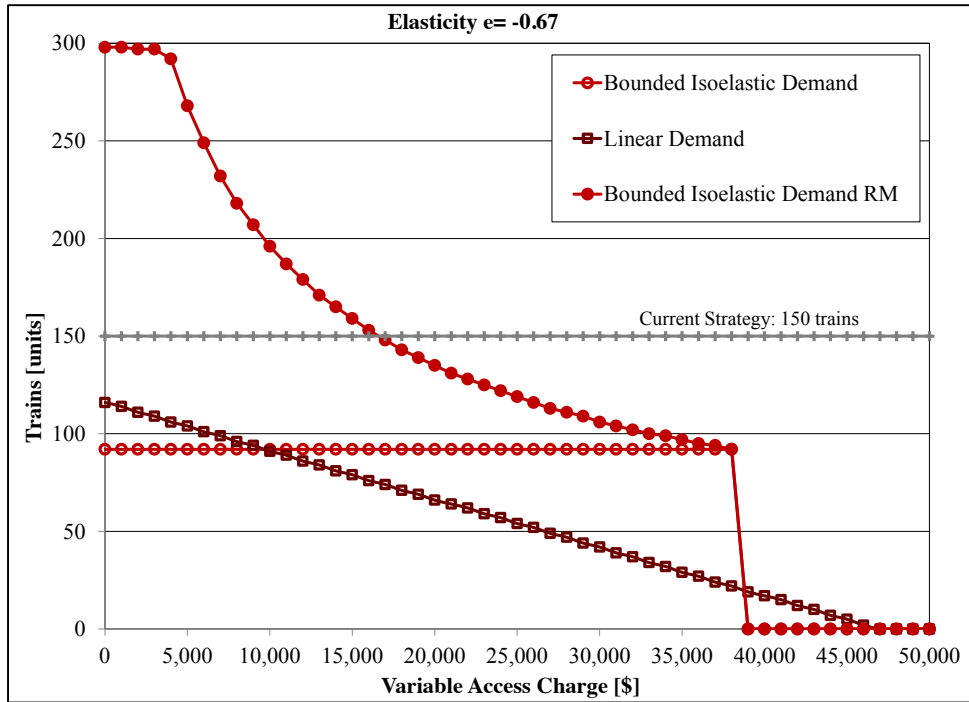
Figure 2-12 Number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -0.67 for different demand functions and fare mechanisms (Source: author)
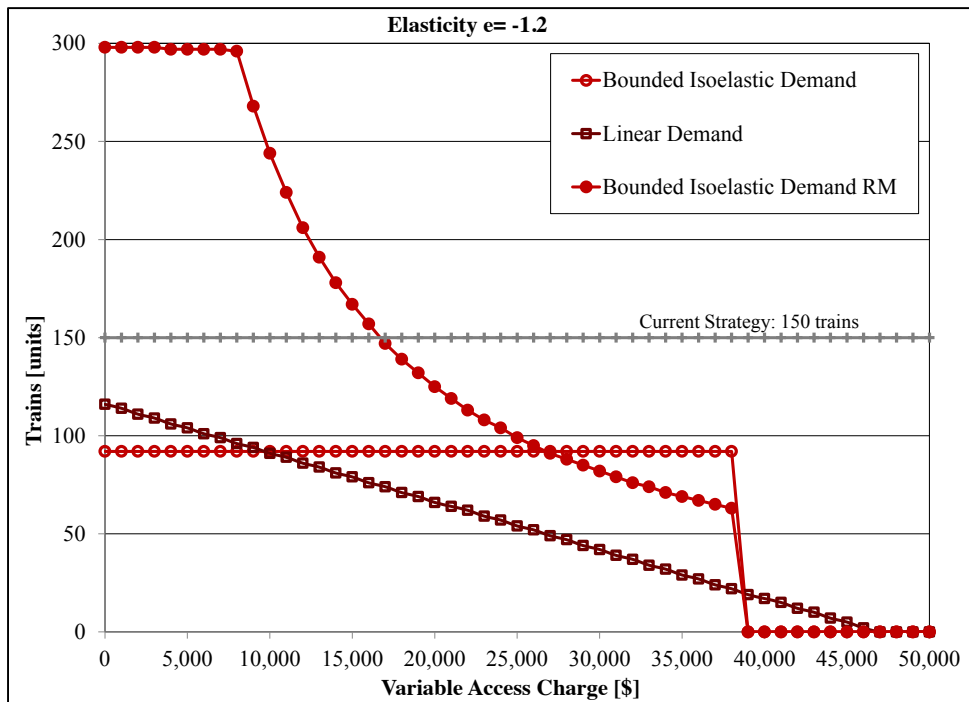
Figure 2-13 Number of trains to be scheduled by a TO as a function of the variable access charges for elasticity value equal to -1.2 for different demand functions and fare mechanisms (Source: author)

2.4.2 Vertically-Separated Regulated Monopolistic Train Operator

These cases assume again vertical separation between TO and the IM. We assume again that the TO is a monopoly in its railway service market and there are no substitute services offered by other TOs. Unlike the previous cases, these cases assume that TO operations are regulated to ensure that the TOs do not charge excessive fares to the end users. We analyze two main cases, one representative of an intercity passenger TO and one representative of a commuter TO.

2.4.2.1 Intercity Passenger TO

Section 2.3.1 analyzes the behavior of intercity passenger monopolistic unregulated TOs that determine the fares and number of trains with the objective of maximizing their profits (see equation (2.3)). In this case however we assume that there is an even-handed regulator that ensures that the TO optimizes service without extracting excessive rents from the end users. That is, we assume that the regulator tries to ensure that the TOs have zero profits (after reimbursing capital at an adequate rate of return). Although this regulation prevents the TOs from making large profits (as in the cases presented in Sections 2.3 and 2.4.1), it also ensures that the access charges that the TO pays to the IM are not excessive (i.e., profits will never be negative in this case, unlike what could happen in the cases presented in Sections 2.3 and 2.4.1). We assume from now on, that intercity passenger demand is a bounded isoelastic function on the fare. As a result, equation (2.3) has to be adjusted in this case.

$$\max_{N,F} N,$$

$$\text{s.t. } s(F,N) + F \cdot \min(d(F), c \cdot N) - fc - vc \cdot N - AC(N) = 0 \tag{2.10}$$

Assuming again that there are no subsidies and that access charges are linear, we first see that the average fare charged to the end users in this case has to be equal to the total costs divided by the total number of travelers. The optimization problem can then be solved by finding the maximum number of trains (n^*) for which the TO can recover costs.

Figures 2-14 and 2-15 show the solution of the optimization problem for different values of variable access charges and for two demand elasticity values (-0.67 and -1.2) assuming that no revenue management mechanism is in place. Figure 2-14 shows the number of trains that a regulated and an unregulated TO would operate and Figure 2-15 shows the fares that they would charge to the end users.
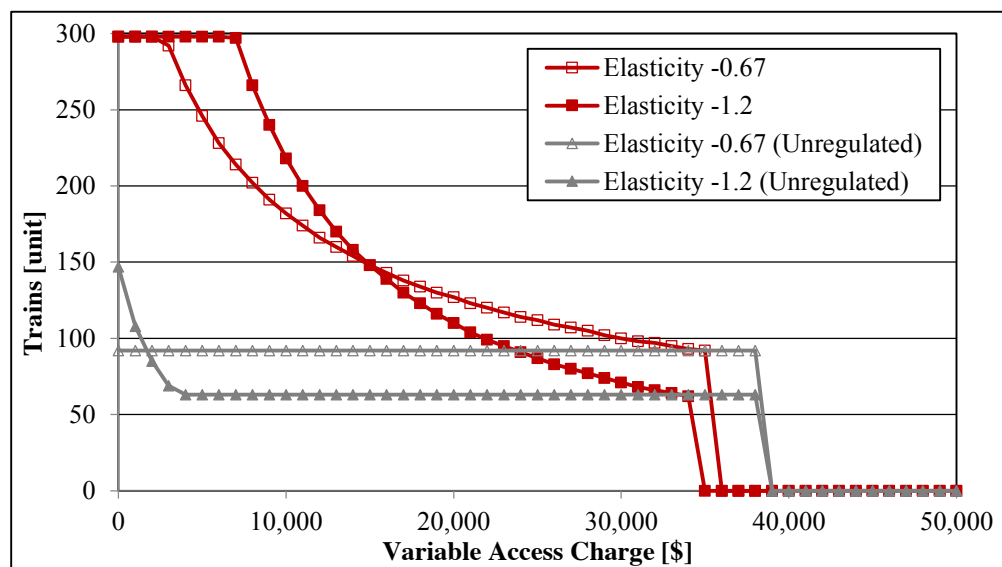
Figure 2-14 Number of trains to be scheduled by a (regulated vs. unregulated) TO as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)
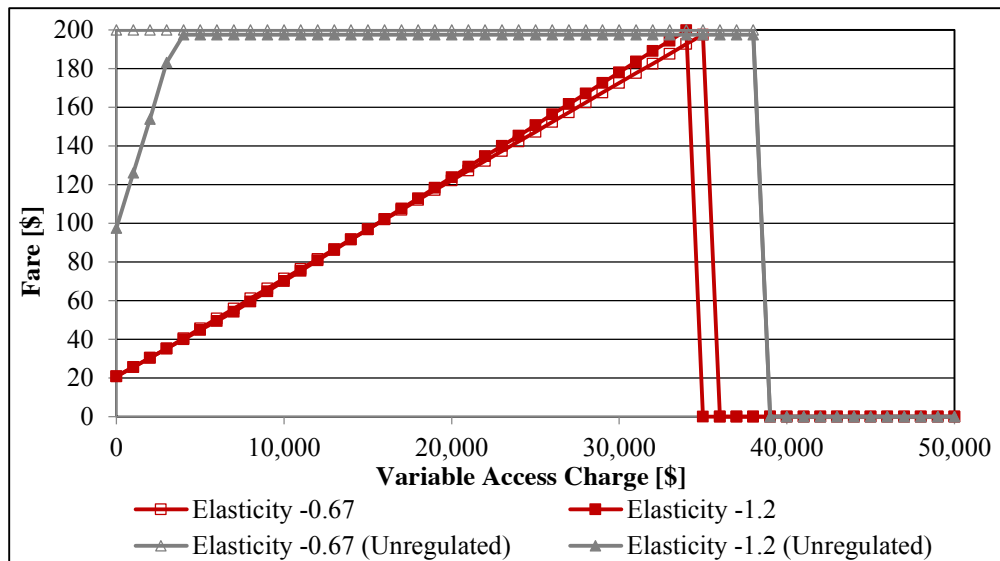
Figure 2-15 Fares charged to the end users by a (regulated vs. unregulated) TO as a function of the variable access charges for elasticity values of −0.67 and -1.2 (Source: author)

There are two important take-away messages from these figures. First of all, as we may expect, regulated train operators operate more services and charge lower fares to the end users. The only exception comes when the variable access charges are so high that the TO is not able to sustainably operate any longer. Second, the demand for scheduling trains of a regulated intercity passenger TO in the NEC is higher than the current number of trains for variable access charges up to $15,000 per train.

2.4.2.2   Commuter TO

This case is inspired by the commuter TOs in the NEC and is different from the previous cases in two senses. First, it assumes that the TOs are regulated by imposing a fare limit. In other words, the regulator does not allow the TOs to change the fares they charge to the end users. Second, it assumes that the demand depends on the frequency of the service. According to (Lago et al., 1981), the demand for commuter services increases when the frequency of service increases and vice-versa.

70

Since the fare is fixed, the elasticity of the demand to the frequency can be defined as $e_n = \dfrac{\Delta d/d_0}{\Delta h/h_0}$ where h is the average headway between consecutive trains. Since the headway is proportional to $1/n$, the elasticity can also be computed as $e_n = -\dfrac{(d-d_0)\cdot n}{(n-n_0)\cdot d_0}$. Therefore, assuming a linear demand function on the frequency, the demand can be determined given the number of trains using $d(n) = (1 - e_n) \cdot d_0 + \dfrac{e_n \cdot d_0 \cdot n_0}{n}$. We can also write $d(n) = \max\left(0, (1 - e_n) \cdot d_0 + \dfrac{e_n \cdot d_0 \cdot n_0}{n}\right)$ to avoid negative demands for low number of services.

The optimal number of trains (n^*) to maximize profits can be determined using equation (2.3) considering that the fare is fixed in this case. Assuming again that access charges are linear $(ac(n) = ac_f + ac_v \cdot n)$ and that the subsidy is a lump sum (paid by the commuter agency to the TO), we can determine that the optimal number of trains would be either:

$$n^* = \sqrt{-\frac{f \cdot e_n \cdot d_0 \cdot n_0}{vc + ac_v}},$$

$$n^* = \frac{(1-e_n)\cdot d_0}{2 \cdot c} - \frac{\sqrt{(1-e_n)^2 \cdot d_0^2 - 4 \cdot c \cdot e_n \cdot d_0 \cdot n_0}}{2c}, \text{ or} \qquad (2.11)$$

$$n^* = 0, \ n^* = \frac{(1-e_n)\cdot d_0}{c}$$

The choice of one number of trains over other would depend on how revenues and cost compare. If revenues obtained from fares are much higher than variable costs, then the optimal strategy to maximize profit would be to maximize revenues. If revenues are comparable to variable costs, the optimal strategy would be to ensure that there is no excess-capacity on the trains. Finally, if variable costs are much higher than the revenues

71

per train, the TO should not operate any train.

Note that this number of trains is independent of the level of subsides and the fixed costs (from operations and access-charges). These values would only affect whether the TO can sustainably operate in the medium term.

As we mentioned before, this case is representative of the situation of the commuter rail TOs in the NEC. According to (MBTA, 2013a; MBTA, 2013b) a TO like the MBTA, the commuter operator in the Boston area, faces fixed operational (direct) costs of $fc = \$435.1k$ per day and variable operational costs of $vc = \$1,666$ per train and per day. The elasticity of the demand with respect to the headway (frequency) is estimated by (Lago et al., 1981) to be equal to $e = -0.41$. As expected, the elasticity is negative because demand increases when headway decreases (i.e., the number of services increases). In 2014, MBTA's average fare ranged from $f_0 = \$7 - \$25$ (we assume an average fare of $f_0 = \$13$), the number of trains averaged $n_0 = 485$ trains per day, with a realized demand of $d_0 = 130.6k$ passengers per day. The train average capacity considered is $c = 350$ passengers, with $80\% +$ load factor. Subsidies $s = \$234k$ per day are considered following (MBTA, 2013a).

Figure 2-16 compares current MBTA profits with the expected profits when the profit maximizing strategy presented in equation 16 is used to determine the number of trains. The results show that higher profits can be unlocked by reducing the number of services, especially when variable costs increase due to access charges. Note that even under the profit maximizing strategy, the TO would not be able to operate if access charges exceed \$2,500 per train per day, since the variable costs of operating the train

would be higher than the revenues obtained. In that case, operating a train would only increase the cost burden for the system.
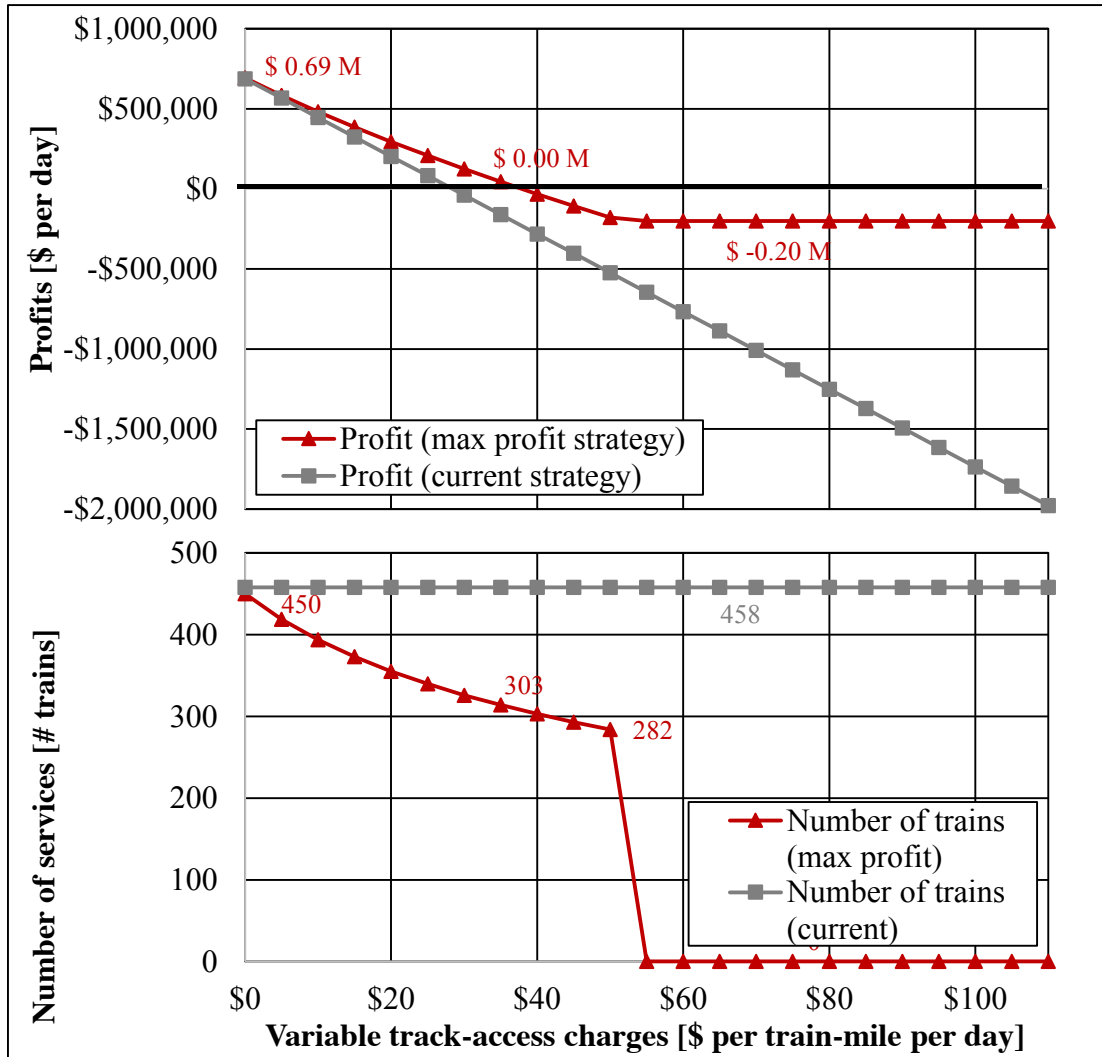
Figure 2-16 MBTA's expected profits and number of trains for different variable access charges
(Source: author)

This analysis shows that under fare limits, a commuter TO similar to the MBTA would be able to make positive profits (after adding the operations subsidy) when variable access charges are low. As a result, a regulation that ensures that the TO always has zero profits would be preferable from the end users' standpoint. However, regulating the end users' fare is much easier for the regulator than ensuring that the TOs have no

profits; what justifies the use of these types of measures when the room to gain profits by the TOs are low.

We also want to compare the variable access charges that a commuter TO could afford with the variable access charges than an intercity passenger TO could afford in the context of the NEC. The previous analysis show that the maximum variable access charges that a commuter TO could afford range between $1,500 and $2,800. We have also shown that the maximum variable access charges that an intercity passenger TO could afford up to $30,000 to $40,000. Our analysis shows that with variable access of $15,000 to $20,000, the number of trains offered by an intercity passenger TO would not be dramatically affected. So these results suggest that intercity TOs in the NEC are able to pay around 10 times higher variable access charges than commuter TOs.

2.4.3   Vertically-Separated Perfectly Competitive Train Operator

These cases assume again vertical separation between TO and the IM. This section assumes that the TOs operate in a perfectly competitive railway service market. An important discussion in these cases is whether TOs may face fixed costs or not. From a theoretical standpoint, if the TOs face any type of fixed costs, it would be more efficient to have a single TO than multiple TOs; and as a result, the market would not be perfectly competitive. Consequently these cases assume no operational fixed costs and no fixed access charges.

The rest of this section presents the analysis for a perfectly competitive intercity passenger service market. Assuming again that the demand is a bounded isoelastic function on the fares, the number of trains and the fares (n^*, f^*) that the TOs would offer in equilibrium and perfect competition can be determined considering that: 1) the optimal

number of trains in equilibrium would be the one at which no TO could be able to schedule an additional train without losing money; and 2) the optimal fares in equilibrium would be the ones at which the TOs would be indifferent between offering the services or not.

Conditions 1) and 2) can be formalized in equation (2.12):

$$\max_{N,F} N,$$

$$\text{s.t. } s(F,N) + F \cdot \min(d(F), c \cdot N) - vc \cdot N - AC_v \cdot N = 0 \tag{2.12}$$

Assuming no subsidies and a bounded isoelastic demand curve, we can analytically solve the problem stated in equation (2.12) to determine that n^*, f^* are either:

$$\begin{array}{ll} f^* = \frac{vc+ac_v}{c}, n^* = \left\lfloor \min\left(\frac{k \cdot (vc+ac_v)^e}{c^{e+1}}, \frac{\overline{d}}{c}\right) \right\rfloor & \text{if } vc + ac_v \leq c \cdot \overline{f} \\ \quad\quad f^* = 0, n^* = 0 & \text{otherwise} \end{array} \tag{2.13}$$

Figures 2-17 and 2-18 show the solution of the optimization problem in the Amtrak-inspired example for different values of variable access charges and for two demand elasticity values (-0.67 and -1.2). Figure 2-17 shows the number of trains that perfectly competitive TOs would operate and Figure 2-18 shows the fares that they would charge to the end users. These results are compared with the number of trains and fares of regulated and unregulated monopolistic TOs. The monopolistic TO's number of trains and fares are slightly different from the ones presented in Sections 2.3.1 and 2.4.2.1, because they assume no fixed costs or access charges for comparability purposes. In particular, the number of trains that a regulated monopolistic TO would operate is slightly higher than the one in Sections 2.3.1 and 2.4.2.1, and the fares charged are slightly lower (since there is no fixed cost to recover).
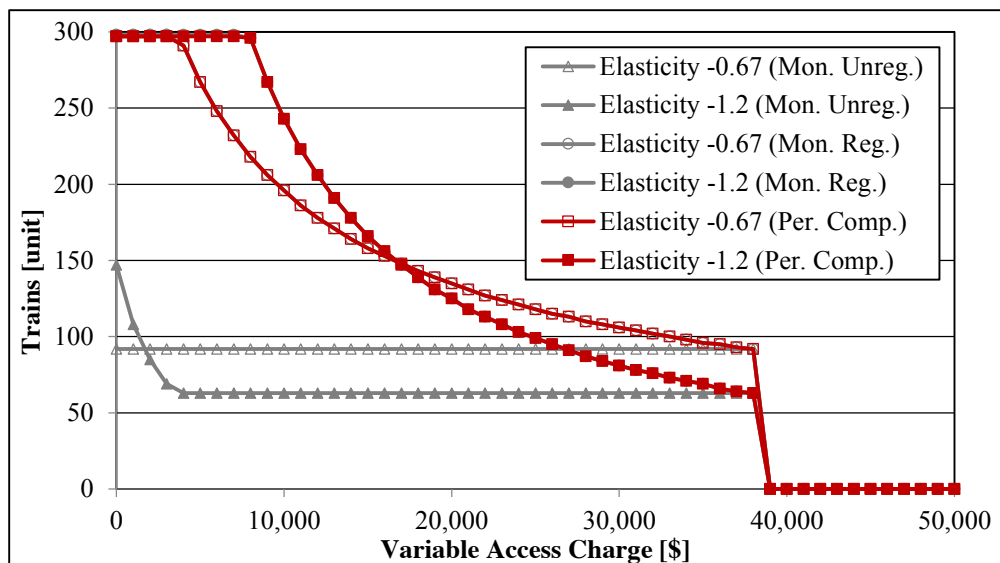
Figure 2-17 Number of trains to be scheduled by (perfectly competitive vs. monopolistic) TOs as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)
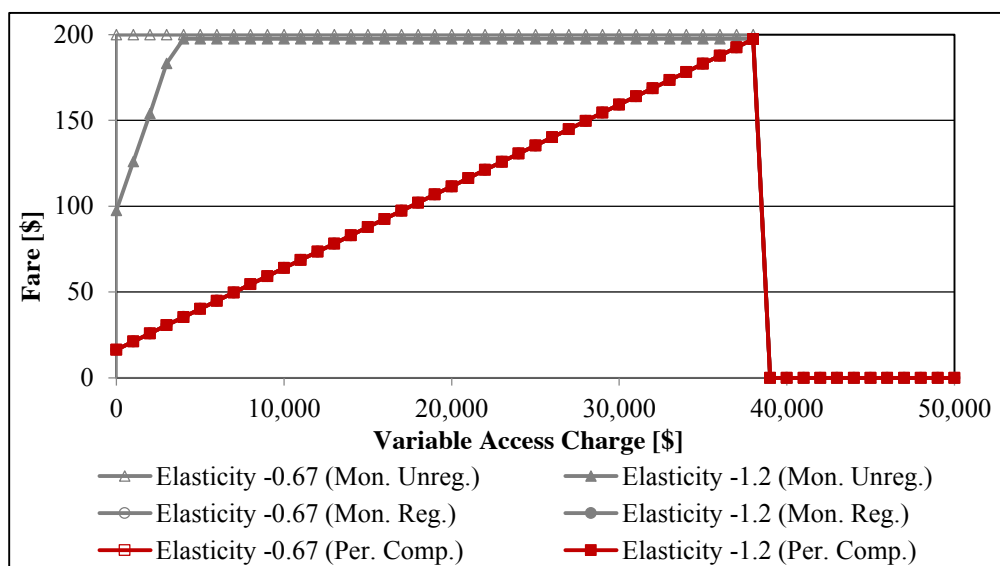
Figure 2-18 Fares charged to the end users by (perfectly competitive vs. monopolistic) TOs as a function of the variable access charges for elasticity values of -0.67 and -1.2 (Source: author)

These results show that perfectly competitive TOs would perform very similarly to monopolistic regulated TOs, as we may expect. The levels of services and fares for these two cases completely overlap in Figures 2-17 and 2-18. The only differences between these cases come from the discrete nature of the number of trains. In some instances a regulated monopolistic TO would be able to offer one more train than a

perfectly competitive market with slightly higher fares and lower train utilization rates. Note however, that an unregulated operator would always operate fewer trains and charge higher fares to the end users. As we showed in Section 2.4.2., the demand for scheduling trains of both regulated intercity passenger TOs and perfectly competitive intercity TOs in the NEC is higher the current number of trains when variable access charges are lower than or equal to $15,000 per train per day. In these cases, the TOs have incentives to operate trains as long as variable access charges are smaller than $38,550 per train.

2.4.4   Vertically-Integrated Railway Company

These cases analyze a vertically-integrated railway company. We use the results of this section as a benchmark for the previous cases. Equations (2.3), (2.4), and (2.5) have to be adjusted to analyze vertically-integrated systems because 1) a vertically-integrated railway company also faces the infrastructure management costs; and 2) track access charges in this case are not necessary (transfer between the TO and the IM that cancels out in a vertically-integrated system).

This section considers two main cases: one where the railway company offers only one type of railway service and another one where the railway company offers two types of (non-substitutable) railway services.

2.4.4.1   Single Type of Service: Intercity Passenger Service

We assume again that the intercity passenger service face a bounded isoelastic demand as a function of the fare charged to the end users. Apart from the pure operations-related costs discussed in Section 2.2, an integrated railway company also faces infrastructure-related costs. At a high level, the infrastructure costs of a railway company can be aggregated in fixed and variable costs. For the context of this research,

we consider fixed costs (fc_{IM}) all costs that do not change in the medium term with the number of trains operated on the infrastructure. We consider variable costs (vc_{IM}) all costs that depend on the number of trains operated on the infrastructure in the short or medium term horizon. As a first order approximation, we will assume that variable costs depend linearly on the number of trains operated.

Equation (2.3) could be adapted to this case by including these costs and eliminating the access charges (internal transfer between the TO and the IM):

$$\max_{N,F}[s(F,N) + F \cdot \min(d(F,N,tt), c \cdot N) - fc - vc \cdot N - fc_{IM} - vc_{IM} \cdot N] \qquad (2.14)$$

Initially, the only difference between equations (2.3) and (2.14) is the fact that the access charges appear in place of the infrastructure related costs. This makes sense, because the objective of access charges is to pass the infrastructure costs on to the TO. As a result, if the access charges scheme just replicated the infrastructure cost scheme (i.e., $AC_f = fc_{IM}, AC_v = vc_{IM}$), there would be no differences between the operational decisions of a vertically-separated TO and the operational decisions of an integrated company. In other words, the vertical separation of the system would introduce no distortion in the operational incentives. Consequently, depending on the regulatory and competitive environment, the integrated railway company would exhibit the same type of behavior than the vertically-separated TOs discussed in Sections 2.3, 2.4.2, and 2.4.3. This finding is consistent with the findings of other network industries (Gomez-Ibanez, 2003; Laffont and Tirole, 1993; Laffont and Tirole, 2000; Perez-Arriaga, 2013). Not surprisingly, many countries have adopted infrastructure marginal costs to price infrastructure capacity (Texeira and Lopez-Pita, 2012; Texeira and Prodan, 2014).

If there are some differences though on the variable component (i.e., $AC_v \neq vc_{IM}$), then a vertically-separated TO would have incentives to operate more trains (if $AC_v < vc_{IM}$) or fewer trains (if $AC_v > vc_{IM}$) than an integrated railway company. Figure 2-19 shows how these differences would result in total utility losses.

Figures 2-19 and 2-20 show the social utility, and the end users, TO, and IM utilities respectively for an integrated railway system, and compared them to those of vertically-separated systems as a function of the variable access charges. We assume $fc_{IM} = \$1m, vc_{IM} = \$10,000$.
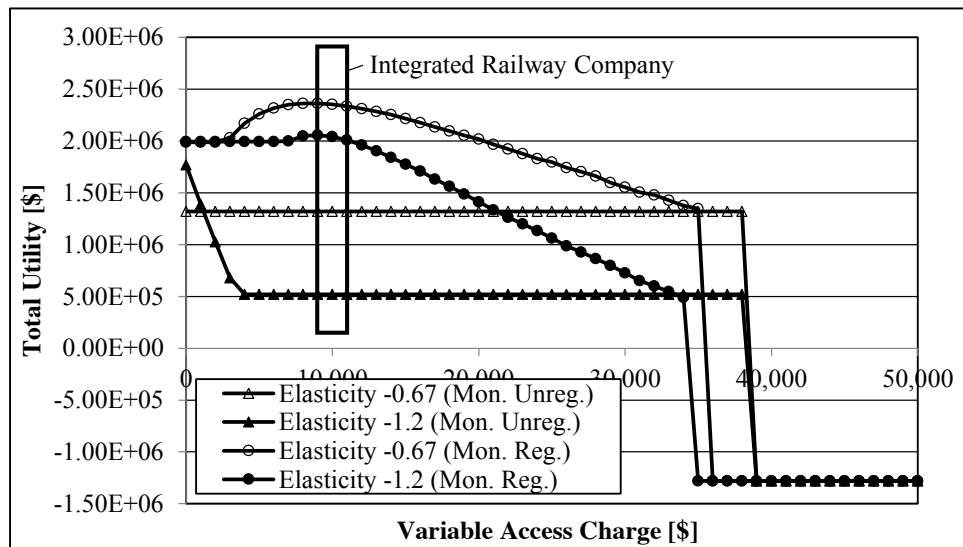
Figure 2-19 Total social utility associated with integrated and vertically-separated monopolistic (unregulated and regulated) TOs as a function of the variable access charges for elasticity values of -0.67 **and -1.2** (Source: author)

Note that the differences between variable infrastructure costs and variable access charges do not only result on different utility distributions between the TO, and the IM; but also in utility losses for the society at large. The regulatory framework and the competitive scheme also affect the utility. As Figures 2-19 and 2-20 show, the greediness of unregulated TOs does not only change the utility distribution between end users and

TOs, but also imposes losses in the total social utility driven by the lost demand from reduced number of trains.
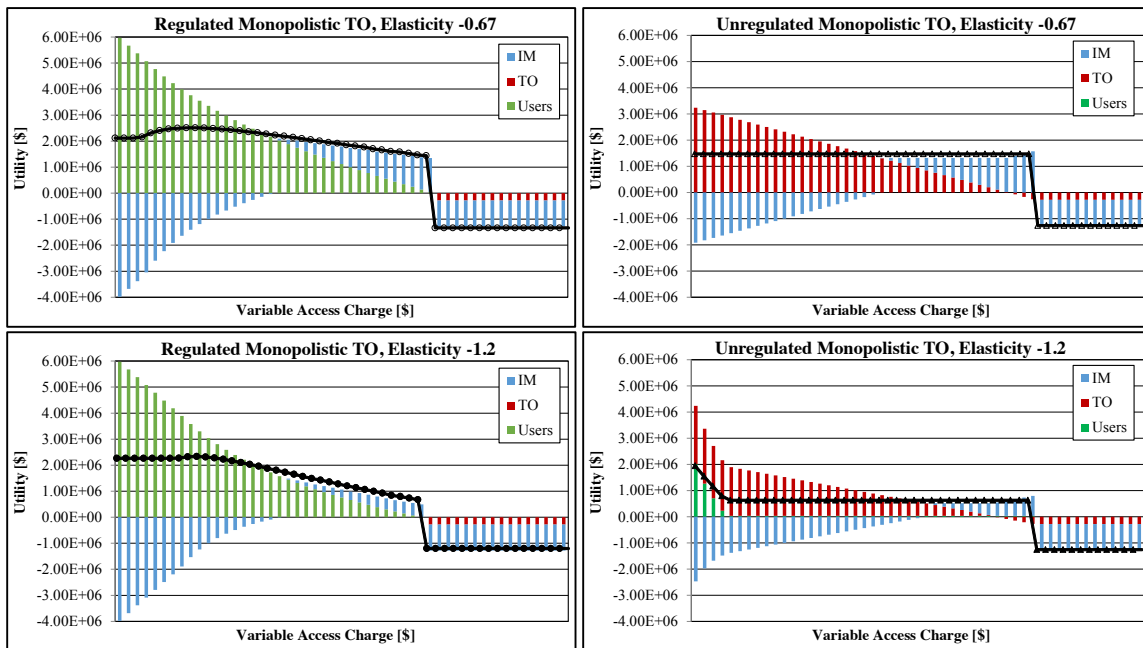
Figure 2-20 End users, TO, and IM's utilities for vertically-separated monopolistic (unregulated and regulated) system as a function of the variable access charges for elasticity values of -0.67 and -1.2
(Source: author)

Any differences in the fixed cost component (i.e., $AC_f \neq fc_{IM}$) would mostly affect the distribution of utility between the TO and the IM (welfare transfers). These differences would not affect the operations of an unregulated monopolistic TO unless they drive the TO profits below their fixed costs. If this is the case, the unregulated monopolistic TO would no longer have incentives to offer any train service (and there would be losses on total social utility). These differences though would also affect the number of trains and fares that a regulated monopolistic TO would operate, slightly reducing social utility. These results are further discussed in the context of Tanzania in Chapter 4.

2.4.4.2   Several Types of Services

This case considers not only the infrastructure-related costs (fc_{IM}, vc_{IM}), but also the operations of different types of services i (such as intercity passenger train services, commuter services, or different types of freight services). We assume that in general, the subsidies, fares, number of trains, infrastructure and operations related variable costs of different types of services may be different (and we indicate that by adding the subscript i). Equation (2.3) could be adapted to this case:

$$\max_{N,F}[\textstyle\sum_i(s_i(F_i, N_i) + F_i \cdot \min(d(F_i, N_i, tt_i), c_i \cdot N_i) - vc_i \cdot N_i - vc_{IMi} \cdot N_i) - fc - fc_{IM}]$$

$$(2.15)$$

Initially, equation (2.15) can be solved independently for each i if: 1) we are able to ensure that there are no interdependencies between the best number of trains in different markets; and 2) we design a mechanism to allocate the fixed cost between the different services. In some instances we can guarantee these two conditions. For example, the optimal levels of service and fares of the various types of services are independent if there are no infrastructure capacity limitations and the services are not substitutes. Furthermore, we have shown that in unregulated monopolistic markets and in perfectly competitive market, the level of operations (provided that the TOs have incentives to operate trains) do not depend on the fixed costs. In this case, the results obtained match the results obtained in the previous sections.

This discussion is further extended in the rest of the dissertation to discuss how the results change 1) when the IM needs to assign fixed costs among the different types of services and 2) when there are infrastructure capacity limitations. Unfortunately, most railway systems fall into those categories. The issues around infrastructure capacity

limitations are addressed by integrating the results of this chapter with the results of the infrastructure manager model from Chapter 3. We illustrate these issues in the context of the NEC (Chapter 5), where intercity passenger TOs, commuter TOs, and freight TOs compete to get access to the infrastructure. The issues around the allocation of infrastructure capacity fixed costs are in the context of the Central Corridor in Tanzania (Chapter 4) where different types of freight service TOs (general cargo and container TOs) share the same infrastructure. While (Laffont and Tirole, 1993; Laffont and Tirole, 200; Perez-Arriaga, 2013) show that the use of marginal infrastructure costs to price capacity often allows for infrastructure cost recovery in other network industries, these results show that infrastructure cost recovery is not possible with marginal infrastructure cost pricing in most railway systems.

2.5 Conclusions

This chapter presents a simple TO Model based on standard TO financial models to discuss how TOs respond to alternative capacity pricing and allocation mechanisms as a function of the institutional and technical context in which the TOs operate. The TOs' response is captured by analyzing three main operational decisions: 1) the passenger fare or freight shipping rate charged to end users; 2) the number of trains operated or number of trains; and 3) the access charges paid to the infrastructure manager to access the infrastructure. Understanding the TOs' response to different access charges is an important step to analyze and compare alternative capacity pricing and allocation mechanisms. The model proposed allows regulators to robustly infer the TO demand for scheduling trains and their ability to pay to access the infrastructure with little information about the TO cost structure and the end users' demand.

There are four main take-away messages from this analysis (in reversed order). First, if the access charges reflect the infrastructure-related costs associated with operations, a vertically-separated TO would make the same operational decisions as an integrated railway system. However, if the access charges do not reflect the costs in which the IM incurs as a result of the operations of trains in the system, the TO would have incentives to provide different levels of service. This often translates into a loss in the total welfare. These implications also stand when different TOs share the same infrastructure. In other words, the use of marginal infrastructure costs to price capacity ensures that TOs make the same decisions than an integrated railway company. However, these mechanisms cannot be used in all cases. Chapters 4 and 5 discuss how to analyze these issues in cases in which there is a need to recover infrastructure costs or when infrastructure capacity is limited.

Second, the introduction of any type of TO regulation or the introduction of competition in the operations result in higher levels of service, lower fares for the end users, and higher levels of total welfare as compared to the operations of unregulated monopolistic operators. The operations of TOs in perfectly competitive markets and the operations of regulated monopolistic TOs are very similar. In some instances though, regulated TOs operate one more train because of the discrete nature of the number of trains.

Third, the number of trains estimate produced by our models depends on the functional form of the demand assumed, on the elasticity, and on the existence of any type of revenue management mechanisms. Although the results obtained are pretty robust to model inputs, a good characterization of the users' demand is important to accurately

estimate the TOs operations. The evidence presented in the literature (Morrison, 1990) and the comparison of the results with current levels of service operated suggest that the demand of the NEC intercity passenger operators used as an example to illustrate the results in the different cases, is best characterized as a bounded isoelastic function of the fare. Nonetheless, the robustness of the results also justifies the approximation of the demand function by a linear function if this approximation simplifies the calculations.

Finally, this research analyzes the maximum access charges that different types of TOs would be able to pay to access the infrastructure. These results are very robust across the different cases studied. The model also anticipates the TOs' response (number of trains) to access charges. This information is used as inputs of the IM Model presented in Chapter 3, allowing the IM to anticipate and understand the operational goals and infrastructure needs of operators on their network. This information is also valuable for regulators, enabling them to understand the performance of the system under alternative mechanisms to price and allocate railway capacity.

Chapter 3 -   The Infrastructure Manager Problem: Determining Train Timetable and Infrastructure Access Charges under Alternative Capacity Pricing and Allocation Mechanisms

"You never change things by fighting the existing reality. To change something, build a new model that makes the existing model obsolete." – B. Fuller (1981)

The train operator (TO) Model proposed in Chapter 2 can be used to anticipate the demand of individual TOs to access the infrastructure under alternative capacity pricing and allocation mechanisms. In congested shared railway systems, there are often conflicts between the services that different TOs would like to operate within the existing infrastructure. The next step to evaluate capacity pricing and allocation mechanisms is thus to analyze which trains can be scheduled within available infrastructure capacity.

However, capacity availability in the railway industry cannot be known in the absence and understanding of infrastructure operations. Therefore assessment of capacity requires the determination of the train timetable, which eliminates any potential conflicts between the TOs' requests to use infrastructure capacity. Although there is a broad literature that proposes train timetabling methods for dedicated railway systems, there are few models that can be used for shared competitive railway systems.

This chapter proposes a train timetabling model for shared railway systems explicitly considering a variable number of trains, with large flexibility margins (TOs' willingness to deviate from their desired timetable), and a variety of train services traveling along different paths. The TOs' demand for scheduling trains is assumed to be exogenous (input from Chapter 2). The model is formulated and solved both as a mixed integer linear programming (MILP) problem (using a commercial solver) and as a

dynamic programming (DP) problem. We solve the DP formulation with a novel algorithm based on a linear programming (LP) approach to approximate dynamic programming (ADP) that can solve much larger problems than commercial MILP solvers.

This model can be used to evaluate the best possible train timetable under alternative capacity pricing and allocation mechanism. We use the results to understand the interactions between capacity planning and capacity operations in shared railway systems. Understanding these interactions is important to be able to design effective capacity pricing and allocation mechanisms. Part of this work has been submitted for publication (Pena-Alcaraz, Webster, and Ramos, 2015a).

The rest of the chapter is structured as follows: Section 3.1 reviews the train timetabling literature in shared railway systems and summarizes the main contributions of this chapter. Section 3.2 describes and formulates the train timetabling problem in shared railway systems, and motivates the assumptions of the chapter. Section 3.3 presents a DP formulation of the problem, and describes the LP-based ADP solution algorithm. Section 3.4 compares the computational performance of the ADP algorithm with the performance of commercial MILP solvers and illustrates the insights obtained using the algorithm to design shared railway systems' timetables for several cases with traffic patterns similar to the traffic of the Northeast Corridor (NEC) in the US. Section 3.5 summarizes the main implications of the results obtained with the model and concludes.

3.1 Train Timetabling Problem for Shared Railway Systems

As mentioned above, the design, assessment, and implementation of capacity pricing and allocation mechanisms at the strategic level are tightly coupled with the infrastructure operations at the tactical level. In other words, capacity utilization cannot be determined in the absence of infrastructure operations because available railway capacity depends on how the infrastructure is operated (Krueger et al., 1999; Pouryousef and Lautala, 2015). The operations in the railway industry are defined by the train timetable that determines the arrival and departure time at every station of all trains scheduled. As a result, the design of the train timetable is a critical step in any capacity planning mechanism. The timetable specifies how the competing demands for infrastructure access are coordinated to meet the infrastructure manager (IM)'s objectives and constraints.

The train timetabling problem has been widely studied in the literature. There are two main approaches to design the best train timetable that meets a set of operational constraints, both based on MILP formulations. (Castillo et al., 2009; Ghoseiri et al., 2004; Liebchen, 2008; Liebchen and Peeters, 2009; Pena-Alcaraz et al., 2011; Zhou and Zhong, 2005) present formulations to compute the train arrival and departure times. Traditionally, these models have been called multi-mode resource constrained project scheduling models. (Cacchiani et al., 2010; Caimi et al., 2009; Caimi et al., 2011; Caprara et al., 2002; Caprara et al., 2011; Cordeau et al., 1998) present formulations that represent the final timetable as a collection of nodes and arcs. Each arc represents possible train arrival and departure times at stations. Infrastructure and operational

constraints are imposed by determining subsets of compatible and incompatible arcs. Traditionally, these models have been called multi-commodity flow models.

We argue, however, that these models cannot be used to analyze capacity planning mechanisms in shared railway systems for three reasons. First, with the exception of (Caprara et al., 2011), the approaches above assume a fixed number of trains to be scheduled on the infrastructure. However, the number of trains to schedule is the main decision variable of the capacity allocation problem in shared railway systems. Second, with the exception of (Caprara et al., 2011) again, the approaches above assume a single TO that tries to schedule trains. This TO could iteratively solve the train timetabling problem, introducing small modifications in each train desired timetable until the resulting timetable meets its needs. In shared railway systems, however, multiple TOs request access to the infrastructure. To ensure that the TOs reveal the value to themselves of each train to be scheduled and to avoid strategic behavior, the IM accepts inputs from the TOs only at specified time-points. As a result, TOs have incentives to provide large flexibility margins around the desired train timetables requested to ensure that the trains they value are scheduled even when there are small conflicts with other trains. The flexibility margin determines how much time TOs are willing to deviate from the desired timetable. Third, most of these models assume that all trains follow the same path. Again, this assumption does not hold when the nature of the services operated in the shared railway system is different. For example, commuter services are typically scheduled around the metropolitan areas, whereas intercity and freight TOs offer services between cities.

This chapter presents a multi-mode resource constrained project schedule formulation for shared railway systems that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. This approach 1) introduces a discrete variable that indicates whether a train can be scheduled or not, 2) uses flexibility margins to ease conflicts, making fast trains travel slowly when there are slow trains ahead and making slow trains wait at sidetracks when fast trains overtake them, and 3) specifies safety constraints (spacing of the trains) for each train path.

These additional considerations make the problem very difficult to solve. From a computational standpoint, the size of the model increases exponentially with the number of stations and the number of trains to schedule. As a result, commercial solvers are only able to solve the problem for small number of trains. Furthermore, most of the techniques developed in the train timetabling literature are designed for traditional single-operator train timetabling problems and cannot be used in this case. Most classical decomposition approaches do not work because of the large number of discrete variables needed to specify which trains are scheduled and to pinpoint the order in which trains go through each station. Any technique that exogenously fixes train order cannot be used here because of the large flexibility margins and because train spacing constraints are specific to each individual train.

To be able to solve the problem in meaningful instances, we propose an alternative class of solution algorithms using ADP techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007). This research develops a novel Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends previous algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). This algorithm

allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap.

In summary, the introduction of shared railway systems requires the design, assessment, and implementation of capacity planning mechanisms to coordinate multiple TOs and the IM. The use of this novel algorithm allow us to solve the train timetabling problem in shared railway systems considering a large number of trains (100 to 150 trains). As a result, we are able to determine the optimal capacity allocation plan given the TO's demand for capacity under alternative capacity pricing and allocation mechanisms. We can use these results to anticipate the answers to relevant policy-type questions such as: how much should intercity TOs pay to be able to schedule services that conflict with commuter train services; whether freight TOs would be able to schedule any trains on the infrastructure, etc. The answers to these questions are central to our ability to design effective capacity pricing and allocation mechanisms.

This research makes both methodological and railway systems-specific contributions. From a methodological standpoint, we present a model that explicitly considers the relevant characteristics of shared railway systems, and offers a novel ADP algorithm for solving this complex train timetable problem for large system sizes that are computationally intractable using commercial software. From a transportation standpoint, the modeling framework and the algorithm developed enable us to simulate optimal decisions by an IM for shared railway systems. These results can be used to answer relevant policy-type questions to design appropriate pricing and allocation mechanisms and to understand the implications of infrastructure shared use.

3.2   Mixed-Integer Programming Formulation

In this section, we formulate the train timetabling problem for shared railway systems under capacity pricing and allocation mechanisms. As we discussed in Chapter 1, there are two main types of market-based mechanisms for capacity pricing and allocation: 1) mechanisms that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not (price-based mechanisms); and 2) mechanisms that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity (capacity-based mechanisms or auctions) (Gibson, 2003). Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which train to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges).

The model presented here determines the optimal set of trains that the IM can accommodate, assuming that an auction mechanism is implemented. Under an auction, at some predetermined frequency, the TOs will have the opportunity to submit bids. Each bid will consist of a list of the trains that the TO wants to schedule on the infrastructure, the desired timetable for each train, and the access charges they are willing to pay to schedule each train. The IM will then determine the set of trains that can actually be scheduled, their timetable, and the access charges that the TOs will pay. We assume that the IM's objective is to maximize revenue and cannot restrict access to the infrastructure beyond the infrastructure constraints (e.g., safety, infrastructure maintenance plans). This thesis assumes that the IM is government owned and not for profit, or in other words, that it does not uses it market power to restrict the access to the infrastructure to the TOs.

We also discuss below how to modify the model to determine the optimal set of trains that the IM can accommodate under alternative price-based mechanisms. The differences between the IM models for each mechanism affect mainly the definition of the parameters and the choice of the objective function. The constraints however are related to the physical characteristics of the infrastructure and remain unchanged across mechanisms. The model formulation is discussed below.

3.2.1 Sets

$i, i \in \{1, \dots, I\}$ train services proposed by the TOs in the bidding process.

$j, j \in \{1, \dots, J\}$ railway system stations.

3.2.2 Parameters

We use again lower-case letters to denote parameters. The information that the TOs provide in the bidding process for every train i is:

ini_{ij} a Boolean matrix that indicates the initial station j from which train i departs.

fin_{ij} a Boolean matrix that indicates the final destination (station j) of train i.

a_i the maximum access price (access charge) that the TO is willing to pay if train i is scheduled. For price-based mechanisms the access price will be predetermined (using, for example, a model to allocate infrastructure-related costs proportionally to infrastructure use) and fixed by the IM depending on the characteristics of the service. It is important to note that the TO will only operate a train if that price is less than or equal to its willingness to pay determined in Chapter 2.

$t_{ij}^{arr}, t_{ij}^{dep}$ the desired arrival and departure time of train i at every station j in the path

of train i.

$\Delta td_i, p_i^{\Delta td}$ maximum acceptable translation, defined as the maximum acceptable difference between the desired timetable and the actual timetable at the initial station (see Figure 3-1) of train i and penalty imposed by the TO if the IM translates the train over the desired timetable. The penalty specifies the reduced access price that the TO is willing to pay.

$\Delta tr_i, p_i^{\Delta tr}$ maximum acceptable change in train i total travel time (see Figure 3-1) and penalty imposed by the TO if the IM increases the travel time of train i at any station with respect to the desired timetable.
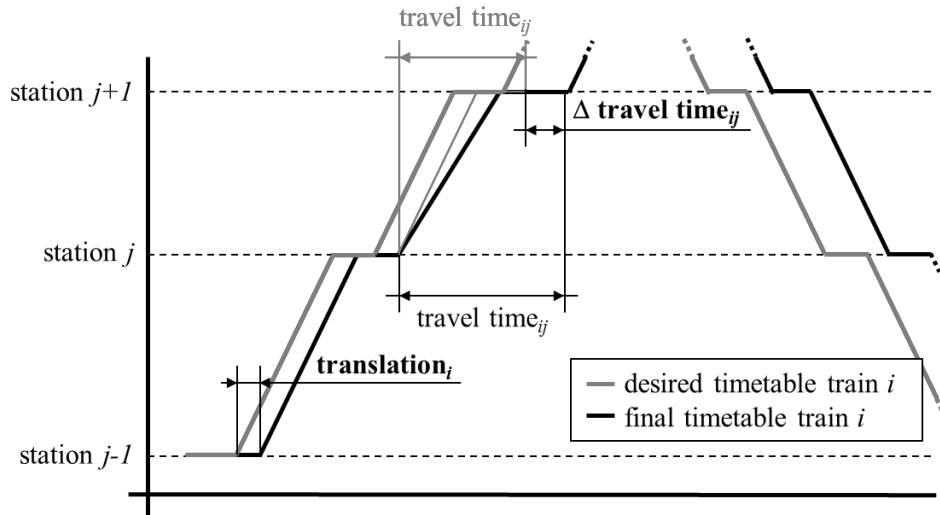
Figure 3-1 Time-space diagram representation of possible changes with respect to the desired timetable (Source: author)

The information about the topology of the line and the type of service is represented by the following two matrices:

$stat_{ij}$ a Boolean matrix that indicates whether train i travels through station j or not.

$next_{ijj'}$ a Boolean matrix that indicates for each train i the station j' that train i will visit immediately after station j. Train i may not stop at station j'.

In addition, the topology of the tracks and the signaling system will determine the minimum safe headway (time elapsed) between consecutive maneuvers at every station:

h_j^{arr}, h_j^{dep} minimum headway between consecutive arrivals/departures to/from station j.

In some cases the minimum safe headway depends also on the type of service and on the characteristics of the rolling stock. If that is the case, the former parameters will have different values for each train pair. The IM can set larger minimum headway to ensure the reliability of the timetable (including time-slack to recover delays in the system).

3.2.3   Variables

We use capital letters for variables. The endogenous decision variables of this problem are:

S_i binary variable that indicates whether train i is scheduled.

$T_{ij}^{arr}, T_{ij}^{dep}$ final arrival and departure time (timetable) of every train i scheduled at every station j in the path of the train.

$\Delta TD_i, \Delta TR_{ij}$ final train i translation and increment of travel time per station j. Note that these variables can be determined knowing $T_{ij}^{arr}, T_{ij}^{dep}$ and vice versa. This research assumes $\Delta TR_{ij} \geq 0$ to ensure that the resulting train timetable is feasible. ΔTD_i can either be positive or negative; so we define the auxiliary positive variable ΔTD_i^+ as the absolute value of ΔTD_i.

$O_{ii'j}$ binary disjunctive variable with value 1 if train i departs before train i' at station j and value 0 otherwhise.

3.2.4 Objective Function

As discussed before, the objective of the problem is to determine which trains should be scheduled and when, in order to maximize the IM's revenue:

$$max\left[\sum_i a_i S_i - p_i^{\Delta td}\Delta TD_i^+ - p_i^{\Delta tr}\sum_j \Delta TR_{ij}\right] \tag{3.1}$$

Alternative objective functions could be defined for different capacity pricing and allocation mechanisms. For example, the functions:

$$max\left[\sum_i S_i\right] \tag{3.2}$$

$$max\left[\sum_i pr_i S_i\right] \tag{3.3}$$

could be used to maximize the number of trains scheduled or the number of priority trains scheduled respectively under price-based mechanisms. In this case pr_i would be a parameter that indicates the priority level of each train i. This priority level can, for example, be proportional to the number of passengers times the miles of the service.

3.2.5 Constraints

The first set of constraints establishes the relation between the desired timetable and the final timetable of every train scheduled:

The departure time from the first station can be determined as:

$$T_{ij}^{dep} = t_{ij}^{dep} + \Delta TD_i, \forall i, j: ini_{ij} \tag{3.4}$$

The travel time between intermediate stations can be determined as:

$$T_{ij'}^{dep} - T_{ij}^{dep} = t_{ij'}^{dep} - t_{ij}^{dep} + \Delta TR_{ij}, \forall i, j, j': next_{ijj'}, fin_{ij'} = 0 \tag{3.5}$$

At the final station, the travel time can be determined using:

$$T_{ij'}^{arr} - T_{ij}^{dep} = t_{ij'}^{arr} - t_{ij}^{dep} + \Delta TR_{ij}, \forall i, j, j': next_{ijj'}, fin_{ij'} \tag{3.6}$$

Note that the arrival time at the initial station is not defined in the timetable, nor is the departure time from the last station.

To ensure that the timetable is feasible, the scheduled stopping and travel time at each station must be greater than or equal to the stopping and travel time in the desired timetable:

$$T_{ij}^{dep} - T_{ij}^{arr} \geq t_{ij}^{dep} - t_{ij}^{arr}, \forall i,j: stat_{ij}, ini_{ij} + fin_{ij} = 0 \qquad (3.7)$$

$$T_{ij'}^{arr} - T_{ij}^{dep} \geq t_{ij'}^{arr} - t_{ij}^{dep}, \forall i,j: next_{ij'j}, fin_{ij} = 0 \qquad (3.8)$$

The maximum translation and increment of travel time for which the TO receives a discount is constrained for each train scheduled. The allowable translation of a train is bounded by a maximum translation defined by the TO:

$$-\Delta td_i \leq \Delta TD_i \leq \Delta td_i, \forall i \qquad (3.9)$$

In addition, the absolute value of the translation ($\Delta TD_i^+ = |\Delta TD_i|$) is determined using the following linear constraints:

$$\Delta TD_i^+ \geq \Delta TD_i, \Delta TD_i^+ \geq -\Delta TD_i \ \forall i \qquad (3.10)$$

The maximum change on travel time is bounded by the maximum increment on travel time specified by the TO:

$$\sum_{j:stat_{ij}} \Delta TR_{ij} \leq \Delta tr_i, \forall i \qquad (3.11)$$

The TO may impose additional conditions within the bid to define the acceptable changes with respect to the desired timetable. That happens when the TO is not interested in operating the train if the departure from or the arrival at one major station is changed. In this case, additional constraints are included to ensure that the timetable respects the TO's requests if the train is scheduled.

The final set of constraints ensures that the timetable proposed by the IM can be accommodated by the existing infrastructure. The IM must ensure first that the difference

between the departure times of every pair of trains scheduled is greater than or equal to the minimum safe headway, so at least one of the following equations must hold:

$$T_{ij}^{dep} - T_{i'j}^{dep} \geq h_j^{dep} \tag{3.12}$$

$$T_{i'j}^{dep} - T_{ij}^{dep} \geq h_j^{dep} \tag{3.13}$$

These conditions can be expressed using the following disjunctive constraints:

$$T_{ij}^{dep} - T_{i'j}^{dep} \geq h_j^{dep} - M_{ii'j}(O_{ii'j} + 2 - S_i - S_{i'}), \forall i, i', j : i < i', stat_{ij}, stat_{i'j} \tag{3.14}$$

$$T_{i'j}^{dep} - T_{ij}^{dep} \geq h_j^{dep} - M_{ii'j}(3 - O_{ii'j} - S_i - S_{i'}), \forall i, i', j : i < i', stat_{ij}, stat_{i'j} \tag{3.15}$$

In these equations $M_{ii'j}$ is a "big enough" number to ensure that one and only one of the equations (3.12) and (3.13) holds. In this formulation we use $M_{ii'j} = h_j^{dep} + t_{i'j}^{dep} - t_{ij}^{dep} + \Delta td_{i'} + \Delta td_i + \max(\Delta tr_i, \Delta tr_{i'})$, which is the smallest possible $M_{ii'j}$ that can be chosen for this problem. The binary disjunctive variable $O_{ii'j}$ is used to automatically activate only one of the constraints depending on the value of the other variables. $O_{ii'j}$ has value 1 if train i departs before train i' at station j. This problem has on the order of $O(I^2J)$ binary variables and is very difficult to solve for large I (number of trains) or J (number of stations) due to a large integrality gap.

Similar constraints are included for inter-arrival times to ensure that the order of the trains is preserved between stations.

$$T_{ij+1}^{arr} - T_{i'j+1}^{arr} \geq h_{j+1}^{arr} - M_{ii'j}(O_{ii'j} + 2 - S_i - S_{i'}), \forall i, i', j : i < i', stat_{ij+1}, stat_{i'j+1} \tag{3.16}$$

$$T_{i'j+1}^{arr} - T_{ij+1}^{arr} \geq h_{j+1}^{arr} - M_{ii'j}(3 - O_{ii'j} - S_i - S_{i'}), \forall i, i', j : i < i', stat_{ij+1}, stat_{i'j+1} \tag{3.17}$$

For these constraints, a value of $M_{ii'j} = h_{j+1}^{arr} + t_{i'j}^{arr} - t_{ij}^{arr} + \Delta td_{i'} + \Delta td_i + \max(\Delta tr_i, \Delta tr_{i'})$ is used.

97

We emphasize that this formulation for shared railway systems differs in three aspects from traditional train timetabling problem formulations. First, it introduces the discrete variable S_i that indicates whether train i can be scheduled or not, which adds to the complexity of the problem. In contrast, the timetabling problem for a vertically-integrated railway will assume that all trains will be scheduled. Second, it uses flexibility margins $\Delta td_i, \Delta tr_i$ to alleviate conflicts. This is necessary because when different TOs are requesting trains, these conflicts are more likely to occur. Large flexibility margins result on high values of $M_{ii'j}$, making the problem hard to solve. Third, this formulation specifies safety constraints (spacing of the trains) for each train path, requiring the definition of the matrices $stat_{ij}, ini_{ij}, fin_{ij}, next_{ij'j}$.

This model is generalizable to other shared railway systems. The same equations will apply, with different parameter values to capture the system-specific information about the topology of the infrastructure, the path of the trains, the safe headways imposed by the signaling system, etc.

3.3 Linear Programming Approach for Approximate Dynamic Programming

As discussed above, the size of the MILP model proposed in Section 3.2 increases rapidly as a function of the number of stations and trains to schedule. We propose a novel solution algorithm using ADP techniques (Bertsekas and Tsitsiklis, 1996; Bertsekas, 2006; Powell, 2007) to tractably solve large timetabling problems in shared railway systems.

Specifically, we propose a Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm that extends the Approximate Linear Programming (ALP) and the

98

Relaxed Linear Programming (RLP) algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). QARLP introduces three main innovations with respect to ALP and RLP algorithms: 1) it incorporates the possibility of learning from previous solutions, allowing the algorithm to improve the solution obtained by refining the sampling strategy in subsequent iterations, 2) it formulates the Bellman equation using Q-factors, and 3) it implicitly samples through the state-action space, enabling the indirect identification of promising areas in the solution space, which is very difficult for large multidimensional problems. This approach decreases the solution time compared to a MILP commercial solver while still ensuring convergence to the optimal solution within a specified optimality gap.

3.3.1 Dynamic Programming Formulation

The problem defined in Section 3.2 can be reformulated as follows.

3.3.1.1 Stages

There are $i = 1, \dots, I$ decision stages (one for each train proposed to be scheduled), and a terminal stage $i = I + 1$.

3.3.1.2 State

The Markovian state variable is the timetable of the trains scheduled so far; that is, a matrix with the departure and the arrival times from/to the stations of all the trains scheduled so far:

$$x_i = \{timetable_{i-1}\}, \forall i \tag{3.18}$$

The timetable is defined as $timetable_{i-1} = \left[T_{i_1j}^{arr}, T_{i_1j}^{dep}; T_{i_2j}^{arr}, T_{i_2j}^{dep}; \dots\right]$, $\forall j$, $i_1, i_2, \dots : S_{i_k} = 1, i_k < i$.

3.3.1.3 Control

At every stage, the control variable indicates whether the IM decides to schedule train i or not, and, if scheduled, the specific timetable of train i at all stations j in the path.

$$u_i = \{S_i, T_{ij}^{arr}, T_{ij}^{dep}\} \in U(x_i), \forall i, j : i \, I, stat_{ij} \tag{3.19}$$

A train can only be scheduled if it does not present any conflict with the trains already scheduled. As we discuss in Section 3.4.1., the present and future value of scheduling each train ensures that the order in which the trains are visited (stages) does not affect the solution obtained.

3.3.1.4 State Transition Function

Given the state and the control at one decision stage, the state in the following decision stage can be computed, which incorporates the timetable of the new train if it is scheduled.

$$x_{i+1} = f(x_i, u_i) = \begin{cases} timetable_{i-1} & if \, S_i = 0 \\ [timetable_{i-1}; T_{ij}^{arr}, T_{ij}^{dep}] & if \, S_i = 1 \end{cases}, \forall i \leq I \tag{3.20}$$

3.3.1.5 Cost Function

The cost associated with a state-control pair is the sum of the penalties minus the revenue obtained if train i is finally scheduled. The sign of the cost function has been chosen to formulate a minimization problem. The cost associated with each state-action pair is evaluated using:

$$g(x_i, u_i) = g(u_i) = -a_i S_i + |\Delta TD_i| p_i^{\Delta td} + \sum_j \Delta TR_{ij} p_i^{\Delta tr}, \forall i \leq I \tag{3.21}$$

3.3.1.6 Bellman Equation

The policy that minimizes the sum of current and future costs at every decision stage can be determined by solving the Bellman equation and calculating the cost-to-go or value function:

$$J_i^*(x_i) = \min_{u_i} g(u_i) + J_{i+1}^*\big(f(x_i, u_i)\big), \forall i \leq I \tag{3.22}$$

$$J_{I+1}^*(x_{I+1}) = 0 \tag{3.23}$$

This equation can be reformulated using Q-factors, which represent the cost-to-go for every feasible state-control pair:

$$Q_i^*(x_i, u_i) = g(u_i) + \min_{u_{i+1}} Q_{i+1}^*(f(x_i, u_i), u_{i+1}), \forall i \leq I \tag{3.24}$$

$$Q_{I+1}^*(x_{I+1}, u_{I+1}) = 0 \tag{3.25}$$

The relation between the cost-to-go function and the Q-factor is:

$$J_i^*(x_i) = \min_{u_i} Q_i^*(x_i, u_i), \forall i \tag{3.26}$$

The optimal policy (timetable) can be determined solving the Bellman equation or the Q-factor Bellman equation using backward induction. However, when the dimension of the state space and/or the dimension of the control space increase, the solution of the exact DP program becomes impracticable because the size of the problem grows exponentially. The benefit of reformulating the MILP model as a DP problem is that we can apply efficient solution algorithms such as the one proposed later in this chapter.

3.3.2 Linear Programming Algorithm

(Borkar, 1988; De Ghellinck, 1960; Manne, 1960) show that solving the Bellman equation (3.22) is equivalent to solving the LP problem proposed in equation (3.27) for any positive vector c because the inequality $J \leq J^*$ holds for every feasible solution J of the problem. The vector c is called the state-relevance weight vector.

$$\max c\mathrm{J}, \text{ s.t. } g_i(u_i) + \alpha \mathrm{J}_{i+1}\big(f_i(x_i, u_i)\big) \geq \mathrm{J}_i(x_i), \forall i \leq I, x_i, u_i \in U(x_i) \tag{3.27}$$

Note that the original problem does not have any discount factor, so we will use

$\alpha = 1$ from now on. This LP problem has as many variables as possible states (value of

the cost-to-go function at each state) and as many constraints as possible state-control

pairs. When the state and control space of the problem are large, this results in a very

large number of variables and constraints.

(Schweitzer and Seidman, 1985; Farias and Van Roy, 2003) proposed a

modification of the previous formulation called the Approximate Linear Problem (ALP):

$$\max c\,\Phi\, r, \text{ s.t. } g_i(u_i) + \Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1} \geq \Phi_i(x_i)r_i, \forall i \leq I, x_i, u_i \in U(x_i) \tag{3.28}$$

where the real value function J_i^* is approximated by a linear combination of basis

functions $J_i(x_i) \cong \sum_{k=1,\dots,R} \Phi_{ki}(x_i)r_{ki} = \Phi_i(x_i)r_i$. In this approximation, there are only

$R \cdot I$ variables (number of basis functions and number of stages). However, the number of

constraints remains the same as in equation (3.27) (one constraint for each state-control

pair).

To reduce the number of constraints in this problem, (Farias and Van Roy, 2004)

proposed a Relaxed Linear Problem (RLP) formulation. RLP proposes a strategy which

samples constraints from the ALP formulation. Farias and Van Roy showed that for an

appropriate probability distribution function Ψ over the set of state-control pairs, the

number of constraints that must be sampled does not depend on the number of state-

control pairs. In particular, to obtain a solution close enough to the optimal solution

obtained using the ALP formulation with $1 - \delta$ confidence level ($\Pr\{|\|J^* - \Phi$

$r_{ALP}\|_{1,c} - \|J^* - \Phi r_{RLP}\|_{1,c}| < \epsilon\} \geq 1 - \delta$), the number of samples required is on the

order of a polynomial in the number of state variables, $1/\epsilon$, and $\log 1/\delta$. Note that these

convergence results are computed over the basis-function approximation, that is, the RLP formulation converges to the best approximation over the basis functions chosen with confidence level $1 - \delta$ within a number of samples that does not depend on the number of state-action pairs. The RLP formulation is:

$$\max c\,\Phi r, \text{ s.t. } g_i(u_i) + \Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1} \geq \Phi_i(x_i)r_i, \forall i \leq I, (x_i, u_i) \in X \qquad (3.29)$$

where X is the set of state-action pairs sampled.

The main drawback of the RLP formulation presented in equation (3.29) is that the convergence results proved in (Farias and Van Roy, 2004) are based on an idealized choice of the probability distribution used to sample the constraints. In particular, the choice assumes knowledge of an optimal policy. Although it is unrealistic to assume that the optimal policy is known a priori, it is possible to obtain a reasonable approximation of the optimal policy by solving the RLP. Applying this idea, we propose the following Adaptive Relaxed Linear Programming (ARLP) algorithm:

Step 0: Set $t = 0$, and sample X_0, giving each state-control pair equal probability to be sampled (Ψ_0 uniform distribution).

Step 1: Solve the problem $\max c\,\Phi r$,

s.t. $g_i(u_i) + \Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1} \geq \Phi_i(x_i)r_i, \forall i \leq I, (x_i, u_i) \in X_t$

Step 2: Set $t = t + 1$. Determine the optimal policy $\pi_{it}, \forall i$ according to the last problem solved $(\pi_{it}(x_i) = \arg\min g_i\big(\pi_t(x_i)\big) + \Phi_{i+1}\big(f_i(x_i, \pi_{it}(x_i))\big) r_{i+1}, \forall i \leq I, x_i)$. Choose the next set of constraints sampled using a probability distribution function Ψ_t that assigns higher probabilities to promising solutions considering the last iteration (i.e., the probability to sample (x_i, u_i) increases as $\Phi_{i+1}\big(f_i(x_i, u_i)\big)r_{i+1}/$

$\Phi_{i+1}\left(f_i(x_i, \pi_{it}(x_i))\right) r_{i+1}$ increases). In general, the variance of the probability

distribution Ψ_t will decrease with t.

Step 3: If $t > T_0$ or the difference between the objective function is smaller than ϵ, stop.

Otherwise go to Step 1.

Algorithm 3-1 Adaptive Relaxed Linear Programming (ARLP) algorithm

The ARLP algorithm iteratively solves a sequence of RLP problems, each with a

manageable number of variables and constraints. This approach takes advantage of the

reduced dimensionality of the RLP formulation while incorporating a mechanism to

refine the sampling strategy Ψ_t using the best approximation of the optimal solution

obtained so far. As a consequence, the convergence of the algorithm would not require

the knowledge of the appropriate probability distribution function Ψ a priori.

However, because the basis function approximation used reduces the

dimensionality of the problem, finding the state-control pair (e.g., timetable) that

corresponds to known basis function values becomes a challenge. In other words,

although it is very easy to determine the value of each basis function for a given state-

control pair, solving the inverse problem (determining a state-control pair associated with

a given basis function value) is extremely difficult in these cases. This is because the

basis functions are a projection from the higher-dimensional state-action space to a

lower-dimensional space, and the mapping from the low-dimensional projection back to

the higher dimensional space is underdetermined. Therefore, there is no straightforward

way to define Ψ_t based on low-cost regions in the basis function space, and to sample

state-action pairs from it.

We solve this problem by 1) reformulating the algorithm using Q-factors instead

of the cost-to-go function, and 2) determining Ψ_t implicitly using a Metropolis Hasting algorithm (Rubinstein and Kroese 2008) that accepts or rejects a state-control sample based on how promising the sample is according to the Q-factor best guess.

To do that, we sample uniformly across the state-action space to determine a candidate state-action pair $(x_i^{'}, u_i^{'})$, we compute the value of its associated Q-factor $(\Phi_i(x_i^{'}, u_i^{'})r_i)$, and compute the ratio of this Q-factor with the range of possible Q-factor values in the latest iteration ($\underline{\Phi_i(x_i, u_i)r_i}$, $\overline{\Phi_i(x_i, u_i)r_i}$):

$$\overline{\phi} = \frac{\Phi_i(x_i', u_i')r_i - \underline{\Phi_i(x_i, u_i)r_i}}{\overline{\Phi_i(x_i, u_i)r_i} - \underline{\Phi_i(x_i, u_i)r_i}} \tag{3.30}$$

We then draw a sample $\overline{\xi}$ from a probability distribution Ξ_t. We accept the state action pair if $\overline{\phi} \geq \overline{\xi}$. To ensure convergence to the optimal solution, the variance of the probability distribution Ξ_t must decrease with t, the probability of accepting any sample must be strictly positive, and the probability of accepting any sample with associated $\overline{\phi} \geq$ 1 must be 1. To ensure that these conditions hold, we determine the probability of accepting a proposed sample based on a sample drawn from $\Xi_t \sim U(-\frac{1}{t}, 1)$. In other words, we accept a sample with probability $\alpha = \min\left(\max(\overline{\phi}, 0) + \frac{1}{t}, 1\right)$. The performance of the algorithms improves when the state-control pairs associated with binding constraints in the previous iteration are retained in future iterations.

We call this algorithm a Q-factor Adaptive Relaxed Linear Problem (QARLP) algorithm.

Step 0: Set $t = 0$, and sample X_0, giving each state-control pair equal probability to be sampled (Ψ_0 uniform distribution).

Step 1: Solve the problem $\max c \Phi r$,

s.t. $g_i(u_i) + \Phi_{i+1}(f_i(x_i, u_i), u_{i+1})r_{i+1} \geq \Phi_i(x_i, u_i)r_i, \forall i \leq I, (x_i, u_i, u_{i+1}) \in X_t$

Step 2: Set $t = t + 1$. Determine the optimal policy according to the last problem solved.

Repeatedly sample state-control pairs until one pair is accepted if the ratio of its

associated Q-factor ($\Phi_i(x_i, u_i)r_i$) and the range of possible Q-factor values in the latest

iteration ($\overline{\phi}$) is greater than or equal to a draw of a probability distribution Ξ_t ($\overline{\xi}$).

Step 3: If $t > T_0$ or the difference between the objective function is smaller than ϵ, stop.

Otherwise save binding constraints from previous iterations and go to Step 1.

Algorithm 3-2 Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm

Note that in this algorithm c, Φ, r have slightly different meanings than in

Algorithm 3-1: c is a positive constant for every state-control pair at every decision stage

and Φ, r are functions of the state-control pair (not only of the state: $\Phi, r = \Phi$

$, r_i(x_i, u_i))$.

3.3.2.1 Capturing the problem structure: choosing basis functions

The choice of basis functions that capture relevant information about the state and

the action while at the same time decreasing the amount redundant information (and

hence the dimensionality of the problem) is a critical design choice of these types of ADP

algorithms. In this research, we use basis functions that capture: 1) the total number of

trains scheduled, as well as the total changes in the TO's desired timetable (state

variable), 2) whether train i is scheduled or not, and the total changes in its desired

timetable (control), 3) the number of conflicts of the trains scheduled so far with the

following trains to be scheduled, and 4) and a constant. This reduces the dimensionality

of the approximate cost-to-function to $R = 8$. That is,

$$\Phi_i(x_i, u_i) = (\sum_{i'<i} S_{i'}, \sum_{i'<i} \Delta TD_{i'}, \sum_{i'<i} \Delta TR_{i'}, S_i, \Delta TD_i, \Delta TR_i, C_i, 1) \qquad (3.31)$$

where C_i is the number of conflicts of the trains scheduled so far with the following

trains: $C_i = \sum_{i'>i} (Y_{i'})$. The variable $Y_{i'}$ has value 1 if the desired timetable of train i'

conflicts with the timetable of any train scheduled so far and 0 otherwise. These basis

functions capture the most relevant features of the problem, and therefore enable us to

achieve a low approximation error around the optimal solution and to differentiate

promising solutions from solutions that are less promising while reducing the

dimensionality.

3.4   Results

In this section, we present the results for the train timetabling problem from both

the MILP and the ADP formulations. Section 3.4.1 presents the computational results of

the chapter, comparing the solution times between the commercial MILP solver and the

QARLP algorithm. Section 3.4.2 illustrates the insights gained by using the model to

design different timetables for a shared railway system.

3.4.1   Computational Results

We begin by presenting the results obtained from solving the timetable problem

for a railway system with the infrastructure represented in Figure 3-2. It consists of a

double-track corridor with 12 stations. Stations 1 and 7 are terminal stations at both ends

of the line. Stations 2-12, 3-11, 4-10, 5-9 and 6-8 represent five stations along the

corridor. We use a different station number for each traffic direction. Traffic moves in the

direction of increasing station numbers in a dedicated track per direction. As a result,

traffic traveling in different directions only interacts at the stations. The system presented

includes the critical characteristics required to represent a corridor such as the NEC, for

which the FRA is currently developing a new capacity pricing and allocation mechanism to foster rail efficiency (Gardner, 2013).
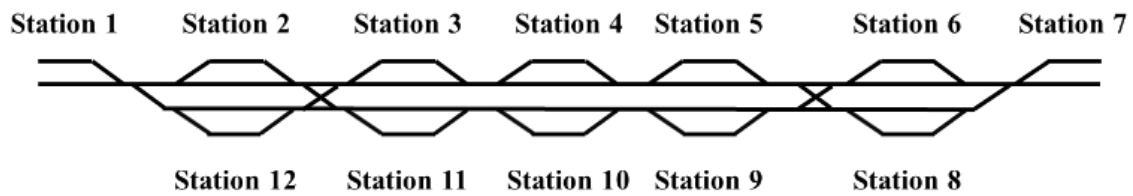
Figure 3-2 Detailed corridor infrastructure (Source: author)

Stations 1, 2 and 12 represent main stations in the same metropolitan area (e.g., Boston), stations 3, 4, 5, 9, 10 and 11 are all in another metropolitan area (e.g., New York), and stations 6, 7 and 8 are in yet a third distinct metropolitan area (e.g., Washington DC). Five types of services are considered: Boston commuter trains traveling around the Boston metropolitan area (stations 1, 2, and 12); New York commuter trains; DC commuter trains; and intercity and freight trains traveling between Boston and Washington DC. Intercity and freight trains may not stop at every station. Freight trains travel the line at speeds much lower than commuter and intercity trains. Intercity trains travel at higher speeds than commuter trains.

At present, around 2,000 commuter trains, 150 intercity trains and 70 freight trains travel around the NEC every day (Amtrak, 2010; Gardner, 2013). In practice, most of the conflicts to schedule trains occur around peak hours; where the IM would have to control for conflicts within sets of around 100-250 trains to make changes in the timetable.

We assume that the commuter TOs (one in each metropolitan area) request scheduling commuter trains every 30 minutes, and that one intercity TO requests operating a train every hour. The number of trains requested by TOs depends on the total

time horizon considered.

The MILP formulation from Section 3.2 is implemented in GAMS 24.1.2 and solved using CPLEX 12.5 on a PC at 2.40 GHz, 4GB, intel core i7, under Microsoft Windows 7 64 bits. To reduce the size of the problem, when the desired arrival and departure times of two trains are very far apart, the value of the binary variable $O_{ii'\ j}$ is fixed a priori (since the relative order in which they pass through the station cannot change). We run CPLEX with options $\text{CHEAT} = 0.05$, $R\text{INSHeur} = 50$, $\text{Threads} = -1$ for faster solution times, and use a 5% optimality gap. A smaller optimality gap may be required if the timetable problem has multiple near-optimal solutions with very different implications for different TOs in terms of which trains are scheduled to ensure that the IM's choice of the trains to schedule is not arbitrary. In practice, for the cases solved for this research, the difference in the objective function between scheduling one additional train or not is large. As a consequence any solution within a 5% optimality gap of the optimal solution ensures that the set of trains scheduled is the same as the set of trains scheduled in the optimal solution unless there are twin trains (TOs willing to pay the same to operate trains with the exact same timetable). In that case neither CPLEX nor the QARLP algorithm would be able to distinguish those trains in the solution and the choice of one solution over other would be random.

We then solve the identical problem using the QARLP algorithm proposed in the previous section. Although theoretically the relative order of the trains does not change the solutions obtained or the convergence speed of the algorithm to the optimal solution, in practice the relative order of the trains may speed up or slow down the process of finding the optimal solution. The results presented in this chapter correspond to cases in

which the relative train order (trains considered at each stage) was randomly assigned.

Table 3-1 shows the number of equations, variables, and discrete variables for problems with several different numbers of requested commuter and intercity trains. Figure 3-3 presents the execution time and the number of iterations required to convergence (within 5% integrality gap) for the MILP and QARLP algorithms. We compute the QARLP algorithm integrality gap using the "best possible" solution bound generated by the MILP commercial software.

Table 3-1 Train timetabling problem size for traffic patterns representative of the NEC Traffic (Source: author)

Number of Trains	Equations	Number of Variables	Discrete Variables
15	970	510	91
30	3,715	1,607	292
60	14,533	5,565	919
120	57,481	20,537	3,145

Note that the QARLP execution time increases as a polynomial function of the number of trains to schedule. In contrast, the MILP solution times increase exponentially. In fact, solving the MILP problem with CPLEX for 75, 90, or 120 trains within a 5% convergence gap is *computationally intractable*. Extrapolating from a regression estimate (Figure 3-3), the solution time using CPLEX for 120 trains would be approximately 46 days. The solutions obtained with the QARLP algorithm for 90 and 120 trains in approximately 20 minutes are better than those obtained for the MILP formulation with CPLEX after 20 hours and 35 hours respectively. In the cases with 15, 30, and 60 trains the solutions obtained with both methods are almost identical.
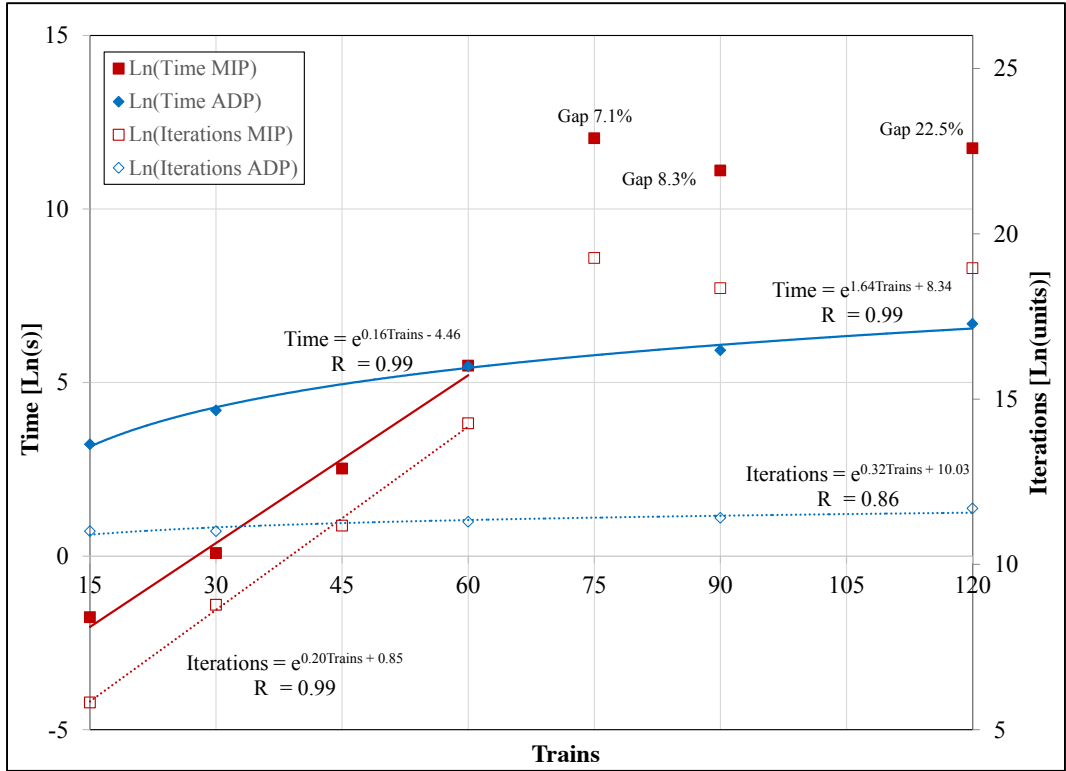
Figure 3-3 Comparison of the execution time and the iterations required to convergence (within 5% integrality gap) of the MILP approach and the QARLP algorithm (Source: author)

3.4.2 Design of Timetables for Systems with Traffic Patterns similar to the US Northeast Corridor's Traffic

In this section, we are interested in analyzing the timetables designed for relevant cases that illustrate trade-offs involved in the capacity planning process in shared railway systems. We analyze the optimal capacity allocation plan (train timetable) to determine how to coordinate different TOs' conflicting demand for scheduling trains. The ability to solve this allocation problem is critical for designing effective capacity pricing and allocation mechanisms. Figures 3-4, 3-5, 3-7, and 3-8 show time-space diagrams for timetables designed by the IM model for cases with different demands for accessing the infrastructure. The y-axes represent distance in miles from station 1 and the x-axes represent time in minutes at which different trains are scheduled to pass through each

point of the line (vs. desired scheduled in dashed line). The horizontal segments represent the stopping times at stations. We assume no interaction between trains traveling in different directions.

Figures 3-4 and 3-5 show the timetable for a case with demand for scheduling an intercity train in the system when commuter trains around the three metropolitan areas operate every 30 minutes. Figure 3-7 shows a case in which two competing intercity TOs request scheduling intercity trains when commuter trains around the three metropolitan areas operate every 30 minutes. These two cases provide information about how much intercity TOs will have to pay to be able to schedule services that conflict with commuter train services. Figure 3-8 shows a case with demand for scheduling a freight train in the system when commuter trains around the three metropolitan areas operate every 1 hour. This case is designed to analyze whether freight TOs would be able to schedule *any* trains on the infrastructure. The IM model proposes the final timetable analyzing the trade-off between eliminating trains and readjusting the desired schedules, according to the objective function in (3.1).

For this example, we assume that each commuter TO pays 1 unit to schedule a commuter service and gets a 3% discount from the original access charge for every minute that one of their train schedules is changed. To analyze the first case, we need to solve a train timetabling problem with 115 commuter trains and 1 intercity train. Figure 3-4 shows the timetable of all the trains scheduled. For clarity purposes, only the schedules of conflicting trains are shown in Figures 3-5, 3-7, and 3-8.

Note that when the IM tries to schedule the intercity train, it will initially conflict with 14 commuter trains (see Figure 3-5). Rescheduling the commuter trains to

accommodate the intercity service requires that the commuter TOs receive a discount of 2.1 units on their total access charges. As a result, the IM would only schedule the intercity train if it represents more than 2.1 units of revenue.

Figure 3-4 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes – including conflicting and non-conflicting trains (Source: author)

Figure 3-5 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes (Source: author)

If the frequency of commuter trains increases, for example to one commuter train every 15 minutes instead of every 30 minutes, the intercity train will initially conflict with 22 commuter trains and will only be scheduled if it represents more than 3.6 units of revenue for the IM (i.e., if the intercity bid is higher than 3.6 units). Conversely, if the frequency of commuter trains decreases to one train every 60 minutes, the intercity train will be scheduled if it represents at least 1.5 units of revenue for the IM. The model can be used to quantify the trade-off between commuter and intercity trains for any other frequency of service (see Figure 3-6). The exact value of the trade-off for low frequencies of commuter services depends on whether there are conflicts among the desired timetables of the trains or not. The results show that the price that an intercity or freight TO will have to bid to be able to schedule a train (minimum access charge) can vary considerably as a

114

function of the congestion of the line. This minimum intercity access charge reflects the congestion rent. The results show that greater cost recovery is expected in congested infrastructure.
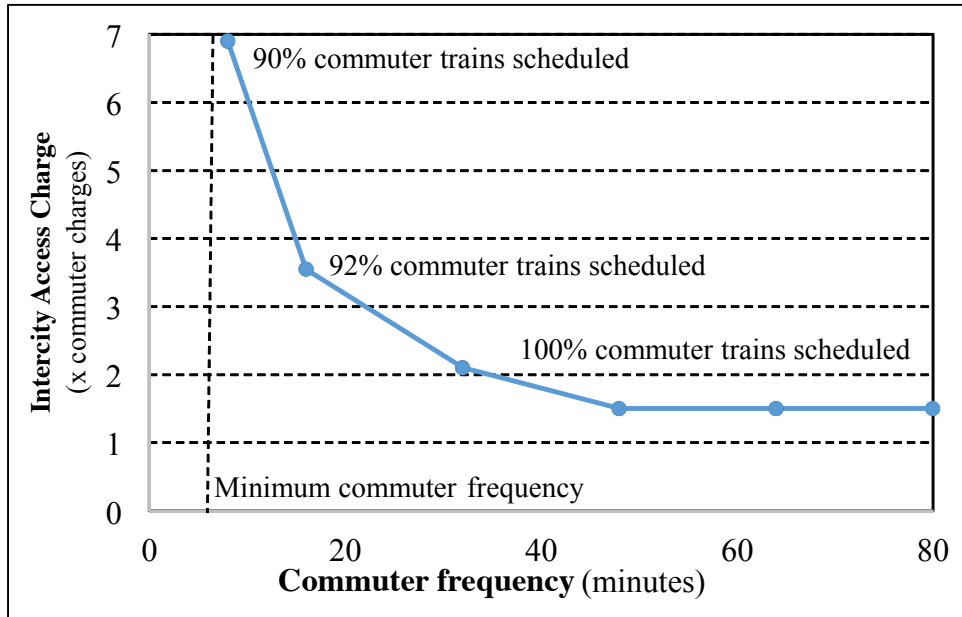
Figure 3-6 Intercity to commuter access charge ratio as a function of the commuter frequency
(Source: author)

We can use the same logic to analyze what would happen if two intercity TOs want to schedule intercity trains at the same time. This case also assumes that commuter TOs try to schedule commuter trains every 30 minutes. We assume again that each commuter TO pays 1 unit to schedule a commuter service and gets a 3% discount from the original access charge for every minute that one of their train schedules is changed. In simple auctions (like the one considered in this chapter), TOs reveal how much access charges they are willing to pay to schedule a train. This case is important because the TO's willingness to pay to access the tracks may change if a competing service is scheduled right before. According to Figure 3-7, scheduling two intercity trains would require changing the desired schedule of 14 commuter trains again. However, the changes in the

115

commuter timetable are much larger (because commuters are overtaken by two intercity trains) as compared to the previous case. As a result, the IM would only schedule the two intercity trains if they represent more than 4.0 units of revenue. If the revenue from scheduling the intercity trains represent between 2.1 and 4.0 units, at most one of the intercity trains would be scheduled.
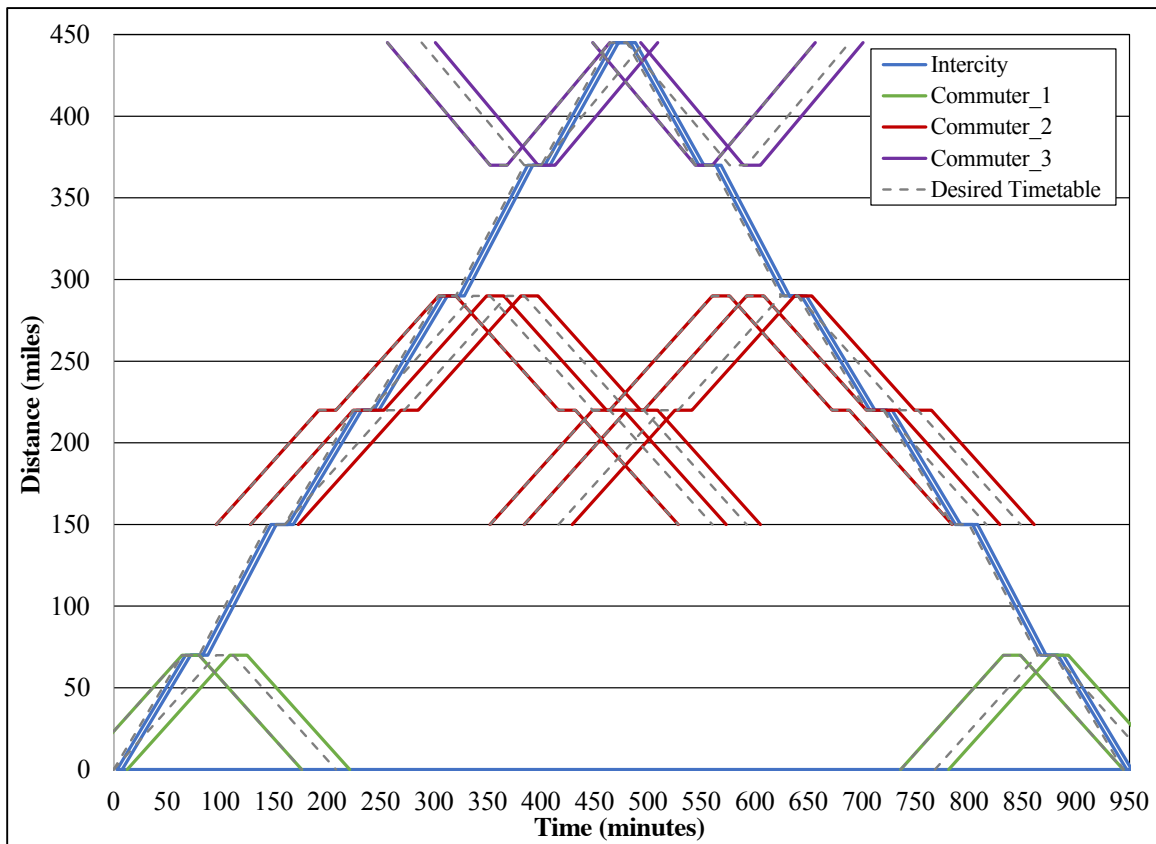
Figure 3-7 Timetable proposed by IM to schedule two intercity trains in a system with commuter trains operating every 30 minutes (Source: author)

Furthermore, note that although both trains would like to depart station 1 at minute 0, one of them will depart at minute 3 and the other one at minute 8. In some cases, none of the TOs may be interested in operating a second intercity service just 5 minutes after another one. These results suggest that intercity TOs may avoid getting their train scheduled just after other intercity train in simple auctions by controlling 1) how flexible

their schedule is, 2) how much discount in the access charge they request if the schedule of the train is changed, and 3) how much they are willing to pay to access the infrastructure. This is important because the railway literature assumes that the value of scheduling trains for TOs can only be captured by combinatorial auctions (Perennes, 2014). However, these results demonstrate that TOs can also avoid getting their trains scheduled right after other competing service in congested systems even when using simple auctions.

The third case considered in this section analyzes whether a TO would be able to schedule trains in the system if it cannot afford to pay high access charges. This discussion can be particularly relevant to understand if freight TOs with low access charges willingness to pay may be able to access the infrastructure in shared railway systems. Figure 3-8 shows that a freight train could be scheduled paying the same access charges as commuter trains (with frequency one train pair hour) if the freight TO is very flexible (in terms of the total allowed translation and increment of travel time it accepts).

The minimum access charge that a freight TO must pay when the line is more congested will depend on how many trains have to be rescheduled to eliminate conflicts. If the commuter TO wants to increase the frequency of commuter service from one train per hour to one train every 30 minutes, the freight train will only be scheduled if the net access charge that the freight TO is willing to pay represents more than 3 units of revenue for the IM (since three commuter services could not be operated). In general, the relative speeds among different types of services have a major impact on the capacity utilization of the system.
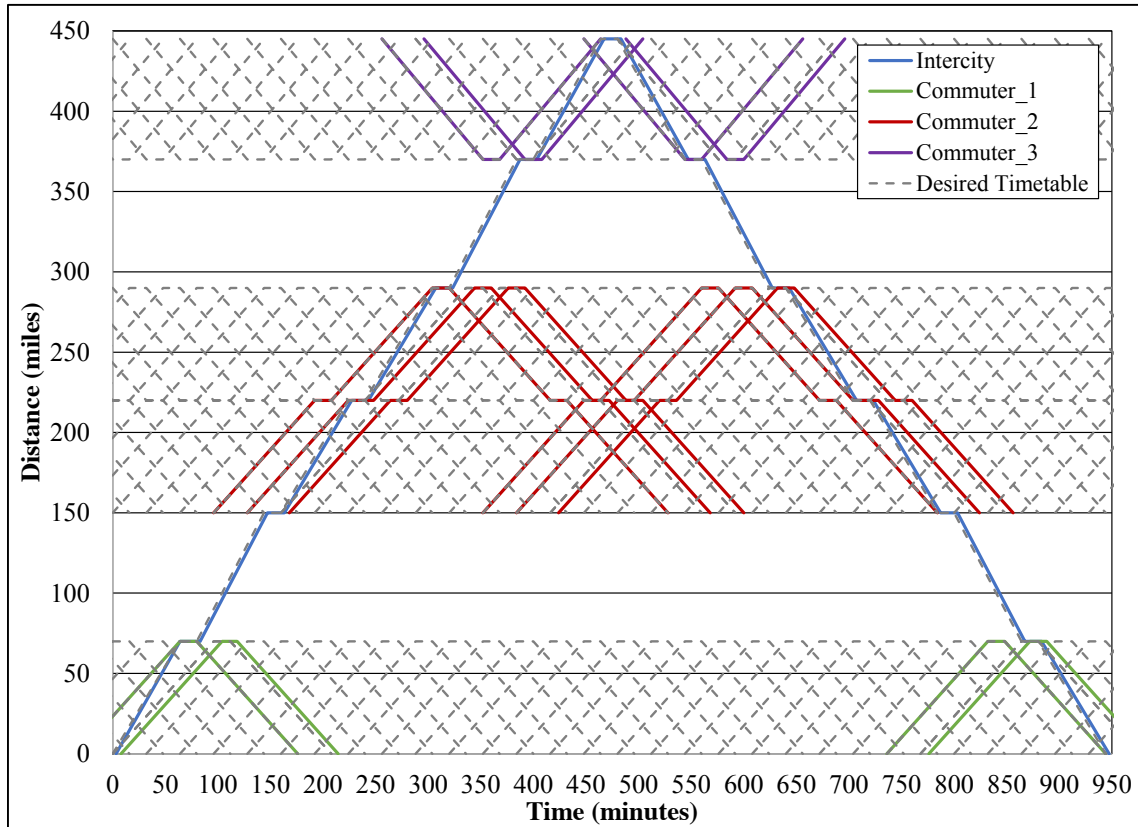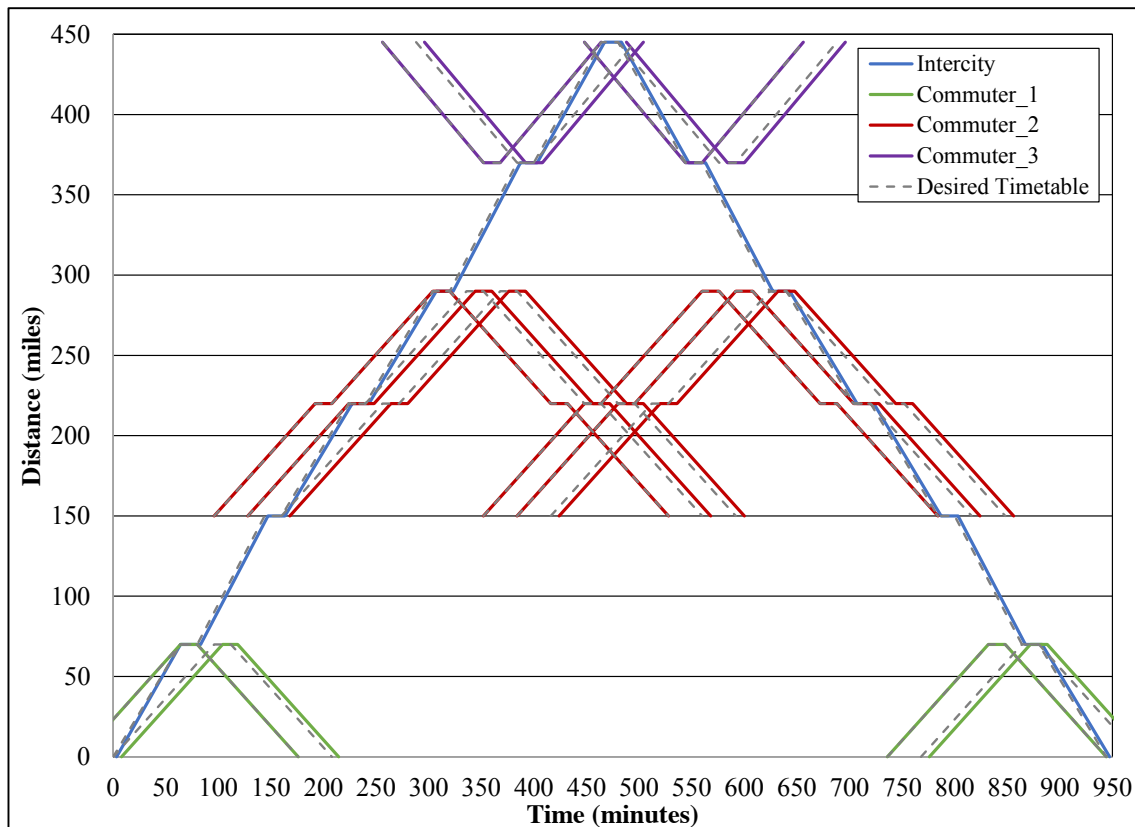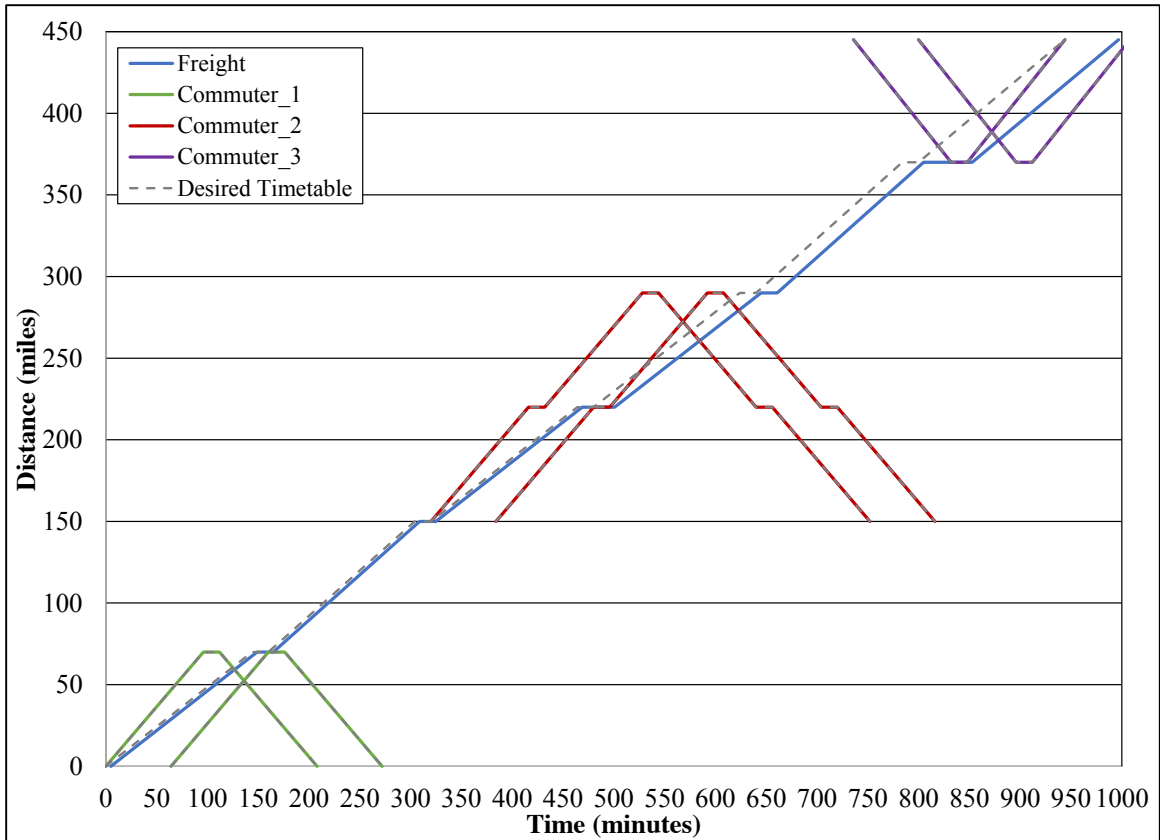
Figure 3-8 Timetable proposed by IM to schedule a freight train in a system with commuter trains operating every 60 minutes (Source: author)

The results in these cases show that intercity and freight services would have to either pay higher access charges or to be more flexible to compete with commuter trains. The exact amount (access charges or flexibility needed) depends on the frequency of commuter services. This strong dependency on the operational variables demonstrates that we need to consider the interactions between capacity planning and operations to evaluate capacity pricing and allocation mechanisms.

3.5   Conclusions

This chapter proposes a train timetabling model for shared railway system that explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths. The model is formulated as both a MILP problem and as a DP

problem. The MILP is solved using commercial software and the DP is solved using a novel algorithm for ADP. The timetables designed with the model are used to evaluate how capacity pricing and allocation may impact different railway system stakeholders. As a result, the contributions of this chapter are both methodological and domain specific.

On the methodological side, the main contributions of the chapter include:

1) The formulation of a train timetabling model for shared railway systems that would allow regulators and decision makers to determine the optimal use of railway infrastructure capacity. This model explicitly considers a variable number of trains, with large flexibility margins, traveling along different paths to analyze the interdependencies between operations and available infrastructure capacity and how they affect the coordination between the TOs and the IM.

2) The development of a novel algorithm for rapidly solving the train timetable problem in shared railway systems, ensuring convergence to the optimal solution within a specified optimality gap. We obtain solutions within 5% of the optimal solution for problem sizes that cannot be solved within a 5% convergence gap using commercial MILP software.

3) The algorithm developed, a Q-factor Adaptive Relaxed Linear Programming (QARLP) algorithm, extends the Approximate Linear Programming (ALP) and the Relaxed Linear Programming (RLP) algorithms developed by (Farias and Van Roy, 2003; Farias and Van Roy, 2004). QARLP introduces three main innovations with respect to ALP and RLP algorithms: 1) it incorporates the possibility of learning from previous solutions, allowing the algorithm to improve the solution obtained by

119

refining the sampling strategy in subsequent iterations, 2) it formulates the Bellman

equation using Q-factors, and 3) it implicitly samples through the state-action space,

enabling the indirect identification of promising areas in the solution space, which is

very difficult for large multidimensional problems. These ideas can be generalized to

efficiently solve other large-scale network optimization problems.

Moreover, the results of the train timetabling model can be used to simulate and

evaluate the best possible behavior of the IM in shared railway systems under different

capacity pricing and allocation mechanisms. The domain-specific contributions of this

chapter are:

4) The modeling framework and the algorithm developed in this chapter enable

us to simulate optimal decisions by an IM for shared railway systems. These results

can be used to answer relevant policy-type questions to understand the implications of

infrastructure shared use.

5) This chapter also shows that the implications of capacity planning

mechanisms depend on the characteristics of the system and the TO demand for

accessing the infrastructure. We propose the use of this model as a tool to allow

regulators and decision makers to better understand the interactions between capacity

planning and operations under alternative capacity pricing and allocation

mechanisms.

This chapter considers the TOs' infrastructure access demand (characterized both

as the demand for scheduling trains and the revealed willingness to pay to access the

infrastructure) as exogenous to the problem. However, the TO's infrastructure access

demand depends on the capacity pricing and allocation mechanism. Section 5.1 (Chapter

5) discusses the integration of the IM model proposed in this chapter with the model of TO bidding behavior developed in Chapter 2 to better quantify the trade-offs between utilization and level of service on the one hand, and infrastructure cost recovered under different capacity pricing and allocation mechanisms. The rest of Chapter 5 uses that framework to analyze alternative capacity pricing and allocation mechanisms in the context of the NEC. These results are valuable to design and evaluate alternative capacity pricing and allocation mechanisms to effectively coordinate the TOs and the IM in shared railway system.

Chapter 4 - Policy Implications for the Central Corridor in Tanzania and Other Shared Railway Systems with Infrastructure Cost Recovery Constraints

"If you want to travel fast, travel alone. If you want to travel far, travel together."
– African Proverb

In 2013, Tanzania's government committed to the implementation of one of the first shared railway systems in Africa (Big Results Now, 2013) as a way to ensure adequate level of rail service by 1) allowing efficient train operators (TOs) to access the infrastructure and operate train services through an open-access policy, and 2) providing sustainable resources through access charges to maintain the infrastructure and keep the system operative in the future. These objectives are critical to prevent future railway systems failures such as the 2001 and 2006 Tanzanian railway system concessions failures (Olievschi, 2013) that resulted in a major underinvestment in rail transportation in the country (Railistics, 2013). This underinvestment critically impacted the operating capacity and the reliability of the railway system, essential to improving accessibility to the East African landlocked countries: Rwanda, Burundi, Uganda, and Western Democratic Republic of Congo (AICD, 2008; Amjadi and Yeats, 1995; Arvis et al., 2010; Raballand and Macchi, 2009).

The implementation of a shared railway system requires new railway regulations that clarify the roles and responsibilities of railway institutions (Railistics, 2013; World Bank, 2014) as well as the design and implementation of a new mechanism to price and allocate railway capacity. This chapter analyzes how alternative capacity pricing and allocation mechanism for freight TOs would affect the performance of the Central

Corridor in Tanzania. We are particularly interested in the number of trains operated in the system and the revenues collected to maintain the infrastructure and recover capital costs. Chapter 2 shows that traditional approaches to price and allocate railway capacity may not work in two cases: 1) when the infrastructure manager (IM) needs to recover part of the infrastructure management fixed costs or 2) when the railway system is congested. Tanzania's Central Corridor falls into the first case. As mentioned above, one of the main purposes of the introduction of shared use in Tanzania is to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational.

An important characteristic of the Central Corridor is that it is not congested. The only current TO, TRL, operates around six trains per week, leaving plenty of spare capacity that could be used by other TOs (World Bank, 2014). As a result, allocating capacity is fairly easy and we can solve the capacity pricing problem independently of the capacity allocation one. In other words, we can use the TO Model to determine the TOs' demand for scheduling trains on the infrastructure. This process can be done for each type of service independently, since there is enough infrastructure capacity to accommodate the demand of the all TOs (Pena-Alcaraz et al., 2014).

The results of this chapter show that the introduction of variable access charges distorts the operational decision of TOs, as predicted in Chapter 2. We then discuss how to avoid this problem with other pricing mechanisms such as the introduction of fixed access charges. We also discuss how to allocate fixed access charges among multiple types of freight TOs and show the need for price discrimination in this context. The results also show that it is not possible to recover infrastructure costs from dedicated

container or general freight traffic in the context of the Central Corridor. However, the shared use of the infrastructure by container and general freight TOs allows the IM to fully recover infrastructure costs. This is one of the benefits of shared railway systems. This work is published in Network Industry Quarterly (Pena-Alcaraz, Perez-Arriaga, and Sussman, 2014).

The rest of the chapter is structured as follows: Section 4.1 presents the main types of capacity pricing mechanisms and discusses how the TO Model presented in Chapter 2 can be used to determine the behavior of TOs under each mechanism. Section 4.2 presents the resulting number of trains that container and general cargo freight TOs would operate under alternative capacity pricing mechanisms. Section 4.3 concludes with some recommendations for capacity pricing mechanisms in shared railway systems with infrastructure cost recovery constraints.

4.1   Capacity pricing mechanisms for shared railway system

The Central Corridor goes from the port (Dar es Salaam) to an inland container terminal (Isaka) that serves as a dry port for Rwanda, Burundi, Uganda, and the Eastern portion of Democratic Republic of Congo (see Figure 4-1). The infrastructure is owned by RAHCO, a publicly owned company. TRL is the only current TO; it operates around six trains per week. Although the corridor is single track, there is plenty of spare capacity that could be used by multiple private companies that have expressed interest in starting operating new services between Dar es Salaam and Isaka (Pena-Alcaraz et al., 2014; World Bank, 2014).

As we discussed in Chapter 1, the implementation of a shared railway systems requires some level of vertical separation between the TOs that operate the trains in the

system and collect the revenues selling transportation services to the final customers and the IM that maintains and manages the infrastructure. Vertical separation requires the definition of a capacity pricing mechanism that determines the access charges that TOs pay to the IM to access and use the infrastructure (Gomez-Ibanez, 2003). The IM uses these revenues to cover infrastructure costs. The use of the state national budget to cover shortfalls is the last resort.
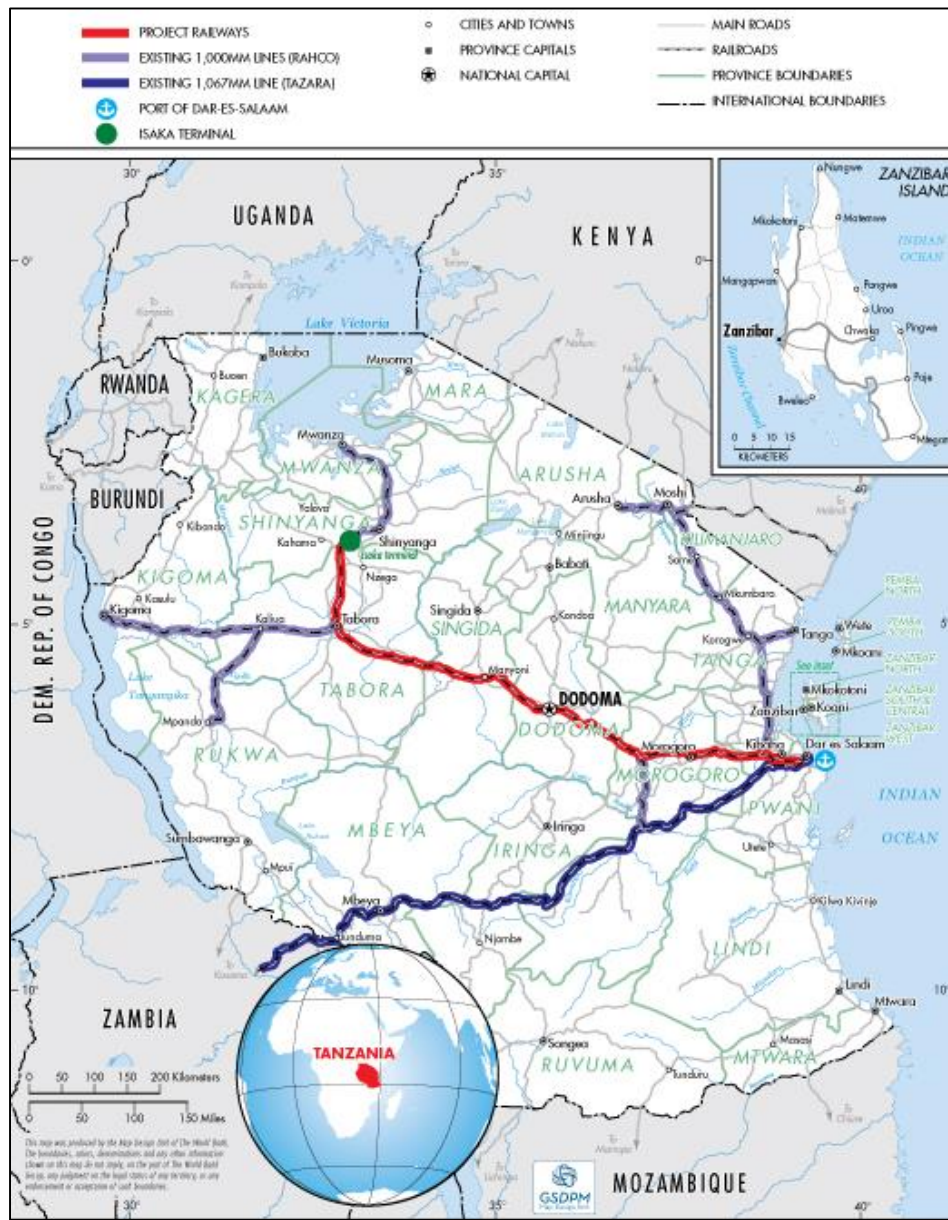
Figure 4-1 Central Corridor (Source: United Nations, 2006, Map No. 3667 from (World Bank, 2014))

The railway literature proposes capacity-based and price-based (also called cost-based) mechanisms to price railway capacity (Gibson, 2003). Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity. However, in cases like the Central Corridor with plenty of excess capacity, the TOs would be able to access the infrastructure paying very low access-charges. As a result, capacity-based mechanisms are not an option to recover infrastructure costs in non-congested systems. There are three cost-based capacity pricing mechanisms designed to allow the IM to recover maximum infrastructure costs: variable access charges, two-part tariffs (variable access charges plus a fixed access charge), and fixed access charges (Gibson, 2003). Under variable access charges, TOs pay some amount per train operated; the charge is in general a function of the type of train, distance, and tonnage. Under fixed access charges, each TO pays an annual lump sum to have a license to operate, regardless of the number of trains the TO operates during the year.

The practice and the broad economic literature in the field recommend the use of variable access charges based on marginal cost plus mark-ups (DB, 2009; Lopez-Pita, 2014; Nash, 2005; World Bank, 2014). However, from an engineering standpoint, infrastructure-related costs in Tanzania are mostly independent of the number of trains. Due to the low number of trains that operate in the system, the infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. As a result, maintenances costs are assumed fixed. In other words, the short-term and long-term infrastructure marginal costs are very low and high mark-ups are required to recover

infrastructure costs. This research analyzes the implications of resulting alternative pricing mechanisms for the system.

For this analysis, we compare the behavior of vertically-separated TOs with the behavior of an integrated railway company (social planner). We assume that both the vertically-separated TOs and the integrated railway company are rational agents, i.e. they determine the number of services per direction per week to operate by maximizing the annual operating margin (operating profits). A vertically-separated TO would only be interested in operating trains if the average annual net cash flow is positive after remunerating any invested capital at an adequate rate of return (no operations subsidies).

We use the financial TO Model developed in Chapter 2 to determine the integrated railway company, vertically-separated TO, and vertically-separated IM's operating margin and cash flow for a representative year under different levels of service. See Chapter 2 for more details about the TO Model and (Pena-Alcaraz et al., 2014; PPIAF et al., 2011; World Bank, 2014) for detailed model assumptions. The integrated railway company faces capital costs associated with the investments in railway infrastructure, variable costs of operating trains (train lease, personnel, fuel), and obtains revenues from transporting freight. The vertically-separated case is similar: the TO faces cost of accessing the tracks (access charges), variable costs of operating trains, and obtains revenues from transporting freight. The IM faces investment costs in railway infrastructure, maintenance costs, and obtains revenues from access charges.

Investment in railway infrastructure includes $300 million investment required to rehabilitate the current Tanzanian railway system (CPCS, 2013; World Bank, 2014) plus periodic investment in maintenance. The revenues of the TOs are determined multiplying

the cargo transported (minimum between the capacity of the trains operated and the demand) by the freight shipping rate. Due to the strong competition from trucks that offer door-to-door transportation services, railway companies have an upper limit on the freight shipping rate they may charge and they have low control over the demand that would likely shift to rail. The state should facilitate strong intermodal integration with the port and with truck companies that provide last mile transportation to/from the terminal rail station to make rail transportation more attractive and increase the utilization of the highly underused railway capacity. All the financial information used in this analysis is publicly available (World Bank, 2014).

4.2   Discussion of the Results

In this section, we discuss the main results obtained for alternative capacity pricing mechanisms designed to recover maintenance and financial infrastructure costs and to ensure that TOs can viably operate (positive profits) in Tanzania in two scenarios: 1) considering only container TOs, and 2) considering both container and general cargo (non-containerized freight) TOs.

4.2.1   Container Traffic

Figure 4-2 shows the annual operating margin and the cash flow for a vertically-separated container TO, for the IM, and for an integrated railway company in Tanzania under variable and fixed access charges when no other type of TO operate in the line. Both access charges have been calculated to recover as much of the infrastructure costs as possible, while ensuring that the operating margin and the net cash flow of the vertically-separated TO are positive. Note that it is not possible to recover all the infrastructure cost ($22.9 million per year in Tanzania) only with container services. The maximum charges

that a vertically-separated TO could viably pay are \$0.035 per ton-km (variable, assuming the TO would operate four trains) or \$19.1 million per year (fixed). We compute these numbers estimating the TO maximum revenues, the variable and fixed costs, and therefore the maximum fixed and variable access charges that the TO can viably pay to achieve an annual net cash flow equal to zero.

The results also show that under variable access charges only, a rational vertically-separated TO would only operate two trains per direction per week while the social planner would operate four. This mismatch happens because when the social planner tries to maximize its operating margin, it operates a train when the additional revenues produced are higher than the additional variable costs (train lease, personnel and fuel). For the social planner, most infrastructure investment cost is a sunk cost: it is already made and it is independent of the number of trains operated in the system. Under variable access charges in contrast, the infrastructure costs are charged as variable costs for TOs. Therefore, a rational TO would only operate a train if the additional revenues produced are higher than the true variable costs plus a share of the infrastructure cost that appears now as an artificial variable cost (the variable access charge).

Under fixed access charges, the infrastructure costs are charged as a fixed cost for TOs. Therefore, this cost will also be a sunk cost for the TO. Consequently, the TO will operate a train when additional revenues produced are higher than the true variable operational costs and there is no mismatch with the number of trains operated by the social planner.

Variable Access Charges *($0.035 per ton-km access charges for container TO)*
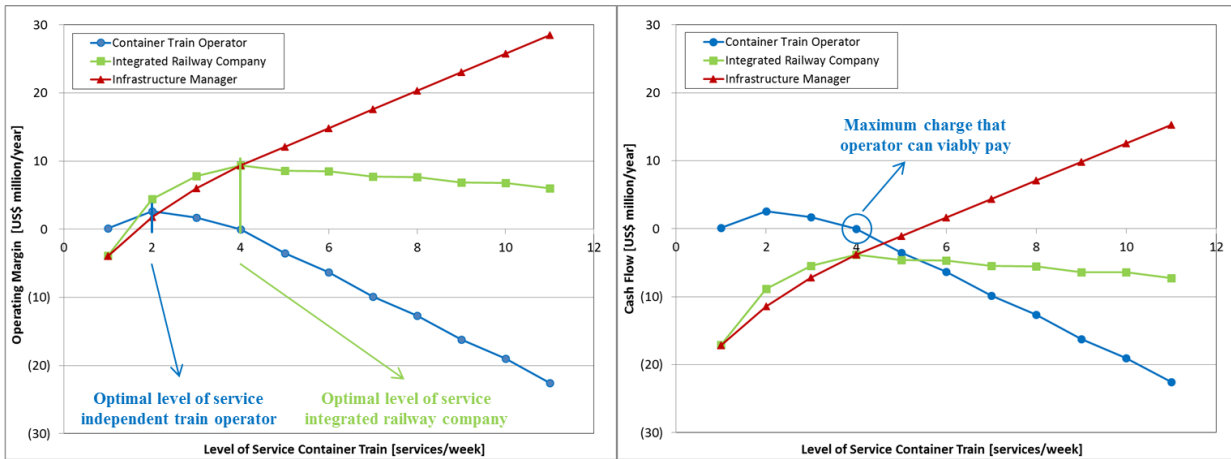
Fixed Access Charges *($19.1 million annual access charges for container TO)*
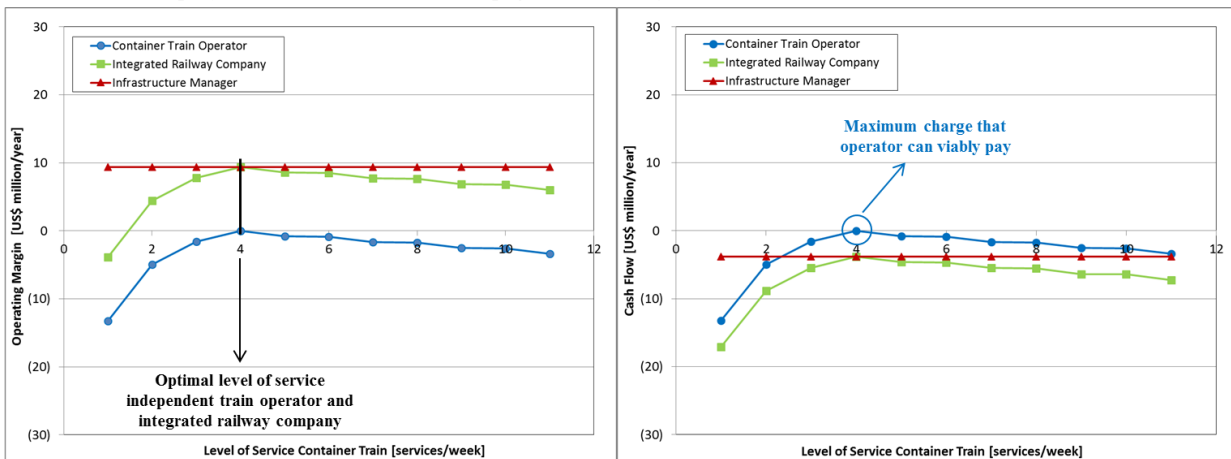
Figure 4-2 Operating margin and cash flow for different levels of service with variable and fixed access charges (Source: author)

4.2.2   Container and General Cargo Traffic

The previous section considers container traffic because container shippers have high willingness to pay to ship containers. Nonetheless, there is plenty unused capacity in the Tanzanian railway system and there are other types of customers interested in transporting non-containerized freight (general cargo) along the corridor. We carried out a similar analysis of costs and revenues for general cargo services (World Bank, 2014) 0.010 per ton-kilometer (variable) or $10.5 million per year (fixed). In both cases an

130

integrated railway company and a vertically-separated TO would operate ten services per week.

Considering these numbers, the IM would need to charge a variable access charge of $0.023 per ton-kilometer (variable, assuming the TO would operate four services) or $12.4 million per year (fixed) to the container TO to recover all infrastructure costs. Note that if the container TO was charged only $10.5 million per year or $0.010 per ton-km it would not be able to recover infrastructure costs (only $21.0 and $15.9 million per year respectively, assuming that a TO would operate four trains with access charges of $0.010 per ton-km). This shows, first of all, that discriminate pricing would be needed to recover infrastructure costs. Although a general cargo TO cannot viably pay as much as a container TO per ton to access the infrastructure, allowing access to the infrastructure to general cargo TOs 1) allows the IM to recover infrastructure costs (not possible only with container TOs), 2) allows container TOs to pay lower charges to access the infrastructure, and 3) improves welfare (for general cargo TOs and general cargo shippers) from a state point of view.

Although these charges are consistent with the industry benchmark (World Bank, 2014), a regulator needs considerable information (operational costs, demand estimates) to determine the maximum access charges that each TO is able to pay. Lower charges would not allow the IM to recover infrastructure costs; higher charges (particularly for general cargo in this case) would not allow TOs to viably operate trains in the system.

With a variable access charge of $0.023 per ton-kilometer, a vertically-separated container TO would only operate three (note that the variable charges are now lower than in Section 4.3.1) train services per direction per week (instead of the four that a social

planner would operate). Under fixed access charges, the number of trains operated by vertically-separated TOs in equilibrium matches the number of trains that an integrated railway company would operate. The main challenge to implement fixed access charges in this case consists of determining the share of infrastructure costs ($22.9 million per year) that each TO should pay. Nonetheless, our computation shows that the number of trains operated by the TOs is robust when the distribution of fixed access charges change: the container and the general cargo TO would be able to pay up to $19.1 million and $10.5 million per year respectively while still being profitable. Any choice such that the annual fixed access charge for the container TO is lower than or equal to $19.1 million, for the general cargo TO is lower than or equal to $10.5 million, and the sum of both charges is $22.9 million would improve number of trains with respect to variable charges while enabling infrastructure cost recovery. This result has important implications: 1) it relaxes the constraint on how much information the regulator needs to determine fixed access charges, and 2) it allows the regulator to design the fixed charge level for TOs with different objectives: such as ensure equity, ensure efficiency, ensure general cargo services.

Under fixed access charges with no variable charges per train, states could implement different mechanisms to allocate operating licenses among potential TOs. First, the regulator could determine a fee (fixed access charge) that a container and a general cargo TO would have to pay to get the license to recover infrastructure costs ($22.9 million per year). If the charges allow the operators to viably operate, they would apply for the license and retain the additional profits ($19.1 or $10.5 million per year minus access charge for each type of TO). Second, when there are several companies

willing to operate trains, the state could implement an auction to allocate the license to operate in each market. If the license is awarded to the TO with higher willingness to pay at each market, the most efficient container and general cargo TOs would bid $19.1 and $10.5 million respectively. In this case, the publicly owned IM would obtain $29.6 million per year (instead of $22.9). The IM can either use the additional revenues to invest in infrastructure in the future or transfer them to the government. If the license is awarded to the TOs that offer best freight shipping rate to customers provided that the IM can recover infrastructure costs, the IM would recover $22.9 million per year, the TOs would recover their costs with some return, and the customers would benefit from a discount in their shipping rate of $6.7 million per year ($29.6 minus $22.9). Table 4-1 summarizes these results. Further options can be explored when more than one license per market (container and general cargo) are allocated.

Table 4-1 Number of container and general cargo trains operated by TOs and revenues raised by IM for different variable and fixed access charges. These numbers are compared with the reference number of trains (operated by a social planner) and the IM costs (Source: author)

variable access charges							
container TO			general cargo TO			IM	
ac_v [$ ton-km]	n [trains]	n reference [trains]	ac_v [$ ton-km]	n [trains]	n reference [trains]	revenues [$m]	costs [$m]
0.000	4	4	0.000	10	10	0.0	22.9
0.010	3	4	0.010	10	10	14.5	22.9
0.023	3	4	0.010	10	10	19.8	22.9
0.030	2	4	0.010	10	10	18.6	22.9
0.046	1	4	0.010	10	10	16.7	22.9
0.092	1	4	0.010	10	10	22.9	22.9
fixed access charges							
container TO			general cargo TO			IM	
ac_f [$m]	n [trains]	n reference [trains]	ac_f [$m]	n [trains]	n reference [trains]	revenues [$m]	costs [$m]
up to 19.1	4	4	up to 10.5	10	10	22.9	22.9

4.3   Conclusions

In this chapter, we analyze different capacity pricing mechanisms designed to recover infrastructure costs (periodic maintenance and financial costs) and to ensure that

TOs can viably operate (positive net cash flow) in Tanzania. The insights derived from this case are useful to design capacity pricing mechanisms for shared railway systems with infrastructure cost recovery objectives in other countries.

First of all, we show that the adoption of variable access charges widely used in the railway industry may create incentives for rational TOs to operate fewer trains than an integrated railway company (social planner). This is consistent with the results of Chapter 2. We show that the use of fixed access charges aligns the behavior of vertically-separated firms with the behavior of an integrated railway company. This result is important in the railway industry because IMs faced important fixed costs, i.e., for the most part infrastructure costs do not vary with the number of trains operated in the system as is generally assumed.

The results obtained also show that discriminate pricing may be needed to be able to recover infrastructure costs when different types of TOs face very different levels of cost and revenues. The results also show the benefits of sharing the infrastructure among different types of TOs: shared use allows for infrastructure cost recovery.

This case also illustrates that regulators need considerable information about the sector (demand and cost) to determine adequate access charge levels that TOs can viably pay. A benefit of introducing fixed access charges is that the number of trains operated by TOs is robust for a wide range of fixed access fees. This relaxes the regulator information need. The ability to achieve a good level of service with a wide range of fixed access charges also allows the regulators and IMs to design effective pricing mechanisms with very different objectives and with very different implications in terms of the welfare distribution among stakeholders. This chapter also discusses why the use of capacity-

based pricing mechanisms would not allow the IM to recover infrastructure costs in non-congested railway systems.

Future work should analyze further how to implement fixed access charges effectively, especially in cases with competing TOs in the same market to avoid barriers to entry. Future research should also determine how these conclusions change with demand uncertainty, elasticity in the demand, and imperfect information.

In this chapter we are able to analyze the capacity pricing problem independently of the capacity allocation problem because the Central Corridor in Tanzania is not congested. Chapter 5 explores the performance of shared railway systems under alternative capacity pricing and allocation mechanisms in instances in which infrastructure capacity is limited and there are important interactions between capacity planning and infrastructure operations.

Chapter 5 - Policy Implications for the Northeast Corridor in the US and Other Congested Shared Railway Systems

"Traffic congestion is caused by vehicles, not by people in themselves."–
J. Jacobs (2002), The Death and Life of Great American Cities

This chapter focuses on the main spine of the Northeast Corridor (NEC) that stretches from Boston, MA to Washington, DC. With over 2,000 trains per day, the NEC is one of the most congested railway corridors in the US. Until now, capacity pricing and allocation in the corridor is managed via bi-lateral contracts. The price that each train operator (TO) pays to access the infrastructure and schedule their trains depends mostly on how much capacity was available when the contract was signed (Gardner, 2013). This imposes two challenges in today's operations: 1) the revenues collected by the infrastructure manager (IM) represent a very small percentage (10%) of the costs in which the IM would need to incur to bring the infrastructure to a state-of-good-repair, and 2) the introduction of new services is extremely complicated. Even if the timetable of some train could be shifted to make room to schedule new trains, rescheduling those trains would require the renegotiation of the contracts. As a result, the FRA required Amtrak and the rest of the NEC commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism (PRIIA, 2008).

NEC stakeholders face two important questions: Which mechanism to price and allocate railway capacity should they implement? What are the implications of such mechanism for each of them and for the overall performance of the system? Chapter 2 points out that in a congested railway systems like the NEC, traditional mechanisms that charge marginal infrastructure costs to TOs and impose simple priority rules to overcome

conflicts may not work. Chapter 1 shows that the implications for the system of alternative mechanisms to price and allocate capacity are still unclear. Furthermore, the interactions between infrastructure planning at the strategic level and infrastructure operations at the tactical level are particularly strong in congested railway systems.

In this chapter, we utilize the framework developed in this thesis to evaluate the performance of the NEC considering both planning and operational aspects. As described above, this framework consists of two models: a TO Model presented in Chapter 2 and an IM Model presented in Chapter 3. The TO Model simulates the behavior of the TOs to determine their demand for scheduling trains on the infrastructure and their willingness to pay for access. The IM Model determines whether that demand can be scheduled in the existing infrastructure. The results obtained are the demand for scheduling trains, the access charges (capacity pricing), and the optimal train timetable (capacity allocation: set of trains scheduled and their timetable).

We then use this information to analyze and compare the performance of a case based on the NEC under two alternative capacity pricing and allocation mechanisms: a price-based cost-allocation and priority-rule mechanism proposed by Amtrak (Crozet, 2004; Gardner, 2013; Nash, 2005; NEC Commission, 2014; Lopez-Pita, 2014; Texeira and Prodan, 2014) and an auction mechanism widely proposed in the railway economic literature (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014). To understand the implications for various stakeholders, we measure performance from the perspective of the IM (cost recovery), the TOs (access charges, trains scheduled), and the end users (number of services, fares). This chapter focuses on the interactions between

intercity and commuter TOs that are responsible for most of the traffic during the very congested peak hours.

The results of this chapter show that there are important trade-offs associated with each these two mechanisms and none of them is superior to the other on all dimensions. We argue that the trade-offs observed cannot be explained solely by the simplifications of this case study. As a result, we recommend NEC stakeholders that they analyze the implications of alternative pricing and allocation mechanisms in detail before locking the system into one of them. Part of this work has been submitted for publication (Pena-Alcaraz, Sussman, Webster, and Perez-Arriaga, 2015b).

The rest of the chapter is structured as follows: Section 5.1 describes the framework inputs and outputs, and discusses how to integrate the TO and the IM Models. Section 5.2 and 5.3 present the results obtained using that framework to evaluate the performance of the NEC under both price-based and capacity-based mechanisms. Section 5.4 compares both mechanisms to price and allocate railway capacity in the context of the NEC. Section 5.5 summarizes the main conclusions of the chapter and identifies lines of future research.

5.1   Shared Railway System Performance Evaluation Framework: Inputs, Outputs, and Model Integration

The NEC (see Figure 5-1) is one of the railway corridors most widely studied in the literature (Archila, 2013; Clewlow, 2012; Kawakami, 2014; Pena-Alcaraz et al., 2013; Sussman et al., 2012; Sussman et al., 2015). However, the implications of new mechanisms to price and allocate railway capacity in this system are still unclear. This section describes the inputs and outputs needed to evaluate the performance of shared railway systems using the framework proposed in this thesis. The main inputs required

138

include both the mechanisms to price and allocate capacity, and inputs of the TO Model and the IM Model. We then discuss how the inputs and outputs of the TO Model and the IM Model relate and how we integrate both models.

Figure 5-1 Northeast Corridor (Source: NEC Infrastructure Master Plan Working Group 2010, from (Sussman et al., 2012))

5.1.1 Mechanisms Selection

As we discuss in Chapter 1, there are three main types of mechanisms to price and allocate capacity: negotiation-based, administrative-based, and market-based mechanisms. The use of market-based mechanisms for capacity pricing and allocation is

preferred in systems like the NEC characterized by capacity scarcity (congestion) and conflicting demand (Perennes, 2014; PRIIA, 2008).

According to (Gibson, 2003), the two main types of market-based mechanisms for capacity pricing and allocation for shared railway systems are 1) price-based and 2) capacity based. Price-based mechanisms are those that determine the price at which capacity will be offered, and let TOs decide whether they are willing to access the infrastructure or not. Price-based mechanisms are typically complemented with priority rules that allow the IM to decide which trains to schedule when there are conflicts (multiple TOs willing to pay the predetermined access charges to schedule conflicting services). This chapter studies a traditional price-based cost-allocation mechanism that assigns infrastructure-related costs proportionally to infrastructure use (Crozet, 2004; Nash, 2005; Lopez-Pita, 2014; Texeira and Prodan, 2014) complemented with priority rules for capacity allocation purposes. This mechanism was proposed by Amtrak and is currently being considered for implementation in the NEC (Gardner, 2013; NEC Commission, 2014).

Capacity-based mechanisms are those that determine the amount of capacity that will be offered, and let the TOs reveal the price that they are willing to pay to use that capacity, e.g. an auction (Affuso, 2003; McDaniel, 2003; Newbury, 2003; Perennes, 2014; Stern and Turvey, 2003). Auction mechanisms have been widely discussed in the literature but have not yet been implemented in any railway system. There are multiple ways to auction the access to the infrastructure (Vazquez, 2003). We could allow TOs to bid for the access to a segment of the infrastructure or for the access to different slots. Auctioning the access to segment presents two problems: 1) the value for a TO of

accessing a segment of the infrastructure is contingent on the TO's ability to get access to the rest of the train path, and 2) TOs may engage in strategies to deliberately overbook some segments to restrict access to the system to other TOs, as occurred in France in 2008 (Barrow, 2012). We also know that different types of trains in shared railway systems need different types of slots: commuter trains travel only around urban areas whereas intercity and freight trains travel between cities, and slow trains cannot use fast trains' slots (and vice-versa). Auctioning slots thus requires the IM to predetermine how many types of slots to allocate to each type of service before receiving the TOs' bids. In this chapter we thus assume that TOs bid for the desired timetable (along the whole train path) of the trains they would like to schedule over the next period of time (typically six months), for the access charges they are willing to pay, and for their flexibility to modify the desired timetable to accommodate other conflicting services. We assume that the TOs can only bid once per period (one-round auction) and they would pay the access charges they bid minus any compensation if the desired timetable is modified (first-price auction), i.e., we consider a complex one-round first-price auction. Having second bidding round would allow the TOs to use railway capacity still available after the first bidding round. This chapter studies a first round auction because in such an auction, TOs would have more incentives to reveal their willingness to pay to access the infrastructure. The literature proposes the use of second-price auctions to ensure that the auction is strategy-proof and the TOs reveal their willingness to pay. However, the bids in this auction are complex and combinatorial (as mentioned above, each TO bids for a combination of timetable, access charges, and flexibility). As a result we cannot guarantee that we have

enough information from the submitted bids to determine the second best price from a similar service for each train scheduled.

The objective of this chapter is to identify trade-offs in the choice among these two alternative capacity pricing and allocation mechanisms for shared railway systems in the context of congested railway systems, and in particular, in the NEC. This chapter focuses on how the introduction of alternative pricing and allocation mechanisms impacts the ability of intercity and commuter TOs to compete for the access to infrastructure capacity. With over 2,000 commuter trains and 150 intercity trains scheduled in the NEC per day (Gardner, 2013), the ability of commuter and intercity TOs to access the infrastructure has a direct impact on the NEC passengers.

5.1.2   Models' Inputs and Outputs

To use the framework proposed we need information about the system to be able to use the TO and the IM Models. The information required for the models can be collected from the annual TOs' financial reports and the IM's network report. As mentioned before, this is a design choice of both models. A model that allows regulators to anticipate the system reaction to a capacity pricing and allocation mechanism should not require extensive information about the railway system that only the TOs and the IM possess.

The main inputs of the TO Model are the TOs' cost and revenue structure. In terms of the costs, the TO Model aggregates all cost sources into fixed and variable costs. In terms of the revenues, the TO Model uses information about subsidies (if any) and end user's demand. According to (Gardner, 2013), there are currently one intercity TO (Amtrak), eight commuter TOs, and four active freight TOs sharing the infrastructure in

142

the NEC. This chapter focuses on the relation between intercity and commuter services during peak hours. We use data from Amtrak's financial report to model the intercity TO's profit structure. We assume that the profit structures of different commuter TOs are similar, so we use data from MBTA's financial report as a proxy to understand the commuter TOs' cost and revenue streams.

As we mentioned in Chapter 2, an intercity TO like Amtrak operating in the NEC faces fixed operational (direct) costs of $fc = \$281k$ per day and variable operational costs of $vc = \$3,425$ per train and day according to (Amtrak, 2014). In 2013, Amtrak's average fare was equal to $f_0 = \$96.5$, the number of trains was $n = 150$ trains per day in average, with a realized demand of $d_0 = 31,250$ passengers per day. The average train capacity was $c = 210$ passengers assuming a physical capacity of 250 seats with 85% load factor (Amtrak, 2014). No subsidies are required for the operations of intercity services in the NEC (Amtrak, 2010; Amtrak, 2012). We assume that the demand of intercity trains depends linearly on the fare, with an elasticity of -0.67 (Morrison, 1990).

According to (MBTA, 2013a; MBTA, 2013b) a TO like the MBTA, the commuter TO in the Boston area, faces fixed operational (direct) costs of $fc = \$435k$ per day and variable operational costs of $vc = \$1,666$ per train and per day. The elasticity of the demand with respect to the headway (frequency) is estimated by (Lago et al., 1981) to be equal to -0.41. In 2013, MBTA's average fare ranged from $f_0 = \$7 - \$25$ (average fare of $f_0 = \$13$ are considered), the number of trains averaged $n_0 = 485$ trains per day, with a realized demand of $d_0 = 130,600$ passengers per day. The train average capacity considered is $c = 350$ passengers, with 80% + load factor. Subsidies $s = \$234k$ per day are considered following (MBTA, 2013a). Commuter TOs in the

NEC are subjected to fare regulation, i.e. they cannot change the fares charged to the end users.

The main inputs of the IM Model are the information about the infrastructure and the TOs' demand for scheduling trains. To capture the main infrastructure characteristics of the NEC, we consider the system presented in Figure 5-2. It consists of a double-track corridor with 12 stations. Stations 1 and 7 represent terminal stations at both ends of the line (Boston and Washington DC respectively). Stations 2-12, 3-11, 4-10, 5-9 and 6-8 represent five stations along the corridor. We use a different station number for each traffic direction. Traffic moves in the direction of increasing station numbers with a dedicated track per direction. As a result, traffic traveling in different directions only interacts at the stations. Stations 1, 2 and 12 represent main stations in the Boston metropolitan area, stations 3, 4, 5, 9, 10 and 11 are all in the New York metropolitan area, and stations 6, 7 and 8 are in the Washington DC metropolitan area. Five types of services can be considered: Boston commuter trains traveling around the Boston metropolitan area (stations 1, 2, and 12); New York commuter trains; Washington DC commuter trains; and intercity and freight trains traveling between Boston and Washington DC. Intercity and freight trains may not stop at every station. Intercity trains travel at higher speeds than commuter trains. The distance between Boston and Washington DC is approximately 450 miles, and the distance traveled by commuter TOs around each city ranges from 40 to 70 miles per direction. Note again that although the infrastructure considered is simple and does not include many intermediate stations such as Philadelphia, New Haven, etc., it contains all the important elements to capture the

144

dynamics of the interaction between commuter and intercity traffic under both capacity pricing and allocation mechanisms.
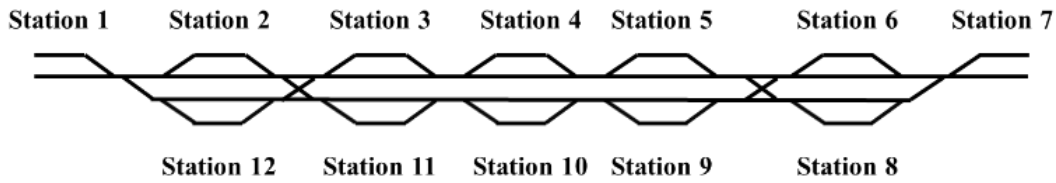
Figure 5-2 Detailed corridor infrastructure (Source: author, Figure 3-2)

The TO Model has two main outputs: the demand for scheduling trains (number of trains that each TO would like to schedule and the access charges that each TO is willing to pay to schedule them) and the fares the TOs would charge to the end users. The IM Model has two main outputs: the final access charges that each TO would pay and the final train timetable (set of trains scheduled and their timetable).

5.1.3   Integration of the Train Operator and the Infrastructure Manager Model

In Section 5.1.2 we mention that the demand for scheduling trains is both an output of the TO Model and an input of the IM Model. There are important differences however between them. Specifically, the demand for scheduling trains considered as an input of the IM Model includes five pieces of information: the number of trains that the TOs ask to schedule, the desired timetable of each train to schedule, the access charges that the TO would pay to schedule each of the trains, information about whether they are flexible to modify the timetable of the train (and by how much) in case of conflict with other train, and how much the IM should compensate them to modify the desired timetable. However the demand for scheduling trains that we get as an output of the TO Model only includes information about the maximum number of trains that each TO would like to schedule and the maximum access charges that each TO is willing to pay to schedule each of the trains.

There are three important observations. First, the outputs of the TO Model are independent of the IM decisions. The TO Model characterizes the maximum number of trains and access charges that the TO would accept. The IM Model would never propose a solution in which the TOs schedule more trains or to pay more than that. In that sense, all the solutions of the IM Model are feasible from the TO perspective.

Second, TOs have incentives to disclose their demand for scheduling trains and their maximum willingness to pay for access to the IM when they do not have market power. In other words, we can use the outputs of the TO Model as inputs of the IM Model. A very interesting line of future research should consider how TOs behave when they have market power and how this behavior will affect the performance of the system following the discussion of Section 5.1.1.

Third, there are three additional IM Model inputs that we cannot get directly from the TO Model: the desired timetable of the trains, how much TOs are willing to change that desired timetable in case of conflict, and how much compensation they should receive for their flexibility to do that. In this chapter we assume that passenger TOs in the NEC are willing to reschedule each train a maximum of 15 minutes and they require a reduction in the access charges of that train equal to 3% per minute modified. These numbers are based on the standards to define passenger train timetables "Introducing a timetable that is easy to remember on the most important lines […] long-distance trains that stops at the main stations only should arrive every 15 minutes" (Kroon et al., 2009), and on the definition of punctuality in in Europe and the US (FRA, 2009; Renfe, 2015). According to (FRA, 2009), an Acela (HSR) train is considered *on-time* if it arrives its endpoint terminal within 10 minutes of the scheduled arrival time. On the other hand, a

Northeast Regional (intercity) train is considered *on-time* if it arrives within 10 minutes for trips less than 250 miles, 15 minutes for trips between 251 and 350 miles, and 20 minutes for trips between 351 and 450 miles. According to (Renfe, 2015), Renfe reimburses 50% of the ticket to HSR passengers that experience a delay of more than 15 minutes.

In terms of the desired timetable, once we know the departure time from the first station, we could use public information from the current timetable to determine the arrival and departure time to all the other stations. In Section 5.3.1 we show that in the NEC if 1) the frequency of commuter trains is higher than or equal to two trains per hour, and 2) TOs prefer to schedule trains uniformly, then the exact desired timetable does not affect the number of trains scheduled nor the access charges that each TO has to pay. This result is based on the fact that a perturbation on the desired timetable of the intercity trains does not affect how many commuter trains would conflict with the intercity train (see Figure 5-4) when the schedule is dense enough (i.e., when the frequency of the commuter is at least 2 commuters per hour). This is a very important result of this research. It suggests that the exact TO's desired timetable affects the final timetable, but it does not affect the number of trains scheduled neither the access charges in many congested shared railway systems. In other words, the exact TO's desired timetable will not be necessary to understand the implications of capacity pricing and allocation in many congested shared railway systems.

5.2   Evaluation of a Price-Based Mechanism: Cost-Allocation and Priority-Rules

In this section we analyze the implications of the price-based cost-allocation and priority-rule mechanism defined in Section 5.1.1 for a shared railway system like the

NEC. We first use the TO Model to anticipate the number of trains that a TO would like to operate for different values of variable access charges and determine the resulting TO profits and IM revenues. We then use the IM Model to analyze if the TOs' demand can be scheduled on the infrastructure.

5.2.1 Train Operator Model Results

Figure 5-3 shows the demand for scheduling train services of an intercity TO in the NEC with Amtrak's cost structure and its resulting profits for different variable access charges. Figure 5-4 shows the demand for scheduling train services of a commuter TO in the NEC with MBTA's cost structure and its resulting profits for different variable access charges. For clarity purposes, we do not show the fares in Figures 5-3 and 5-4 although we use them to compute the profits of the TOs. For the intercity TO, the fares increase from $128 to $237 as access charges increase. For the commuter TO, the fares do not change because they are regulated. Figures 5-3 and 5-4 use a distance of 50 miles for commuter trains (Boston, MA to Cranston, RI) and a distance of 450 miles for intercity trains (Boston, MA to Washington, DC) to compute the variable access charges per mile (Gardner, 2013). We first see that the number of trains that the TO would like to schedule decreases as variable access charges increase.

Figures 5-3 and 5-4 show the maximum variable access charges that both intercity and commuter TOs are able to pay. Table 5-1 summarizes this information. In particular, the maximum access-charge that an intercity TO like Amtrak would be able to pay is $102 per train-mile per day, which is equivalent to $46,000 per train per day for Boston to Washington. The maximum variable access charge that a commuter TO like MBTA would be able to pay is $52 per train-mile per day, which is equivalent to $2,578 per train

148

per day. With higher variable access charges the TOs would be better-off not operating any trains. That means that an intercity TO is able to pay two (2) times as much as a commuter TO in a per mile basis, or almost eighteen (18) times as much as a commuter TO in a per train basis. We also determine the maximum sustainable access charges (access charges for which the TOs would have 0 profits after reimbursing capital at an adequate rate of return). That means that their finances allow them to continue operations over the medium term without additional operational subsidies.
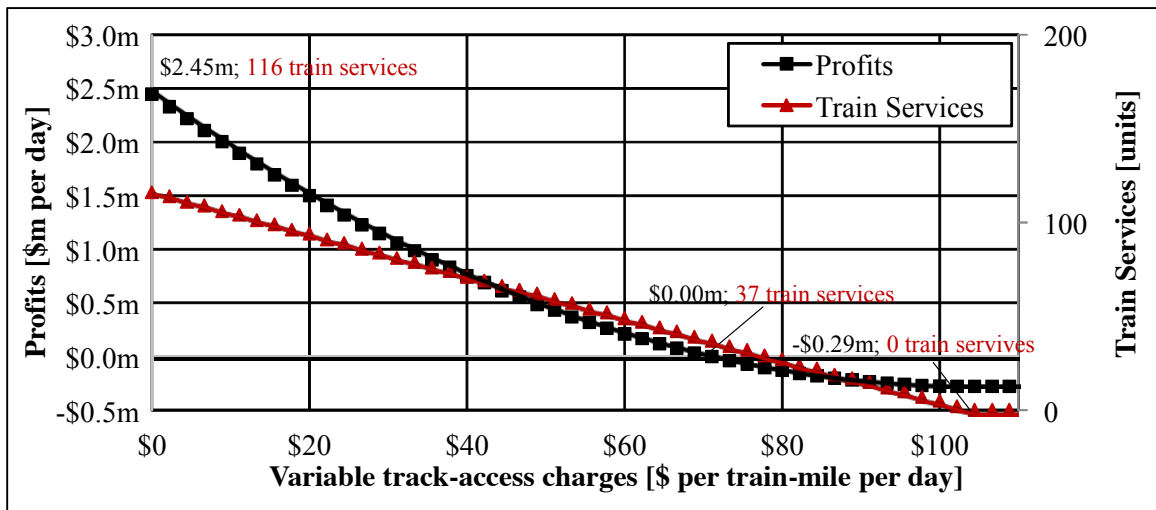
Figure 5-3 NEC intercity TOs expected profits per day and number of trains per day for different variable access charges (Source: author)
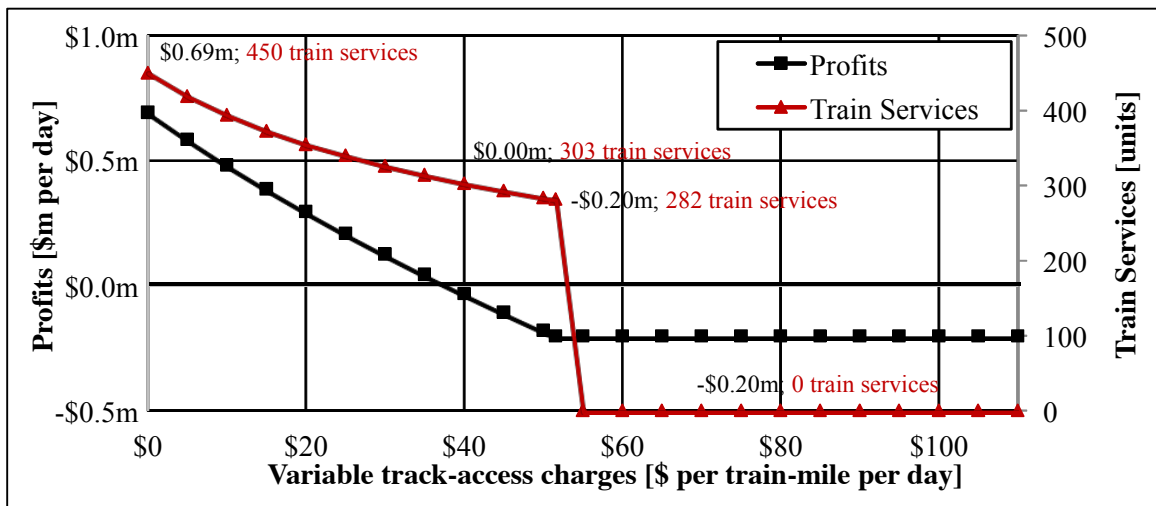
Figure 5-4 NEC commuter TOs expected profits and number of trains per day for different variable access charges (Source: author)

Table 5-1 TOs' expected profits and number of trains to schedule per day for different variable access charges (Source: author)

	reference point	no access charges			maximum access charges			maximum sustainable access charges		
	profits (n=0) [$m]	ac_v [$ tr-mi]	N [trains]	profits [$m]	ac_v [$ tr-mi]	N [trains]	profits [$m]	ac_v [$ tr-mi]	n [trains]	profits [$m]
Intercity	-0.28	0.00	116	2.47	102	1	-0.28	71	37	0.00
Commuter	-0.20	0.00	450	0.69	52	282	-0.20	40	303	0.00

Although both figures only show the response of TOs to variable access charges (per train), we can use them to determine the response of the TOs to fixed access charges too. As discussed in Section 2.2 and equation (2.5), a fixed access charge in addition to the variable access charge would not change the TO operational decisions (number of trains and fares) as long as the resulting TO profits are greater than $s(0) - fc$, (the profits the TO would obtain operating 0 trains). If the resulting profits were lower than the profits operating no trains, a profit maximizing TO would be better off operating no trains (and paying no access charges). Otherwise, the TO would see the fixed access charge as a fixed lump sum that would not change the optimality conditions in the profit equation and as a result, would not change its operational decisions either.

The maximum fixed access charges that the intercity TO would be willing to pay would be $2.75 million per day ($2.47 million + $0.28 million). With this access charge, it would still operate 116 trains, and its profits would be –$0.28 million. If the access charge increases beyond that point, the TO would not operate any trains. Likewise, the maximum fixed access charges that the Boston and Washington DC commuter TOs would be willing to pay are $0.99 million per day ($0.69 million + $0.43 million – $0.23 million). The New York City commuter TO would be willing to pay up to $1.98 million per day (because the commuter trains around New York City travel double the distance than the commuter trains in Boston and Washington DC in the NEC, part in the New

York City – Boston and part in the New York City –Washington DC segment). According to (Gardner, 2013), the NEC should invest $51.9 billion (uninflated) from 2010 to 2030 to bring the system to a state of good repair. That means that the IM would need to invest $7.10 million per day. That means that the maximum possible recovery with any mechanisms is $6.71 million per day ($2.75 million + 2×$0.99 million + 1.98 million), i.e., the maximum possible recovery considering intercity traffic and commuter traffic around Boston, New York City, and Washington DC is 95%.

The main advantages of including fixed access charges are that they allow for maximum IM revenue collection while they do not affect the TO's operational decisions (Pena-Alcaraz et al., 2014). However, the use of fixed access charges may create barriers of entry for new competitors. Determining the right fixed access charges for different TOs is also challenging. The rest of this chapter thus assumes that there are no fixed access charges. Note again thought that as shown in Chapters 2 and 4, the use of variable access charges higher than the variable infrastructure maintenance costs would result in operational levels that are suboptimal from a social welfare standpoint.

Table 5-2 shows these same results from the perspective of the IM. According to (Gardner, 2013), a cost-allocation model would allow the IM to charge TOs for the use of the infrastructure. The access charges would depend on the segment in which the trains are scheduled and on their infrastructure needs. The resulting access charges for intercity and commuter TOs would be comparable (as opposed to those of freight trains that do not use passenger stations). According to (Texeira and Prodan, 2014) variable access charges in other countries vary from $0 per train mile (e.g. Estonia and Norway) to $50 - $100 per train mile (in France and Netherlands).

Table 5-2 IM expected revenues for different variable access charges and resulting TOs' demand for scheduling trains per day assuming three commuter TOs (Source: author)

ac_v [$ train-mi-day]	revenues – IM [$m]	n – intercity TO [trains]	n – commuter TO [trains]
currently	0.76	153	458 (x3)
0	0.00	116	450 (x3)
25	2.69	88	340 (x3)
50	4.19	60	284 (x3)
52	4.28	59	282 (x3)
75	1.08	32	0
100	0.23	5	0
125	0.00	0	0

We can see that the revenues that the IM obtains do not always increase when access charges increase, since the TOs' demand for scheduling trains drop. These results suggest that the maximum revenues that the IM would be able to collect from these intercity and commuter TOs is $4.28 million per day when it charges $52 per train-mile. At this point, the intercity TO would be contributing $1.37 million, the Boston and Washington DC commuter TOs would be contributing $0.73 million each, and the New York commuter TO would be contributing $1.45 million (because New York commuter trains travel double the distance than Boston and Washington DC ones as mentioned above). Note that with the current pricing and allocation contracts, the IM revenues are equal to $0.76 million per day, so the IM raises revenues equal to only 10% of the basic infrastructure costs required to bring the system to a state of good repair. The results also show that this price-based mechanism would allow the IM to recover up to 60% of infrastructure costs ($4.28 million / $7.10 million per day).

5.2.2 Infrastructure Manager Model Results

The TO Model anticipates the response of the TOs when each of them optimizes their operational decisions on their own. As a result, the revenues and profit presented in Figures 5-3 and 5-4 and Tables 5-1 and 5-2 assume that the TOs are able to schedule all

the trains on the infrastructure. This may not be the case in a congested corridor like the NEC. In this case, however, Table 5-2 shows that TOs' demand for scheduling trains under this mechanism is always lower than the current level of operations (presented in the first row of Table 5-2). In other words, we know that there is enough capacity in the corridor to schedule all the trains that TOs would like to schedule under this price-based mechanism. Consequently, when we use the IM Model in this case, we see that the IM should be able to schedule all the trains in the system if TOs are willing to adjust their desired schedule to accommodate conflicting trains. Note however that TOs with priority to access the infrastructure may not have incentives to be flexible on their trains scheduling preferences. In other words, when there are conflicts since they have priority to access the infrastructure, priority TOs know that their trains will be scheduled independently of their flexibility, whereas the trains of other TOs may only be scheduled if all TOs are flexible. This will thus have a direct impact on the other TOs ability to schedule trains in the system, their profits, and on the total revenues collected by the IM. The IM Model can thus be used to understand the implications of different priority rules for the system. This is an important consideration for this mechanism, because the priority rules grant some TOs priority over others.

5.3   Evaluation of a Capacity-Based Mechanism: Auction

In this section we analyze the implications of the auction mechanism defined in Section 5.1.1 for a shared railway system like the NEC. In this case we use first the IM Model to determine the minimum access charges that an intercity TO have to pay (as a function of the commuter TO access charges) to be able to schedule an intercity train as a function of the commuter frequency. We then use that information and the results from

153

the TO Model to estimate the number of trains that a TO would operate and the access charges it would pay. We use the results of both models to estimate the final TO profits and IM revenues.

5.3.1   Infrastructure Manager Model Results

We start analyzing the optimal capacity allocation plan (train timetable) to determine how to coordinate different TOs' conflicting demands for scheduling trains. Figure 5-5 shows the time-space diagram for the optimal timetable designed using the IM Model in a case in which an intercity TO tries to schedule one train and three commuter TOs try to schedule commuter trains around Boston, New York, and Washington DC every 30 minutes. The y-axes represent distance in miles from station 1 and the x-axes represent time in minutes at which different trains are scheduled to pass through each point of the line (vs. the desired timetable in dashed line). There are no interactions between trains traveling in different directions. The IM Model proposes the final timetable analyzing the trade-off between eliminating trains and readjusting the desired timetable, according to the objective function in equation (3.1). We can use this information to determine how much intercity TOs will have to pay to be able to schedule services that conflict with commuter services. For this example, we need to solve a train timetabling problem with 115 commuter trains and 1 intercity train. For clarity purposes, only the schedules of conflicting trains are shown in Figure 5-5.

As Figure 5-5 shows, the intercity train would initially conflict with 14 commuter trains. Rescheduling the commuter trains to accommodate the intercity service requires that the commuter TOs receive a total discount equivalent to the access charges of 2.1 commuter trains. As a result, the IM would only schedule the intercity train if the intercity

TO pays access charges higher than 2.1 commuter trains, i.e., if its bid is higher than 0.33 times that of the commuter TOs per train-mile (considering the miles traveled by intercity and commuter trains). This number does not change with the desired timetable of the intercity train because the intercity train still conflict with 14 commuter trains. The same results are obtained when the frequency of commuter trains is higher than 2 trains per hour. For these frequencies, the number of conflicting trains does not depend on the intercity train desired timetable, nor the total discount that the commuter TOs should receive because the timetable is dense and uniform. That means that we do not need to know the exact desired timetable to determine the relationship between intercity and commuter train bids in the NEC if the frequency of commuters is greater than 2 trains per hour.
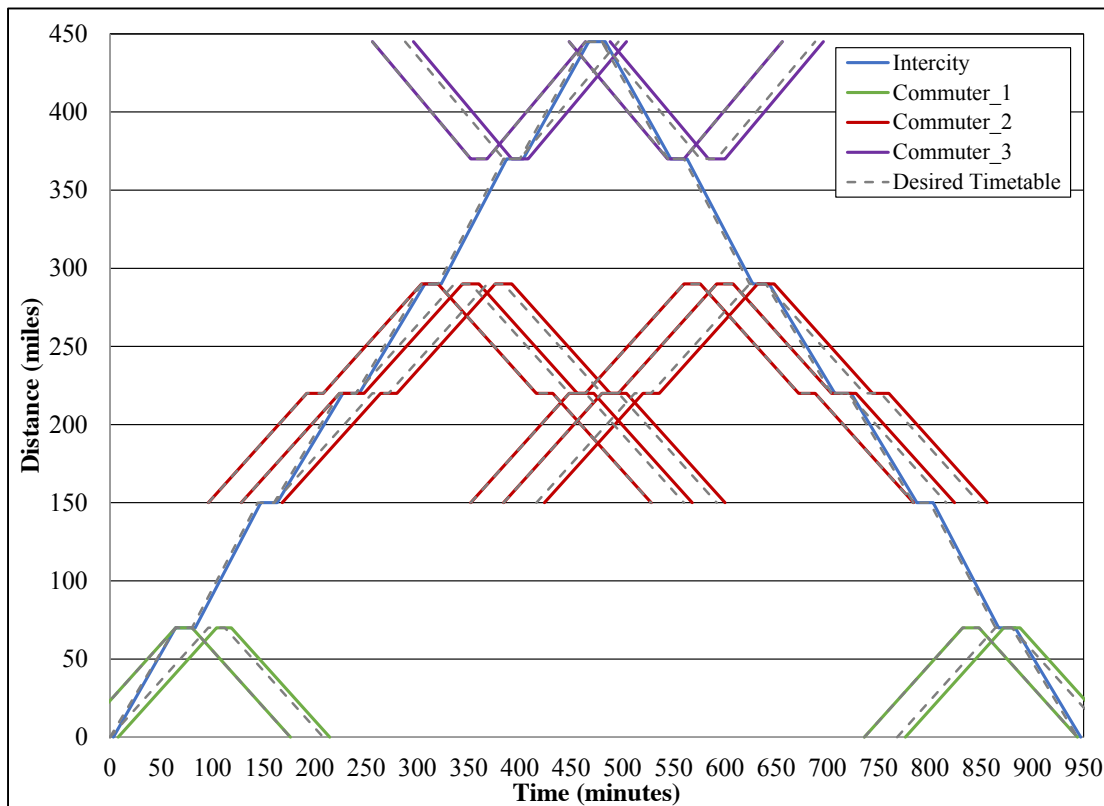
Figure 5-5 Timetable proposed by IM to schedule an intercity train in a system with commuter trains operating every 30 minutes (Source: author, Figure 3-5)

155

If the frequency of commuter trains increases, for example to one commuter train every 15 minutes instead of every 30 minutes, the intercity train will initially conflict with 22 commuter trains and will only be scheduled if the intercity TO pays access charges equivalent to the access charges of 5.3 commuter trains (i.e., if the intercity TO bid is higher than 0.82 times that of the commuter TOs' per train-mile). Conversely, if the frequency of commuter trains decreases to one train every 60 minutes, the intercity train will be scheduled almost always (the IM would need to compensate the commuter TOs with a total compensation that ranges between 0 and 0.4 times the access charges paid by a commuter trains depending on the desired timetable, what translates in bids higher than 0.00 or 0.06 times that of the commuter TOs per train mile). The model can be used to quantify the trade-off between commuter and intercity trains for any other frequency of service (see Table 5-3).

Table 5-3 Minimum intercity to commuter access-charge per train-mile bid ratio to ensure that their train is scheduled as a function of the commuter frequency (Source: author)

commuter frequency [minute]	ratio [per train-mi]	commuter trains scheduled [%]
5	5.86	73%
10	0.84	100%
15	0.82	100%
30	0.33	100%
60	0.00-0.06	100%

Table 5-3 shows that, when the system is congested, an intercity TO may have to pay more than a commuter TO to schedule a train: the intercity TO has to pay between 0.82 and 5.86 times the access charges of commuter TOs per train-mile or between 5.27 and 37.67 times the commuter TOs' access charges per train. This minimum intercity access charge reflects the congestion rents. The results show that greater cost recovery is expected in congested infrastructure. The frequency of commuter services in the NEC

today ranges between 10 and 30 minutes. A first analysis of the results suggests that, with current levels of traffic, intercity TOs have an advantage over commuter TOs to access the infrastructure under this capacity-based mechanism.

5.3.2   Train Operator Model Results

How can we know whether commuter TOs would be able to compete to access the infrastructure with the intercity TO? Fortunately, we can use the TO model (equation (2.4)) to determine and compare the maximum access charges that each TO would be willing to pay as a function of the number of trains they want to schedule. Figure 5-6 summarizes these results, the maximum variable access charges that an intercity and a commuter TO with the cost structure of Amtrak and MBTA respectively could bid as a function of the number of trains to schedule. Similarly to the cost-based approach, the results show that the intercity TO ability to pay to access the infrastructure is almost double that of the commuter TO counterparts. Note that the TOs' willingness to pay to access the infrastructure decreases with the number of train services with the exception of the commuter TOs' willingness to pay when they schedule between 0 and 280 train services. This happens because end users' demand for commuter services is elastic to the frequency and increases substantially when more trains are scheduled (or decreases substantially when the frequency of commuter trains is very low). As a result, commuter TOs have incentives to ensure a minimum service frequency.

We need to make one adjustment before we use these results as inputs of the IM Model. The TO Model assumes that all the trains have the same OD pair. However, the 150 intercity services that (Gardner, 2013) mentions, include for instance Boston to New York services and New York to Washington DC services that we count in the IM Model

157

as a single service. They also average the number of trains during the day without considering differences between peak and off-peak hours. We use the following equivalences between frequency and number of trains: 118 intercity services in the TO Model are equivalent to 1 train per hour in the IM Model and 450 commuter services in the TO Model are equivalent to 6 commuter trains per hour in the IM Model (Amtrak, 2014; Amtrak, 2015; MBTA, 2013a; MBTA, 2015).
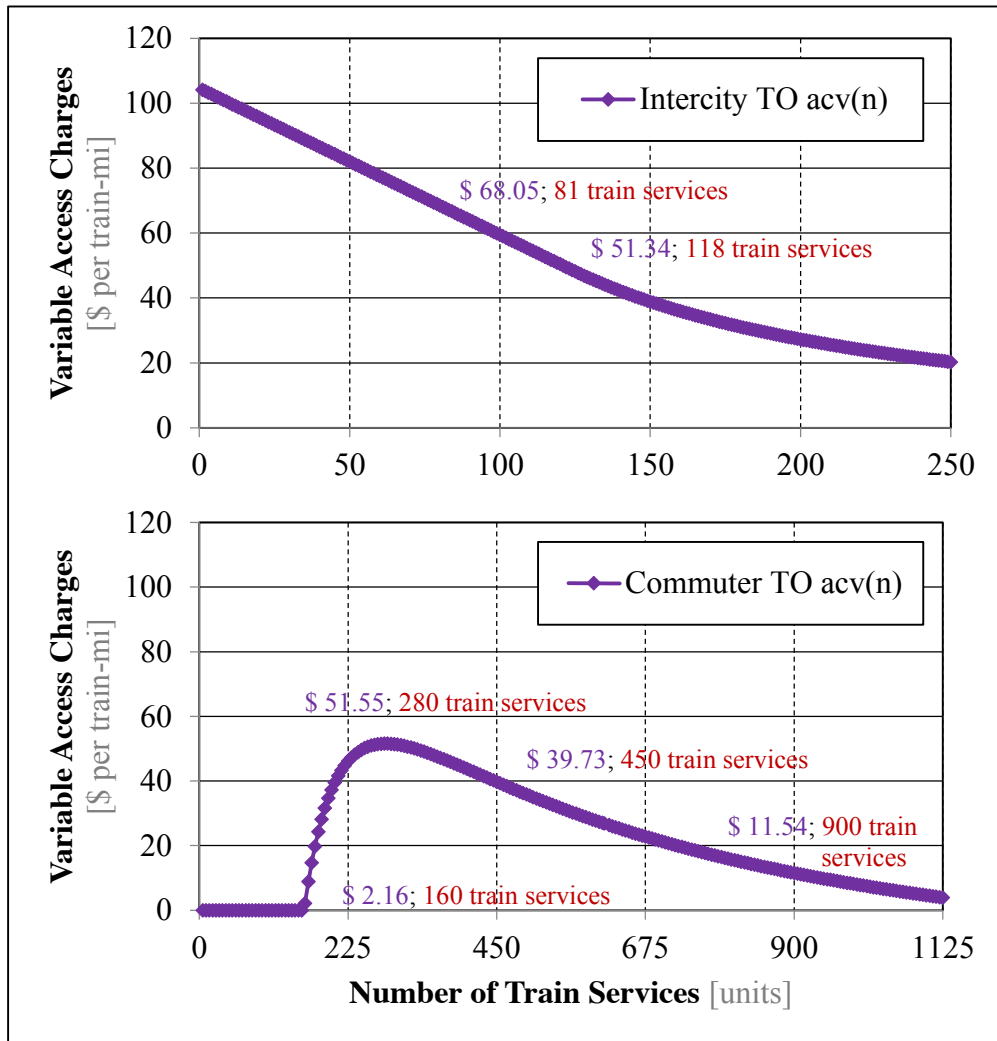
Figure 5-6 NEC intercity and commuter TOs maximum willingness to pay for access (maximum variable access charges) as a function of the number of trains to schedule (Source: author)

Table 5-4 shows the result of combining the information in Table 5-3 and Figure 5-6. The first three columns show the bids of each commuter TO as a function of the

desired frequency (number of trains to schedule). The next three columns show the bid of an intercity TO that tries to schedule 1 train per hour. The last two columns determine how many trains of each type can be scheduled and compute the resulting revenues for the IM (again, assuming three commuter TOs, one in the Boston area, one in the New York City area, and one in the Washington DC area).

Table 5-4 TOs' demand for scheduling trains for different variable access charges and resulting IM expected revenues per day assuming three commuter TOs (Source: author)

commuter frequency [minute]	n commuter TO [trains]	ac_v commuter TO [$ train-mi]	n intercity TO [trains]	ac_v intercity TO [$ train-mi]	n commuter n intercity [trains]	revenues IM [$m]
5	900	12	81	68	657, 81	4.00
10	450	40	118	51	450, 118	6.30
15	300	51	118	51	300, 118	5.80
16	280	52	118	51	280, 118	5.61
28	160	2	118	51	160, 118	2.80
30	150	0	118	51	0, 118	2.73

Considering that the NEC infrastructure needs in the NEC amount to $ 7.10 million per day from 2010 to 2030, the results for this mechanism show that the IM would be able to recover up to 89% ($6.30 million / $7.10 million) of infrastructure costs with intercity and commuter train services around Boston, New York City and Washington DC. This is a substantial recovery ration considering 1) that currently the IM only recovers 10% of infrastructure costs and 2) that the maximum potential for recovery with these services amounts to 95% as mentioned in Section 5.2.1.

Note that the intercity to commuter TO bid ratio exceed the ratio presented in Table 5-3 when the commuters frequency (headway) is bigger than 5 minutes. In other words, the intercity TO is almost always able to schedule all the intercity services. These results confirm that the intercity TO in the NEC is usually in better position to access the tracks than the commuter TO under an auction mechanism with current levels of service. If the frequency of commuter trains were to increase by 85%, with 5 minutes headways,

the intercity TO would not be able to schedule trains if it bid less than $68 per train-mile (5.86 x $12 per train-mile). Using the TO Model, we can determine that the intercity TO is able to bid over $68 per train-mile if its demand for scheduling trains is equal to 81 trains or less.

In this case, the commuter TO would only be able to schedule 657 trains (73% of 900). This equilibrium is stable because none of the TOs would want to schedule more trains. As Figure 5-6 shows, the commuter TO would be willing to pay higher access charges for 657 trains than for 900 trains. As a result, scheduling only 657 trains at the 900-train access charge level translates to extra profit for the TO. Similar results are obtained for the intercity TO bids and for all other commuter TO bids with more than 280 trains. Between 160 and 280 commuter trains the equilibrium is not stable because the demand for commuter services would significantly decrease due to the amount of service reduction, and also the commuter TO profits when not all trains are scheduled.

5.4 Comparison of Price-Based and Capacity-Based Mechanisms in the Northeast Corridor

The previous sections discuss the operational decisions of intercity and commuter TOs under a price-based (cost-allocation and priority-rule) mechanism and a capacity-based (auction) mechanism in the NEC. The results are summarized in Tables 5-2 and 5-4 respectively.

Table 5-5 shows the number of trains that each TO scheduled and its profits for different access charges, together with the revenues raised by the IM under both mechanisms. Although there is not a one-to-one comparison between both mechanisms, we can compare both sides of the table.

Table 5-5 Distribution of revenues (rev., in $m), profits ($\pi$, in $m) and number of trains scheduled (n) per day comparison for different access charges (ac$_v$, in $ per train-mile) (Source: author)

Price-Based Mechanism							Capacity-Based Mechanism						
intercity TO			commuter TO			IM	intercity TO			commuter TO			IM
ac$_v$	n	π	ac$_v$	n	π	rev.	ac$_v$	n	π	ac$_v$	n	π	rev.
0.0	116	2.5	0.0	450	0.7	0.0	68.1	81	-0.3	11.6	657	0.2	4.0
25.0	88	1.3	25.0	340	0.5	2.7	51.3	118	-0.3	25.0	640	-0.2	5.9
50.0	60	0.5	50.0	284	-0.2	4.2	51.3	118	-0.3	50.0	325	-0.2	6.0
51.3	59	0.4	51.3	282	-0.2	4.3	51.3	118	-0.3	51.3	297	-0.2	5.8
51.6	59	0.4	51.6	282	-0.2	4.3	51.3	118	-0.3	51.6	280	-0.2	5.6
53.3	57	0.4	53.3	0	-0.2	1.4	51.3	118	-0.3	51.6	280	-0.2	5.6

These results show that the revenues collected by the IM under the capacity-based mechanism studied are higher than the revenues collected when a price-based mechanism with similar charges for intercity and commuter TOs is implemented. The capacity-based mechanism does not only allow the IM to collect higher revenues (around 20% more than using the price-based mechanism), but also results in higher number of trains scheduled (20% more) and higher total welfare (also 20% higher) as compared to the price-based mechanism. Note that these differences are significant considering the robustness of the TO Model (discussed in Chapter 2) and the precision of the IM Model (discussed in Chapter 3). These advantages, however, have a cost for the TOs, who receive much lower profit (losses in most cases). Consequently, the results of the auction mechanism are not resilient to uncertainty in transportation demand, since the TOs would prefer not to operate trains if the profits decrease even further. This auction mechanism may thus require the design of a procedure to redistribute revenues to ensure that TOs can sustainably operate over the medium term.

5.4.1   Response to Uncertainty

It is important to note that both the TO Model and the IM Model considered in this research are deterministic. They intend to capture the essence of a normal day of operations. However, the TOs and the IM face several sources of uncertainty when

making their operational decisions. The two most important sources of uncertainty in the NEC in the medium term are 1) the condition of the infrastructure, 2) the end users' demand for transportation services.

The first source of uncertainty is particularly critical until the NEC reaches a state of good repair that ensures that the infrastructure is reliable. With today's backlog of maintenance work, the need of last-minute maintenance and interventions has a direct impact on the capacity on the corridor and on the TOs ability to schedule trains. This uncertainty gets amplified under price-based mechanisms. Any problems with infrastructure availability under price-based mechanisms would reduce the number of services operated by TOs that often operate the minimum number of services that allows them to be profitable. This lack of infrastructure capacity unevenly affects those TOs with lower priority assigned. Under capacity-based mechanisms, the TOs will still make profits even if some trains are not scheduled due to infrastructure availability problems. The IM under the auction mechanism studied would have important incentives to avoid uncertainty on the infrastructure capacity availability, since fewer trains scheduled would lower its ability to recover infrastructure costs as compared to the deterministic case.

The uncertainty in the demand for transportation has a major impact on the expected revenues that the TOs would collect. The fact that the price-based mechanism ensures high TO profits makes this mechanism more resilient to demand uncertainty than the auction mechanism, where the TOs operational decision will probably change if the TOs expect a very uncertain demand. Note that there is also uncertainty in the demand distribution. Passengers do not arrive homogeneously during the day; they instead concentrate around some particular times. As a result we may expect TOs scheduling

some more trains than the ones that the model indicates. For example, while the commuter TO Model shows that the optimal number of trains to schedule is 397, MBTA currently runs 485 trains on the infrastructure. Although the model already considers a load factor of 80% to accommodate part of this demand, scheduling 485 commuter trains would result in an average load factor of 65% (industry benchmark for commuter services). This would result in a higher operational cost to accommodate the same demand. The need to offer 485 vs. 397 trains would depend on the exact distribution of the demand. The load factor of the model can be adjusted to consider this uncertainty. This uncertainty will propagate to the expected profits. As a result, we may expect to see a lower TOs' willingness to pay to access the infrastructure even for the same number of trains to schedule, and hence a lower IM ability to recover infrastructure costs than in the deterministic case.

5.4.2   *Mechanism Implementation*

In terms of implementation, price-based mechanisms are easier to implement than capacity-based mechanisms. As mentioned above the priority rules have important implications for NEC commuter and intercity passengers and TOs, especially if the IM does not have access to sophisticated methods to solve the train timetabling problem. Nonetheless, determining which trains to schedule under a price-based mechanism, once the priority rules and the demand for scheduling trains are known, is easy to understand for all stakeholders.

However, the implementation of an auction mechanism *requires* the IM to be able to solve the train timetable problem proposed in Chapter 3 to ensure transparency in the capacity allocation process. As we discussed in Chapter 3, solving the train timetabling

problem is difficult in railway systems with large numbers of trains and stations. Note also that understanding the implications for infrastructure utilization of capacity rules also *requires* the IM to solve the optimal train timetable given the operators demand for scheduling trains. In other words, being able to solve the train timetabling problem is critical to evaluate both price-base and capacity-based mechanisms.

5.4.3 Other Considerations: Gaming the Mechanisms

As we mentioned before, these results also assume that TOs do not have market power or do not take advantage of their power to game the mechanism to their interest and disclose their willingness to pay to access the infrastructure. However, TOs often have market power in the railway industry. The results of this chapter show, for instance that commuter TOs' ability to pay for access is much lower than the one of their intercity counterpart. This is particularly important under auction mechanisms where the TOs have incentives to keep lowering their bids while their trains get scheduled to maximize their profits. In particular, intercity TOs could use the framework proposed in this thesis using only publicly available data to replicate the results of Table 5-3 to understand that they could get their trains scheduled paying much less than their maximum willingness to pay. These aspects are beyond the research of this thesis. Note in any case, that the framework proposed in this chapter would allow the regulator to anticipate the results of the auction and to investigate any variation with respect to these numbers.

5.5 Conclusions

This chapter shows how to use the framework developed in Chapters 2 and 3 of this thesis to evaluate the performance of a congested shared railway system based on the NEC under two alternative capacity pricing and allocation mechanisms considering both

planning and operational aspects. The two alternative capacity pricing and allocation mechanisms evaluated are a cost-allocation and priority-rule mechanism, which was proposed by Amtrak and is currently being considered by the NEC Commission and the different TOs in the NEC, and a capacity based (auction) mechanism. Section 5.1.3 discusses how to integrate both the TO Model and the IM Model to be able to anticipate the demand for scheduling trains, set the access charges (capacity pricing), and set the final train timetable (capacity allocation: set of trains scheduled and their timetable).

The results of the chapter show that there are important trade-offs between the mechanisms analyzed. The capacity-based mechanism studied results in almost 20% more IM cost recovery and 20% more trains scheduled as compared to the price-based mechanism studied in the NEC. However, it also results in lower profits for the TOs. The price-based mechanism, on the other hand, is easy to implement (does not requires the IM's ability to solve a sophisticated the train timetabling problem) and ensures higher profits for the TOs, making the TOs more resilient to uncertainty in end users' transportation demand. Note again that this comes with a cost to end users (who will have fewer trains) and to the IM (who will obtain fewer revenues from access charges). The price-based mechanism is not very resilient to uncertainty in infrastructure capacity availability. Under the capacity-based mechanism, intercity TOs are in better position than commuter TOs to access the tracks with current NEC levels of service. The priority level of each TO is a design choice in price-based mechanisms, but this choice has important implications for NEC commuter and intercity passengers and TOs that need to be benchmarked against the optimal capacity allocation determined with the IM Model.

This analysis also benchmarks the IM cost recovery of the price-based mechanism and the capacity-based mechanism studied, 60% and 89% respectively, with the maximum IM cost recovery achievable with any capacity pricing and allocation mechanism and the current IM cost recover, 95% and 10% respective. This shows that 1) both mechanisms allow for much greater IM cost recovery than the current capacity pricing and allocation mechanism implemented, and 2) the capacity-based mechanism designed IM cost recovery level (89%) is very close to the maximum IM cost recovery achievable in the NEC (95%) considering intercity services and commuter services around Boston, New York City and Washington DC. Slightly higher IM cost recovery numbers could be achieved considering other commuter services in the NEC (Philadelphia, Connecticut) and freight services too.

To our knowledge, this is the first research that compares the performance of price-based and capacity-based mechanisms in the same railway system. The results show that neither of these two mechanisms is superior to the other on all dimensions. A better understanding of these trade-offs is necessary to design effectively coordinate shared railway systems. We believe that the stakeholders in the NEC should carefully analyze the implications of alternative pricing and allocation mechanisms before locking the system into one of them.

Although this chapter focuses on the interactions between intercity and commuter TOs, the framework proposed is valid to analyze other questions such as the implications of the mechanisms for freight TOs, for the end users, or for the whole system. Future research should consider the variety of services operated in the NEC (services with different speeds and stops, serving different OD pairs), the distribution of the revenues

collected by the IM among the different NEC infrastructure owners, and the effects of TO's market power to refine the understanding of the trade-offs among alternative pricing and allocation mechanisms. Future research should also analyze how these results change in the context of other congested and non-congested shared railway systems.

Chapter 6 - Conclusions

"If we knew what we were doing, it would not be called research." – attributed to A. Einstein

In this thesis we developed a framework to analyze the performance of shared railway systems under alternative capacity pricing and allocation mechanisms. We then use this framework to understand the implications of representative capacity pricing and allocation mechanisms in the Central Corridor in Tanzania and the Northeast Corridor in the US. In this chapter Section 6.1 summarizes the work presented in this thesis and the main findings of our research, Section 6.2 presents the main conclusions and recommendations, and Section 6.3 describes possible directions for extensions and further research.

6.1 Summary of Thesis

Recently, governments have started promoting the use of shared railway systems. Shared use allows for a more efficient utilization of existing railway infrastructure but requires a strong coordination between the infrastructure manager (IM) and the train operators (TOs). Such coordination, in turn, requires capacity planning mechanisms that determine which trains can access the infrastructure at each time, capacity allocation, and the access price they need to pay, known as capacity pricing. This is particularly challenging in the railway industry because there are strong interactions between capacity planning and infrastructure operations.

We start in Chapter 1 by presenting a literature review of alternative capacity pricing and allocation mechanisms. We draw two important conclusions in relation to the

existing research in shared railway systems at a macroscopic level. First, we conclude that capacity pricing and allocation mechanisms used for coordination purposes in shared railway corridors are getting more heterogeneous. Second, we observe that different mechanisms are evaluated using different metrics. As a result, the comparative performance of different mechanisms to price and allocate railway capacity is still unclear. This thesis aims to help fill this literature gap by 1) developing a framework to evaluate the performance of shared railway systems under alternative capacity pricing and allocation mechanisms (Chapters 2 and 3), and 2) using this framework to understand the implications of representative capacity pricing and allocation mechanisms in the Central Corridor in Tanzania (Chapter 4) and the Northeast Corridor (NEC) in the US (Chapter 5). This thesis does not answer the question of which capacity pricing and allocation is best, because there is no unambiguous answer. The selection of the most appropriate mechanism to price and allocate railway capacity depends on the shared railway systems goals.

Chapter 2 presents a TO Model that anticipates TOs' demands for infrastructure use under alternative capacity pricing and allocation mechanisms. The focus of this thesis is in the interactions between railway infrastructure operations and available infrastructure capacity. As a result, the TO Model proposed is simple by design. The main objective is to allow the regulator and the IM to anticipate the response of the TOs relying only on information that is readily available to them. Nonetheless, the structure of the model is consistent with the standard financial models of TOs used to analyze investments in the railway industry. The results obtained in Chapter 2 show that the TO demand for infrastructure use estimates are robust to model inputs. In other words, the

TOs' demand for using the infrastructure does not change much with small changes in the inputs of the model (cost and demand estimates). This suggests that the level of detail of the model is adequate to capture the interactions between the TOs and the IM. This model analyzes each TO independently of other TOs. Once we know TOs' demand for scheduling trains, we need to determine if there is capacity available to schedule all the services. We deal with this question in Chapter 3.

Chapter 3 presents train timetabling problem for shared railway systems that determines which of these trains can be scheduled on the tracks considering the topology of the line, safety constraints, and other technical aspects of the infrastructure. This model explicitly captures the interrelation between infrastructure operations and available infrastructure capacity. From a computational standpoint, the size of the IM Model increases rapidly with the number of stations and the number of trains to schedule. This thesis proposes a novel solution algorithm based on a linear programming approach to approximate dynamic programming (QARLP algorithm) to be able to solve the problem in meaningful instances. This algorithm allows us to decompose and solve large problems that are intractable with MILP commercial solvers while still converging to a solution within an optimality gap.

The economic literature have long suggested the use of traditional mechanisms that price capacity with the marginal infrastructure costs and use simple priority rules to allocate capacity in case of conflict. In Chapter 2 we compare the operational decisions of vertically-integrated railway systems with those of a vertically-separated profit-maximizing TO. We show that the operational decisions of a profit-maximizing TO match the decisions of an even-handed railway industry regulator when variable access

charges reflect variable IM costs to operate trains on the infrastructure. These results justify the use of traditional mechanisms to price and allocate capacity that have been adopted in most countries.

However, Chapter 2 also shows that there are two cases in which these mechanisms cannot be used: 1) when the IM needs to recover part of the infrastructure management fixed costs as it happens in Tanzania, or 2) when the railway system is congested as it happens in many US corridors. In fact, most railway systems fall into at least one of these two categories. This motivates the choice of the two case studies of this thesis and the use of the framework developed to understand the trade-offs associated with the use of alternative mechanisms in these cases. To our knowledge, this is the first effort to compare alternative mechanisms to price and allocate capacity in the same shared railway system.

We illustrate the first case in Chapter 4 in the case of the Central Corridor in Tanzania. Due to the low number of trains operated in the system today, infrastructure maintenance costs do not increase (for all practical purposes) when more trains are operated. Therefore, maintenances costs are assumed fixed. If access charges are set following the traditional approach, operators would not have to pay to access the infrastructure. However, it is critical to ensure that the IM is able to raise revenues to maintain the infrastructure and keep the system operational. As a result, the IM has to allocate infrastructure fixed costs among TOs through the access charges. Chapter 4 first shows that the introduction of variable access charges distorts the operational decision of TOs and then discusses how to avoid this problem with other price-based mechanisms such as the introduction of fixed access charges. We also discuss how to allocate

infrastructure costs among different types of TOs and conclude that charging different access charges to different types of TOs is beneficial for all the stakeholders. Chapter 4 also discusses why the introduction of capacity-based mechanisms in non-congested shared railway systems does not allow the IM to recover costs.

Chapter 5 then analyzes alternative capacity pricing and allocation mechanisms in the context of the very congested NEC, in the US. In this case, we need both the TO Model to anticipate how each TO will respond to capacity pricing and allocation mechanisms and the IM Model to determine the final allocation of infrastructure capacity. Until now, capacity pricing and allocation in the corridor has been managed via bi-lateral contracts negotiated between the IM and the TOs. However, the limitations of this negotiation-based mechanism motivated the FRA's request to Amtrak and the rest of the commuters and freight railway companies to agree on a new capacity pricing and allocation mechanism by the end of 2015. Chapter 5 considers two representative mechanisms to price and allocate railway capacity: a price-based (cost-allocation and priority-rule) mechanism proposed by Amtrak and a capacity-based (auction) mechanism. The results of Chapter 5 show that there are important trade-offs associated with each mechanism and none of them is superior to the other on all dimensions. NEC stakeholders should carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking the system into one of them.

6.2 Conclusions and Recommendations

At the beginning of this thesis we mentioned that the implementation of shared railway systems requires the design and implementation of capacity pricing mechanisms. These mechanisms are the rules needed for coordinating the multiple agents that share the

infrastructure. In this thesis, we analyze the performance of shared railway systems under alternative mechanisms to price and allocate railway capacity. This section summarizes three main conclusions of this work and recommends some courses of action bases on these conclusions.

The first conclusion of this research is that the implications of capacity pricing and allocation mechanisms for shared railway systems are still unclear. While this thesis tries to offer clarity in this area, there is still much work to be done. In that sense, we join (Drew and Nash, 2011; Nash, 2010) in recommending to academics that they invest in this research topic. Any progress in research that contributes to a better understanding of the implications of alternative mechanisms to price and allocate capacity could immensely help practitioners and policy makers. This is particularly important in a context in which several countries are currently restructuring their railway sector to allow shared use.

The second conclusion of this research is that sharing railway infrastructure capacity is not straightforward. In the railway industry, as compared to other network industries, there are very strong interactions between capacity planning and infrastructure operations. Chapter 5 shows that capturing this interactions is critical to implement capacity-based mechanisms and to understand the implications of price-based mechanisms in the railway industry. Despite these differences, regulators and policy makers rely on the lessons learned from other network industries. Although these lessons are useful and can serve as guiding principles to design mechanisms to price and allocate railway capacity, Chapter 2 shows that they often do not work in the railway industry. We thus recommend that policy makers are cautious and question the validity of

assumptions based on other industries. We also recommend to academics that they reach other communities beyond their domains doing research in these topics. A better understanding of what works and what does not work across network industries and why would also be very valuable for practitioners and policy makers.

The third conclusion, on a more positive note, is that the implementation of adequate capacity pricing and allocation mechanisms can mitigate the coordination problems of shared railway systems while maintaining the benefits of shared infrastructure use in the railway industry. Chapter 2 shows that the introduction of TO competition enabled by shared use may have similar effects to the introduction of regulation to ensure that TOs behave as even-handed integrated railway companies. Chapter 4 shows that shared use may allow the IM to recover more infrastructure costs than those recovered in dedicated corridors by enabling the entrance of new TOs that offer profitable services that the current TO does not provide. In a context in which the NEC and many other systems are moving ahead with the implementation of new capacity pricing and allocation mechanisms, we conclude with three more recommendations. This research shows important trade-offs among alternative mechanisms to price and allocate railway capacity. We recommend the use of the framework developed in this thesis to identify *personalized* mechanisms to price and allocate capacity, aimed at the specific characteristics of the systems. At the same time, we recommend that practitioners and policy makers consider alternative mechanisms to price and allocate railway capacity before locking their system into one of them. Even if those cases where stakeholders have to make a decision soon, we recommend that they allow for some flexibility to adapt the mechanism implemented as they gather more information and better understand the

implications of alternative mechanisms for their systems. Finally, we recommend railway companies and regulators that they measure the performance of their systems using a wide variety of performance metrics and to share information and best practices with other railway systems. The data form these experiments will contribute to the improved understanding and management of shared railway systems.

6.3   Future Research

In this section we identify four lines of future research that we find particularly relevant to better understand the implications of capacity pricing and allocation mechanisms in shared railway systems. These lines include 1) additional validation of the framework developed, 2) the extension of the models and algorithms proposed, 3) the utilization of the framework developed to answer other related and relevant shared railway systems research questions, and 4) the development of a broader understanding of capacity pricing and allocation across network industries.

There are three different ways to validate the framework developed in this thesis. First we could validate the models with the results obtained. This thesis uses current operational data to calibrate the models. This prevents us from using that same data to validate the models. Nonetheless, this thesis extensively discusses the results obtained, compares it with industry benchmarks, and carries out multiple sensitivity analysis to understand the sensitivity of the results to the models' inputs. Second, we could validate the models comparing them to other models already validated in the field. In that sense, the TO Model presented is based on a TO financial model used extensively to analyze investments on the railway system. The IM Model presented is also based on train timetable models widely adopted in the railway industry. Third, we could further validate

the models using them to analyze additional capacity pricing and allocation mechanisms in other shared railway systems. In this direction, the author collaborates with other students in the MIT Regional Transportation and HSR Research group that are using a similar framework to analyze the performance of shared railway systems in California and in Europe. California is working towards the implementation of a *blended* HSR system in which commuters and high-speed trains will share the infrastructure in San Francisco (Levy, 2015). Many European countries have experienced changes in the mechanisms to price and allocate railway capacity in the last 10 years (Prodan, 2015) and thus represent excellent natural experiments to calibrate and further validate the framework proposed in this thesis. The author has also participated in studies to implement shared railway systems in Tanzania and India in collaboration with the World Bank. Future data from any of these projects would be useful to further validate this framework and to improve our understanding of the implications of capacity pricing and allocation mechanisms for shared railway systems.

The results of the thesis show that there are important trade-offs between alternative mechanisms. Consequently, we recommend shared railway systems' stakeholders that they carefully analyze the implications of alternative capacity pricing and allocation mechanisms before locking their system into one of them. We discuss three possible directions to extend the models and the algorithm developed in this thesis to further analyze these issues, together with the main challenges that we envision:

- The TO Model proposed can be extended to capture more institutional, technical, and operational details of the shared railway systems. This extensions could consider, for instance, the effects of modeling multiple OD

pairs, substitutable services, different time periods during the day, more detailed demand models, minimum number of service constraints imposed by Public Service Obligations, maximum number of services due to rolling stock limitations, funding availability, etc. As we discuss in Chapter 2, these extensions would require more information about TOs' costs and revenues and about the end users' transportation demand. However, they would also allow us to anticipate TOs' operational decisions with more accurate estimates of the train capacity and the interactions between services. The extensions discussed would not impose a challenge to solve the underlying TO Model optimization problem and determine the TO operational decisions. This thesis solves this problem analytically. Most of the extensions discussed would facilitate the determination of the operational decisions by imposing bounds on the feasible space. Others may require the use of numerical optimization methods, but there is ample room to add complexity to this model from a computational standpoint. This thesis focuses on the TO-IM relation to capture the interactions between infrastructure operations and available infrastructure capacity. These extensions would connect this work with a broad field of research that studies the details of the end users-TO relation (Bebiano et al., 2014; Ben-Akiva and Lerman, 1985; Lago et al., 1981; Morrison, 1990).

- While the IM Model proposed is fairly comprehensive, this thesis uses it to analyze the allocation of capacity in simple instances (the infrastructure details are aggregated and only consider a few stations). The IM Model can be

easily parameterized to consider the topology of the infrastructure in more detail. However, the size of the model increases rapidly with the number of stations and services. The instances analyzed in this thesis are at the limit of what current MILP commercial solvers can handle. The two main options to handle this increased dimensionality are 1) to use the timetables in simple instances as a starting point to develop more detailed timetables and 2) to use the novel algorithm proposed in Chapter 3 to solve the resulting optimization problem.

- Finally, in this direction, the algorithm proposed in Chapter 3 has proven promising to solve large scale network optimization problems. Chapter 3 shows that choice of the basis functions takes advantage of the problem structure and is thus problem specific. Further research to study this algorithm, to develop principles to choose the basis functions, and to incorporate the algorithm in commercial packages that could be used by practitioners would be extremely useful to efficiently solve the train timetabling problem and other large network optimization problems.

This thesis presents a framework that allows regulators, IMs, and TOs to analyze and compare alternative capacity pricing and allocation mechanisms. This framework could be used in the future to analyze two other related and relevant aspects of shared railway systems. From a prescriptive standpoint, the ability to analyze and understand alternative mechanisms to price and allocate railway capacity is critical to design effective mechanisms to coordinate shared railway systems. From a descriptive standpoint, the extension of the results of this thesis considering that TOs in railway

178

systems generally have market power is also critical to analyze, compare, and design capacity pricing and allocation mechanisms.

There is a broad field of research that analyzes the design of mechanisms to price and allocate capacity in network industries (Affuso, 2003; Greve and Pollitt, 2013; McDaniel, 2003; Newbury, 2003; Parkes, 2001; Perennes, 2014; Stern and Turvey, 2003; Vazquez, 2003). The desirable properties of a good mechanism to price and allocate railway capacity are: strategy proofness, allocative efficiency, individual rationality, budget balance (Greve and Pollitt, 2013), and transparency (Vazquez, 2003). This thesis shows that the important interactions between infrastructure operations and available infrastructure capacity affect the performance of shared railway system under alternative capacity pricing and allocation mechanisms. As a result, it is important to design mechanisms that consider these interactions and respond to the overarching goals of the system's stakeholders (Perennes, 2014), and to use the framework proposed in this thesis together with the expertise of the mechanism design field to the implications of such mechanisms for the performance of shared railway systems.

With respect to the second question, the integration of the models developed in this thesis assumes that TOs reveal the IM their real demand for scheduling trains and their willingness to pay for access. As discussed in Chapter 5, this assumption is not valid when TOs and IMs exercise market power. The framework proposed in this thesis could be complemented with game theory and industrial organization concepts to analyze the performance of shared railway systems under mechanisms to price and allocate capacity when TOs and IMs exercise their market power. This line of research would rely on behavioral economic models of the TOs and IMs to capture their response. These models

would determine how the outputs of the TO Model relate with the inputs of the IM Model in equilibrium. The comparison of these results with the results obtained in this thesis are important to understand the impact of the market power of railway system stakeholders in the performance of shared railway systems. These findings will also have important implications for the design of efficient mechanisms to price and allocate railway capacity.

Finally, the work of this thesis is part of a larger research field that analyzes infrastructure planning, management, and operations in different network industries (Gomez-Ibanez, 2003; Jacquillat, 2015; Laffont and Tirole, 1993; Laffont and Tirole, 200; Perez-Arriaga, 2013; Sussman, 2000; Vaze, 2011). Although the practical experiences of regulators in one network industry are used in practice when regulating other network industries; the research bodies on different network industries are mostly disconnected. Any efforts 1) to identify lessons that are transferable among industries and 2) to understand how differences in the institutions and in the characteristics of the physical systems among industries translate into modeling differences would be very useful to further connect these bodies of research. We envision that these efforts could be both deductive, using conceptual models of different network industries to analyze the responses of such industries to alternative capacity pricing and allocation mechanisms; and inductive, analyzing the response to alternative capacity pricing and allocation mechanisms in different case studies across network industries.

In a context in which shared-use is becoming more relevant in our economies, this thesis analyzes the prospects to effectively share railway infrastructure. We discuss that while sharing railway infrastructure is not straightforward; understanding alternative rules to coordinate agents is a first step to being able to design adequate rules that unlock

the potential benefits of shared use. We thank readers for their attention and hope that some parts of this thesis can help in the design and analysis of effective rules to share infrastructure in network industries.

Appendix A - The Infrastructure Manager Problem: Detailed Train Timetable

This appendix includes the detailed timetable of the first case study presented in Chapter 3 and represented in Figures 3-4 and 3-5.

Intercity train timetable

Table A-1 Intercity train timetable. 12:00pm corresponds to minute 0 in Figures 3-4 and 3-5 (Source: author)

Intercity	
Station	**Train 1**
Station 1	12:00 PM
Station 2	12:25 PM
Station 3	12:50 PM
Station 4	1:15 PM
Station 5	1:40 PM
Station 6	2:05 PM
Station 7	2:30 PM
Station 8	2:55 PM
Station 9	3:20 PM
Station 10	3:45 PM
Station 11	4:10 PM
Station 12	4:35 PM
Station 1	4:55 PM

Commuter train timetable

Table A-2 Commuter train timetable. 12:00pm corresponds to minute 0 in Figures 3-4 and 3-5 (Source: author)

Commuter_1

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18
Station 1	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:08 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM
Station 2	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:44 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM	6:24 PM	6:56 PM	7:28 PM
Station 12	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:58 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM
Station 1	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:34 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:48 PM	9:20 PM

Station	Train 19	Train 20	Train 21	Train 22	Train 23	Train 24	Train 25	Train 26	Train 27	Train 28	Train 29	Train 30	Train 31	Train 32	Train 33	Train 34	Train 35	Train 36
Station 1	6:24 PM	6:56 PM	7:28 PM	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:56 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 2	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:32 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM
Station 12	8:16 PM	8:48 PM	9:20 PM	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:48 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 1	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:24 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM

Commuter_2

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18	Train 19	Train 20	Train 21	Train 22
Station 3	5:04 AM	5:36 AM	6:08 AM	6:40 AM	7:12 AM	7:44 AM	8:16 AM	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:48 PM	3:12 PM	3:44 PM	4:16 PM
Station 4	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:08 PM	4:40 PM	5:04 PM	5:36 PM	6:08 PM
Station 5	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:44 PM	6:16 PM	6:40 PM	7:12 PM	7:44 PM
Station 9	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	6:00 PM	6:32 PM	6:56 PM	7:28 PM	8:00 PM
Station 10	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:52 PM	8:24 PM	8:48 PM	9:20 PM	9:52 PM
Station 11	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:48 PM	9:28 PM	10:00 PM	10:24 PM	10:56 PM	11:28 PM

Station	Train 23	Train 24	Train 25	Train 26	Train 27	Train 28	Train 29	Train 30	Train 31	Train 32	Train 33	Train 34	Train 35	Train 36	Train 37	Train 38	Train 39	Train 40	Train 41	Train 42	Train 43
Station 3	4:48 PM	5:20 PM	5:52 PM	6:24 PM	7:04 PM	7:28 PM	8:00 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 4	6:40 PM	7:12 PM	7:44 PM	8:16 PM	8:56 PM	9:20 PM	9:52 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 5	8:16 PM	8:48 PM	9:20 PM	9:52 PM	10:32 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM
Station 9	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:48 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM	5:36 AM	6:08 AM	6:40 AM	7:12 AM
Station 10	10:24 PM	10:56 PM	11:28 PM	12:08 AM	12:40 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM
Station 11	12:00 AM	12:32 AM	1:04 AM	1:44 AM	2:16 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM	7:28 AM	8:00 AM	8:32 AM	9:04 AM	9:36 AM	10:08 AM	10:40 AM

Commuter_3

Station	Train 1	Train 2	Train 3	Train 4	Train 5	Train 6	Train 7	Train 8	Train 9	Train 10	Train 11	Train 12	Train 13	Train 14	Train 15	Train 16	Train 17	Train 18
Station 7	8:48 AM	9:20 AM	9:52 AM	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:56 PM	5:20 PM	5:52 PM
Station 8	10:24 AM	10:56 AM	11:28 AM	12:00 PM	12:32 PM	1:04 PM	1:36 PM	2:08 PM	2:40 PM	3:12 PM	3:44 PM	4:16 PM	4:48 PM	5:20 PM	5:52 PM	6:32 PM	6:56 PM	7:28 PM
Station 6	10:40 AM	11:12 AM	11:44 AM	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:48 PM	7:12 PM	7:44 PM
Station 7	12:16 PM	12:48 PM	1:20 PM	1:52 PM	2:24 PM	2:56 PM	3:28 PM	4:00 PM	4:32 PM	5:04 PM	5:36 PM	6:08 PM	6:40 PM	7:12 PM	7:44 PM	8:24 PM	8:48 PM	9:20 PM

Station	Train 19	Train 20	Train 21	Train 22	Train 23	Train 24	Train 25	Train 26	Train 27	Train 28	Train 29	Train 30	Train 31	Train 32	Train 33	Train 34	Train 35	Train 36
Station 7	6:24 PM	6:56 PM	7:28 PM	8:08 PM	8:32 PM	9:04 PM	9:36 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM
Station 8	8:00 PM	8:32 PM	9:04 PM	9:44 PM	10:08 PM	10:40 PM	11:12 PM	11:44 PM	12:16 AM	12:48 AM	1:20 AM	1:52 AM	2:24 AM	2:56 AM	3:28 AM	4:00 AM	4:32 AM	5:04 AM
Station 6	8:16 PM	8:48 PM	9:20 PM	10:00 PM	10:24 PM	10:56 PM	11:28 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM
Station 7	9:52 PM	10:24 PM	10:56 PM	11:36 PM	12:00 AM	12:32 AM	1:04 AM	1:36 AM	2:08 AM	2:40 AM	3:12 AM	3:44 AM	4:16 AM	4:48 AM	5:20 AM	5:52 AM	6:24 AM	6:56 AM

References

ADB, (2014). *Managing for development results: Rail infrastructure tariffs – enabling private development in Mongolia's railway sector*. Asian Development Bank.

Affuso L., (2003). Auctions of rail capacity? *Utilities Policy*, Vol. 11, pp. 43-46.

Agosta B., (2015). *Development of a methodology for evaluating investments in infrastructure for the sustainable exploitation of shale energy in Argentina*. MIT Master Thesis.

AICD, (2008). *Transport prices and costs in Africa: A review of the main international corridors*. World Bank Working Paper 14.

Amjadi, A., A.J. Yeats, (1995). *Have transport costs contributed to the relative decline of Sub-Saharan African exports?* World Bank Policy Research Working Paper 1559.

Amtrak, (2010). *The Northeast Corridor infrastructure master plan*. NEC Master Plan Working Group: Washington, DC.

Amtrak, (2012). *The Amtrak vision for the Northeast Corridor: 2012 update report*. Washington, DC.

Amtrak, (2014). *FY2014 budget and business plan*. *FY 2015 budget request justification*. *FY2014-2018 five year financial plan*. Washington, DC.

Amtrak, (2015). *Amtrak system timetable*. Washington, DC.

Archila A.F., (2013). *Intercity Passenger Rail Productivity in the Northeast Corridor: Implications for the Future of HSR*. MIT Master Thesis.

Arvis J.F., G. Raballand, J.F. Marteau, (2010). *The cost of being landlocked: logistics costs and supply chain reliability*, World Bank.

Barrow K., (2012). SNCF fined 60.9m euro for anti-competitive behavior. *International Railway Journal*, December 18, 2012.

Bebiano B.M., O. Petrik, F. Moura, R.F. Allard, J. de Abreu e Silva, (2014). *Simulating the impact of the high-speed network on the performance of other transport operators: the HSR on the Lisbon-Oporto link*. In: Proceedings of the 93rd Annual Meeting of the Transportation Research Board. Washington, DC.

Belli P., J. Anderson, H. Barnum, J. Dixon, J.P. Tan, (2001). *Economic analysis of investment operations*. World Bank Institute, Washington, DC.

Ben-Akiva, M.E., S.R. Lerman, (1985). *Discrete choice analysis: theory and application to travel demand*. (Vol. 9). MIT Press.

Bertsekas D., (2006). *Dynamic programming and stochastic control*. Ed. Athena Scientific, Belmont, MA.

Bertsekas D., J.N. Tsitsiklis, (1996). *Neuro dynamic programming*. Ed. Athena Scientific, Belmont, MA.

Big Results Now, (2013). *Executive summary of the transport lab national key results area report*, 2013. Big Results Now, Tanzania Development, Vision 2025, Tanzania.

Borkar V., (1988). A convex analytic approach to Markov decision processes. *Probability Theory Related Fields*, Vol. 78, pp. 583-602.

Cacchiani V., A. Caprara, P. Toth, (2010). Non-cyclic train timetabling and comparability graphs. *Operations Research Letters*, Vol. 38, No. 3, pp. 179-184.

Caimi G., D. Burkolter, T. Herrmann, F. Chudak, M. Laumanns, (2009). Design of a railway scheduling model for dense services. *Networks and Spatial Economics*, Vol. 9, pp. 25-46.

Caimi G., F. Chudak, M. Fuchsberger, M. Laumanns, R. Zenklusen, (2011). A new resource-constrained multicommodity flow model for conflict-free train routing and scheduling. *Transportation Science*, Vol. 45, No. 2, pp. 212-227.

Caprara A., M. Fischetti, P. Toth, (2002). Modeling and solving the train timetabling problem. *Operations Research*, Vol. 50, No. 5, pp. 851-861.

Caprara A., L. Kroon, P. Toth, (2011). Optimization problems in passenger railway systems. *Wiley Encyclopedia of Operations Research and Management Science*.

Castillo E., I. Gallego, J. Ureña, J. Coronado, (2009). Timetabling optimization of a single railway track line with sensitivity analysis. *TOP*, Vol. 17, No. 2, pp. 256-287.

Clewlow, R.R.L., (2012). *The climate impacts of HSR and air transportation in the Northeast Corridor*. MIT Ph.D. Thesis.

Cordeau J.F., P. Toth, D. Vigo, (1998). A survey of optimization models for train routing and scheduling. *Transportation Science*, Vol. 32, No. 4, pp. 380-404.

CPCS, (2013). *Financial and economic analysis of Tanzania Intermodal and Rail Development Project (TIRP)*, Report prepared for RAHCO and The World Bank, CPCS Report 12124, Canada.

Crozet Y., (2004). European railway infrastructure: towards a convergence of infrastructure charging? *International Journal of Transport Management*, Vol. 2, pp. 5-15.

Czerny A.I., (2010). Airport congestion management under uncertainty. *Transportation Research Part B*, Vol. 44, pp. 371-380.

DB, (2009). *Feasibility study for the Isaka-Kigali/Keza-Gitega-Musongati railway project*. Study Summary Report prepared for Multinational Tanzania, Rwanda, Burundi, DB, Germany.

De Ghellinck G., (1960). Les problemes de decisions sequentielles. *Cahiers du Centre d'Etudes de Recherche Operationnelle,* Vol. 2, pp. 161-179 (In French).

Drew J., (2006). *Rail freight: The benefits and cost of vertical separation and open access*. Ms. Thesis, Institute for Transport Studies, University of Leeds.

Drew J., C.A. Nash, (2011). *Vertical separation of railway infrastructure - does it always make sense?* Institute for Transport Studies, University of Leeds, Working Paper 594.

Farias D.P., B. Van Roy, (2003). The linear programming approach to approximate dynamic programming. *Operations Research*, Vol. 51, No. 6, pp. 850-865.

Farias D.P., B. Van Roy, (2004). On constraint sampling in the linear programming approach to approximate dynamic programming, *Mathematics of Operations Research*, Vol. 29, No. 3, pp. 462-478.

FRA, (2009). *Metrics and standards for intercity passenger rail service*. Federal Railroad Administration, Department of Transportation, FRA-2009-0016.

Gardner S., (2013). Creating capacity for growth – new approaches for managing the future of the NEC. Socioeconomic and Financial Aspects of Intercity Passenger Rail Subcommittee. *2013 Annual Meeting of the Transportation Research Board*, Washington, D.C.

Ghoseiri K., F. Szidarovszky, M.J. Asgharpour, (2004). A multi-objective train scheduling model and solution. *Transportation Research, Part B: Methodological*, Vol. 38, No. 10, pp. 927-952.

Gibson S., (2003). Allocation of capacity in the railway industry. *Utilities Policy,* Vol. 11, pp 39-42.

Gomez-Ibanez J.A., (2003). *Regulating infrastructure: monopoly, contracts, and discretion*. Harvard University Press, Cambridge, MA, USA.

Gomez-Ibanez J.A., G. de Rus, (2006). *Competition in the railway industry: an international comparative analysis*. Edward Elgar Editors, Northampton, MA, USA.

Greve T., M.G. Pollitt, (2013). Network Procurement Auctions. Cambridge Working Papers in Economics.

Jacquillat, A., (2015). *Integrated allocation and utilization of airport capacity to mitigate air traffic congestion*. MIT Ph.D. Thesis.

Kawakami, S., (2014). *Application of a Systems-Theoretic Approach to Risk Analysis of High-Speed Rail Project Management in the US*. MIT Master Thesis.

Kroon L., D. Huisman, E. Abbink, P.J. Fioole, M. Fischetti, G. Maroti, A. Schrijver, A. Steeenbeek, R. Ybema, (2009). The new Dutch timetable: The OR revolution. *Interfaces*, Vol. 39, No. 1, pp. 6-17.

Krueger H., P.A. Farrington, H. Black Nembhard, D.T. Sturrock, G.W. Evans, (1999). *Parametric modeling in rail capacity planning*. WSC'99. 1999 Winter Simulation Conference Proceedings. `Simulation - A Bridge to the Future', 2.

Laffont J.J., J. Tirole, (1993). *A theory of the incentives in regulation and procurement*. Cambridge: MIT Press.

Laffont J.J., J. Tirole, (2000). *Competition in telecommunications*. Cambridge: MIT Press.

Lago A.M., P. Mayworm, J.M. McEnroe, (1981). Transit service elasticities: Evidence from demonstrations and demand models. *Journal of Transport Economics and Policy*, Vol. 15, No. 2, pp.99-119.

Levy S., M. Pena-Alcaraz, A. Prodan, J.M. Sussman, (2015). Analyzing the financial relationship between railway industry players in shared railway systems: the train operator's perspective. *Transportation Research Record*, in press.

Levy S., (2015). *Capacity Challenges on the San Francisco Peninsula Corridor--How Local Decisions Have Statewide Impacts*. MIT Master Thesis.

Liebchen C., (2008). The first optimized railway timetable in practice. *Transportation Science*, Vol. 42, No. 4, pp. 420-435.

Liebchen C., L. Peeters, (2009). Integral cycle bases for cyclic timetabling. *Discrete Optimization*, Vol. 6, No. 1, pp. 98-109.

Lopez-Pita A., (2014). *UIC study of European railway infrastructure charges for freight*. UIC Report.

Manne A.S., (1960). Linear programming and sequential decisions. *Management Science*, Vol. 6, No. 3, pp. 259-267.

Martland C.D., (2011). *Toward more sustainable infrastructure: Project evaluation for planners and engineers*. Ed. John Wiley & Sons, Boston, MA.

MBTA, (2013a). *Financial statements, required supplementary information and supplementary information*. Boston, MA.

MBTA, (2013b). *Commuter rail executive summary*. Boston, MA.

MBTA, (2015). *MBTA commuter rail schedules and maps*. Boston, MA.

McDaniel T., (2003). Auctioning access to networks: evidence and expectations. *Utilities Policy*, Vol. 11, pp. 33-38.

Morrison S.A., (1990). The value of Amtrak. *Journal of Law and Economics*, Vol. 33, No. 2, pp. 361-382.

Munnell A., (1992). Policy watch: Infrastructure investment and economic growth. *The Journal of Economic Perspectives*, Vol.6, No. 4, pp. 189-198.

Nash C.A., (2003). *Best practices in shared-use high-speed rail systems*. Mineta Transport Institute Report.

Nash C.A., (2005). Rail infrastructure charges in Europe. *Journal of Transport Economics and Policy*, Vol. 39, No. 3, pp. 259-278.

Nash C.A., (2010). European rail reform and passenger services – the next steps. *Research in Transportation Economics*, Vol. 29, No. 1, pp. 204-211.

NEC Commission, (2014). *Northeast Corridor commuter and intercity rail cost allocation policy*. Passenger Rail Investment and Improvement Act of 2008, Section 212(c), Amended on December 17, 2014.

Newbury D.M., (2003). Network capacity auctions: promise and problems. *Utilities Policy*, Vol. 11, pp. 27-32.

Olievschi V.N., (2013). *Rail transport: Framework for improving railway sector performance in Sub-Saharan Africa*. World Bank, Sub-Saharan Africa Transport Policy Program (SSATP), Working Paper No. 94.

Parkes D.C., (2001). *Iterative combinatorial auctions: Achieving economic and computational efficiency*. UPenn Ph.D. Thesis.

Pena-Alcaraz M., A. Fernandez, A.P. Cucala, A. Ramos, R.R. Pecharroman, (2011). Optimal underground timetable design based on power flow for maximizing the use of regenerative-braking energy. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, Vol. 226, No. 4, pp. 397-408.

Pena-Alcaraz M., S.J. Carlson, A.F. Archila, N. Stein, J.M. Sussman, (2013). *Analysis of High-Speed Rail Implementation Alternatives in the Northeast Corridor: the Role of Institutional and Technological Flexibility*. Proceedings of the 92nd Annual Meeting of the Transportation Research Board. Washington, DC.

Pena-Alcaraz M., J.M. Sussman, I. Perez-Arriaga, (2014). Capacity pricing schemes to

implement open-access rail in Tanzania. *Network Industry Quarterly*, Vol. 16, No. 2, pp. 18-22.

Pena-Alcaraz M., (2015). Capacity pricing and allocation in shared railway systems. *Intercity Rail Passenger Systems Update*, TRB, No. 23, pp.13-15.

Pena-Alcaraz M., M.D. Webster, A. Ramos, (2015a). *Train timetable design for shared railway systems using a linear programming approach to approximate dynamic programming*. Engineering Systems Division Working Paper Series. Submitted for publication.

Pena-Alcaraz M., J.M. Sussman, M.D. Webster, J.I. Perez-Arriaga, (2015b). *Comparison of price-based and capacity-based mechanisms for pricing and allocating railway capacity in the Northeast Corridor*. Engineering Systems Division Working Paper Series. Submitted for publication.

Perennes P., (2014). Use of combinatorial auctions in the railway industry: Can the "invisible hand" draw the railway timetable? *Transportation Research Part A: Policy and Practice*, Vol. 67, pp. 175-187

Perez-Arriaga I.J., L. Olmos, (2009). *A comprehensive approach for computation and implementation of efficient electricity transmission network charges*. MIT Center for Energy and Environmental Policy Research, Vol. 7.

Perez-Arriaga I.J., (2013). *Regulation of the power sector*. Chapters 4, 5, and 6. Edited by Perez-Arriaga, Springer Verlag.

Pouryousef H., P. Lautala, (2015). *Capacity evaluation of directional and non-directional operational scenarios along a multiple-track U.S. corridor*. Proceedings of the 94th Annual Meeting of the Transportation Research Board. Washington, DC.

Powell W., (2007). *Approximate dynamic programming: Solving the curses of dimensionality*, Ed. Willey, NJ.

Pozzo di Borgo P., (2005). *Railway concession experiences in Sub-Saharan Africa*. World Bank.

PPIAF, World Bank, TRS, (2011). *The Railway Reform Toolkit*. World Bank.

PRIIA, (2008). *Passenger Rail Investment and Improvement Act of 2008*. Federal Railroad Administration.

Prodan A., (2015). *Railway infrastructure pricing and interaction of infrastructure managers and train operators*. Working paper.

Raballand G., P. Macchi, (2009). *Transport prices and costs: The need to revisit donor's policies in transport in Africa*. World Bank.

Railistics, (2013). *Institutional assessment of Tanzania railway system report*. Report prepared for RAHCO and The World Bank, Railistics, Tanzania.

Renfe, (2015). *Renfe's punctuality commitment*. Renfe's website www.renfe.com Accessed on March 26, 2015.

Rubinstein R.Y., D.P. Kroese, (2008). *Simulation and the Monte Carlo Method*, Second Edition. John Wiley & Sons, Hoboken, N.J., chapter 6.

Rubio F.J., (1999). *Metodologia de asignacion de costes de la red de transporte en un contexto de regulacion abierta a la competencia*. Ph.D. Thesis, Comillas Pontifical University (in Spanish).

Sakamoto R., (2012). *High speed railway productivity: How does organizational restructuring contribute to HSR productivity growth?* MIT Master Thesis.

Stern J., R. Turvey, (2003). Auctions of capacity in network industries. *Utilities Policy*, Vol. 11, pp. 1-8.

Sussman J.M., (2000). *Introduction to transportation systems*. Artech House, Boston, MA.

Sussman J.M., A.F. Archila, S.J. Carlson, M. Pena-Alcaraz, N. Stein, (2012). *Transportation in the Northeast Corridor of the U.S.: A multimodal and intermodal conceptual framework*. Report for JITI.

Sussman J.M., T.P. Doyle, J.C. Moody, M. Pena-Alcaraz, (2015). *The evaluation of international HSR markets for JR East: A systems approach*. Report for Japan Railways East.

The Economist, (2015). Africa rail and roads: all aboard. *The Economist*, February 28, 2015, Middle East and Africa Edition.

Texeira P.F., A. Lopez-Pita, (2012). *UIC study on railway infrastructure charges in Europe, high-speed*. UIC Report.

Teixeira P.F., A. Prodan, (2014). Railway infrastructure pricing in Europe for high-speed and intercity services: State of the practice and recent evolution. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2448, Railroads, pp. 1-10.

Vaze V., (2011). *Competition and congestion in the National Aviation System: multi-agent, multi-stakeholder approaches for evaluation and mitigation*. MIT Ph.D. Thesis.

Vazquez C., (2003). *Modelos de casacion de ofertas en mercados electricos*. Ph.D. Thesis, Comillas Pontifical University (in Spanish).

Weitzman M.L., (1974). Price vs. quantities. *Review of Economic Studies*, Vol. 41, No. 4, pp. 477-491.

World Bank, (2008). *Infrastructure and growth in developing countries: Recent advances and research challenges*. Development Research Department Report, World Bank.

World Bank, (2014). *Project appraisal document for Tanzania intermodal and rail development project*. World Bank.

Zhou X., M. Zhong, (2005). Bicriteria train scheduling for high-speed passenger railroad planning applications. *European Journal of Operational Research*, Vol. 167, No. 3, pp. 752-771.