3-10-2019

Road Pricing for Congestion Management and Infrastructure Financing

DA XU

University of Windsor

Road Pricing for Congestion Management and Infrastructure Financing

by

Da Xu

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Industrial and Manufacturing Systems Engineering Graduate Program
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada

2019

Road Pricing for Congestion Management and Infrastructure Financing

by

Da Xu

APPROVED BY:

Y. Nie, External Examiner
Northwestern University

W.Li
Odette School of Business

L. Oriet
Department of Mechanical, Automotive & Materials Engineering

A. Alfakih
Department of Mathematics & Statistics

G. Zhang, Co-Advisor
Department of Mechanical, Automotive & Materials Engineering

X. Guo, Co-Advisor
Odette School of Business

January 17, 2019

DECLARATION OF ORIGINALITY

I hereby certify that I am the sole author of this dissertation and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

ABSTRACT

Road pricing has two distinct objectives, to alleviate the congestion problem, and to generate revenue for transportation infrastructure financing. Accordingly, road pricing studies can be roughly classified into two branches with overlapping, one on congestion pricing and the other on toll roads. This dissertation contributes to both branches of road pricing studies. Three topics are discussed. The first two are related with congestion pricing and the third one is related with infrastructure financing.

The first topic is that we study the optimal single-step coarse toll design problem for the bottleneck model where the toll level and toll window length have maximum acceptable upper bounds and the unconstrained optimal solution exceeds the upper bounds. We consider proportional user heterogeneity where users' values of time and schedule delay vary in fixed proportions. Three classic coarse tolling models are studied, the ADL, Laih and braking models. In the ADL model, toll non-payers form a mass arrival at the bottleneck following the last toll payer. In the Laih model, there is a separated waiting facility for toll non-payers to wait until the toll ends. In the braking model, toll non-payers can choose to defer their arrival at the bottleneck to avoid paying the toll. We find that, in the ADL and the Laih models, the optimal solution chooses the maximum acceptable toll level and toll window length. The ADL model further requires the tolling period to be started as late as possible to eliminate the queue at the toll ending moment. In the braking model, if the upper bound of the toll window length is too small, no toll should be charged. Otherwise the optimal solution chooses the maximum acceptable toll window length and may choose a toll price less than the maximum acceptable level.

The second topic is that we develop a new coarse tolling model to address the coarse tolling problem during morning peak hour. An "overtaking model" is proposed by considering that toll payers could overtake those braking commuters (toll non-payers) to pay toll to pass the bottleneck. This would allow commuters to brake and in the meanwhile can make the bottleneck fully utilized during the tolling period, i.e., eliminate the somewhat unrealistic unused tolling period in the braking model. The overtaking model systematically combines the Laih model and the braking model together, capturing both of their properties. Specifically, the overtaking model reduces to the Laih model when the unit overtaking cost approaches zero, and reduces to the braking model when the unit overtaking cost is too high. An unconstrained optimal tolling scheme is developed, and we find out that, unlike the ADL and the Laih models, in the overtaking model, the tolling scheme causing capacity waste could be better than tolling scheme without capacity waste. It is found that, the optimal tolling scheme is affected by the unit overtaking cost. One critical unit overtaking cost is defined. For a small unit overtaking cost, the optimal tolling scheme is similar to that of the Laih model, i.e., featured by no queue exists at the toll starting and ending moments and no capacity waste exists; for a large unit overtaking cost, the optimal tolling scheme is to set the toll high enough to prevent users from overtaking and thereby make the model reduce to the braking model. In the latter case, although the unused tolling period (as in the braking model) can

be fully utilized through lowering the toll to make commuters overtake, the system cost will be increased by doing so.

The third topic is that we investigate the profit maximizing behavior of a private firm which operates a toll road competing against a free alternative in presence of cars and trucks. Trucks differ from cars in value of time (VOT), congestion externality, pavement damage, and link travel time function. We consider mixed travel behaviors of cars and trucks in that trucks choose routes deterministically, while cars follow stochastic user equilibrium in route choice. We derive the equilibrium flow pattern under any combination of car-toll and truck-toll, and identify an integrated equilibrium range within which each road is used by both cars and trucks. We find that, depending on the per-truck pavement damage cost, the firm may take a car-strategy, a truck-strategy, or a car-truck mixed strategy. The perception error of car users, the VOTs and traffic demands of cars and trucks are critical in shaping the firm's strategy.

ACKNOWLEDGEMENT

I would like to express my gratitude to my advisors and committee members for their support and guidance through my PH.D. study.

My greatest thanks to my parents for their caring and support.

I would like to thank all the ones who offered me help to my life or research.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Road pricing are direct charges levied for the use of roads, such as flat tolls, distance-based tolls or time varying tolls. The tolls are aimed to alleviate congestion levels of certain road area or discourage use of certain types of vehicle that can cause serious environment pollution. These toll charges are used to generate revenue for the government or private firms to finance the road infrastructure construction and maintenance. Especially, the toll can be used as a road management tool to control peak hour travel demand as to reduce traffic congestion level. It can also help to reduce other social and environmental negative externalities generated by vehicles, such as air pollution caused by gas emissions, noise and road accidents.

1.1 Road pricing for congestion alleviation

Road congestion causes enormous economic costs. Nowadays, congestion has become a very crucial problem in most urban areas in the world. To expand transport infrastructure alone does not solve the congestion problem, because construction of new roads or adding more capacity to the existing roads can not keep pace with the increase in population and vehicle use. Road pricing as a way of alleviating traffic congestion has become more and more appealing to the policy makers. The theoretical idea of road pricing has dates back to the 1920's (Pigou, 1920; Knight, 1924). In practice, congestion pricing schemes has become popular around the world for decades.

The world's first toll road appeared in Singapore's core central business district in 1975. The tolling scheme is referred as the Singapore Area Licensing Scheme. In 1998, the tolling scheme is converted to a one hundred percent free-flowing Electronic Road Pricing system. To manage traffic demand in the city, Singapore introduced variable pricing based on congestion levels in 2007. The government also levies high annual road tax, charge new vehicle registration fees and implement a quota system for new vehicles.

In 2004, Austria implemented a distance based tolling scheme called Go-Maut for vehicles over 3.5 tons on motorways. If vehicles that is less than 3.5 tons need to enter the Austrian motorway network, they have to buy a sticker. With the sticker, drivers can enjoy paying lower tolls and use most of motorway network in Austria. The toll is set around 8 euros for 10 days. But to use certain routes, such as long tunnels and expensive routes through the Alps, vehicles have to pay extra tolls.

In Norway, electronic urban tolling scheme is implemented on the main road corridors into Bergen in 1986, Oslo in 1990 and Trondheim in 1991. In Bergen, tolling scheme is operated as charging a cordon toll at every entry points that lead to the city central area. In Oslo, the toll was initially intended to generate revenue to finance the road infrastructure construction, but it was found out that tolling reduced traffic demand by around 5%. The Norwegian government announced the rules of congestion tolling in cities in October 2011. The road pricing system in Norway is mainly used to reduce greenhouse gas emissions, control air pollutions and alleviate traffic congestion. In November 2015, motorways into downtown areas in five more cities of Haugesund, Kristiansand, Namsos, Stavanger, and Tonsberg started to implement congestion tolling scheme in Norway.

Sweden introduced the Stockholm congestion tax (also referred to as Stockholm congestion charge) in August 1, 2007. The congestion toll is collected as a tax levied on most vehicles entering and exiting central Stockholm. The primary purpose of the congestion tax is to reduce traffic congestion and environmental pollution in central Stockholm. The toll revenue will be used for road constructions and maintenance in Stockholm. The toll price has been increased since January 2016 in the inner-city parts of Stockholm. Meanwhile, the congestion tolling are also implemented in Essingeleden.

From December 2015, the Beijing Municipal Commission of Transport started planning to introduce congestion charges to the city. The congestion pricing scheme is intended to be a time varying toll based on real time traffic flows and vehicles' emission data. Different vehicles will be charged different fees according to time of the day and areas of the city. The tolling is aimed to alleviate the congestion level and improve the air quality of the city, since vehicle emissions is responsible for 31% of the city's smog sources. In order to control the air pollutions, Beijing has already carried out a driving restriction scheme based upon the last digits on the license plates as well as a vehicle quota system since 2011. The city is now considering to start charging congestion tolls by 2020.

In 1995, US implemented the high occupancy toll lane (HOT lane) system in California's 91 express lanes in Orange County. The 91 express lanes is a ten mile road that have both high occupancy lanes and general toll lanes. The HOT lane is a type of roadway that is free to high occupancy vehicles which should have three or more than three passengers and selected exempt vehicles such as motorcycles and zero emission vehicles. If other vehicles want to use the HOT lane, they are required to pay a toll. On 91 express lanes, there are no toll booths to collect tolls. All vehicles are required to install a transponder to pay the toll electronically. The 91 Express Lanes implements a variable tolling scheme based on the time of day. During peak hours, the toll is much higher than the off-peak hours. For instance, on Friday, the eastbound toll from 3 pm to 4 pm is as high as 13.2 dollars but at 9 pm it drops to 5.2 dollars. The westbound toll from 7 am to 8 am is 10.15 dollars. After 9 am it drops to 6.55 dollars.

Brazil federal government enacted the Urban Mobility Law In January 2012. The law is intended to authorize municipalities to conduct congestion pricing to mange traffic flows and reduce traffic congestion. The enactment of the law is to encourage people to use public transportation as well as reduce air pollution. The law stipulates that revenues generated from congestion tolling can only be used for urban infrastructure like public transportation and non-motorized modes, and to finance public subsidies to transit fares. The law has been in effect since April 2013.

Although road pricing is supported by many solid theories and nowadays has been supported by many advanced technologies for implementation, it also has long been treated as a political issue. Congestion pricing proposals are frequently declined due to public rejection. Examples include Manhattan area in New York and Great Manchester in United Kingdom.

There are several reasons that public cannot accept congestion pricing. For the public, congestion pricing is just like another tax levied by government. As summarized by Cervero (1998), "middle-class motorists often complain they already pay too much in gasoline taxes and registration fees to drive their cars, and that to pay more during congested periods would add insult to injury. In the United States, few politicians are willing to champion the cause of congestion pricing in fear of reprisal from their constituents". Commuters' surplus is transferred to the government in the form of the toll revenue if the congestion pricing is conducted. Thus, commuters are not happy even if the society as a whole (including the government and the users) is benefitted. Furthermore, congestion pricing can cause social inequity problem, i.e. people with relatively low income tend to avoid tolls thus can be deprived of driving freedom. Due to this reason, congestion pricing is often considered as an elitist policy which prices the poor off the roads so that the wealthy can move about unencumbered (Cervero, 1998). In addition, spatial equity issue might also arise, namely that congestion pricing's impact may be different for people living in different areas. For these reasons, congestion pricing as a policy often encounters impedance from the public, and its implementation faces more political issue other than a theoretical or technological problem.

Due to this public acceptance reason, the maximum toll level or tolling period imposed on a road may have upper bounds. A tolling scheme under stringent political constraints that can maximumly alleviate congestion level are in strong need for the government to implement road pricing. Part of this dissertation is devoted to rigorous analysis on how to set up a constrained optimal tolling scheme that can minimize congestion and be accepted by the public as well.

1.2 Road pricing for infrastructure financing

Private provision of public roads such as tunnels and bridges is increasing around the world. These projects are based on build-operate-transfer (BOT) contracts. The BOT contract is defined as that the private sector would build and operate the road at its own expense and in return receive toll revenues for a period of time and then the road will be transferred to the government. The BOT contract helps governments to construct and maintain roads if it has budget constraints, while still retaining public ownership of the roads in the long term. Private toll roads are becoming more popular. A primary reason is that the private sector is more efficient than the public sector, and therefore builds and operates facilities at less cost than the public sector. Also, the public sector, under taxpayers' resistance, is very difficult to finance facilities. In addition, if new road space is provided as an "add-on" to an existing network system, and if road users find it worthwhile to patronize this new road and pay charges, and if the charges cover all costs (including congestion and environmental costs), all may gain benefit, and there would be no obvious losers. Even those who do not use these new roads would benefit from reduced congestion on the old ones (Mills, 1995). For these reasons, the private provision of public roads is becoming more acceptable for the public.

Privately operated toll roads are increasing around the world. For instance, France is well known for its toll roads that 8,000 km of the 11,000 kilometers in France are private toll roads. The toll roads were granted as franchises to private firms. The free roads are directly administered by the national government. Tolls are either a coarse toll or based on the distance driven. The latter case is the most common for long distances. Travellers take a ticket from an automatic machine when they enter the toll road and pay the toll according to the distance when they exit. In Canada, the 108 km Highway 407 ETR through the Greater Toronto Area is a private toll road under a 99-year lease agreement with the Ontario provincial government. The highway collects tolls electronically. Travellers must install a transponder in their car to pay the toll. For those who do not have a transponder in their vehicle are tracked by automatic number plate recognition. Then the toll bill will be sent to the address of the plate on file. Toll rates are determined by both the 407 ETR and the Province of Ontario for each of the respective sections they own. The government set limitations in the 407 ETR lease contract for controlling traffic volumes to make the toll price reasonable and acceptable. While toll has been increasing annually against the requests of the provincial government, resulting in several court battles and the public's complaints that the toll is too high. According to Canadian Shipper (2002), the Ontario highway 407 is charging a very high toll for trucks. Almost no trucks are using highway 407. Now, it is mainly serving commuters and obtaining its profit by attracting passenger cars only. It has been criticized to levy a high toll on trucks to discourage truck use. Regulations from the government is expected to be imposed on highway 407 to attract more trucks to alleviate the truck volume on other roads and improve the transportation network's efficiency.

Profit maximization is typically the goal of the private firm which operates the road. Many factors can affect the private toll roads' profitability, such as infrastructure construction, maintenance costs, operating costs, regulation and political constraints on tolls, and competition from other alternative roads and transportation modes. For the government, understanding the profit-oriented behavior of the firm is necessary for choosing suitable regulations. Part of this dissertation is devoted to investigating how a private firm can achieve profit maximization on a private toll road.

1.3 Objectives

The main objectives of this research are

To study bottleneck coarse tolling problem during morning commute peak hour and propose a constrained optimal tolling scheme that can minimize total system cost under the consideration that due to public acceptance reason, the maximum toll level and toll window length imposed on a bottleneck have upper bounds.

To develop a new coarse tolling model to study coarse tolling by considering that toll payers could overtake those braking commuters (toll non-payers) to pay toll to pass bottleneck. The proposed new model "overtaking model" could capture commuter's travel behavior of braking and overtaking together, thus generating more realistic insights than previous models.

To investigate how a private firm can achieve profit maximization on a private toll road. Understanding such profit-oriented behavior of the firm is necessary for choosing suitable regulations for the government.

1.4 Outline of the dissertation

This dissertation is organized as follows: Chapter 2 carries out a literature review on previous studies related to this dissertation. Besides the basic idea of road pricing, two branches of literatures most relevant to this thesis are reviewed: bottleneck coarse tolling and toll road profit maximization.

In Chapter 3, we investigate how to achieve constrained optimization in bottleneck model. We consider the most realistic case that the political constraint on toll road is stringent: the public maximum acceptable toll window length is less than the unconstrained optimal toll window length and maximum acceptable toll price is less than the unconstrained optimal toll level. Three bottleneck models are studied: mass arrival model, separated waiting lane model and braking model.

In Chapter 4, a new coarse tolling model "overtaking model" is proposed. It is assumed that toll payers could overtake those braking commuters (toll non-payers) to pay toll to pass bottleneck. The overtaking model systematically combines separated waiting lane model and braking model together, capturing both of their properties. An unconstrained optimal tolling scheme of overtaking model is developed.

In Chapter 5, we investigate how to achieve profit maximization of a private toll road which competes against a free alternative in presence of cars and trucks. Trucks differ from cars in value of time (VOT), congestion externality, pavement damage, and link travel time function. We consider mixed travel behaviors of cars and trucks in that trucks choose routes deterministically, while cars follow stochastic user equilibrium in route choice.

Finally, a summary of the major finding of this dissertation and several suggestions for future research are presented in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

Previous studies related to this thesis are reviewed in this chapter. We first briefly review the basic idea of road pricing, and then we focus on two branches of literatures most relevant to this dissertation: bottleneck coarse tolling and toll road profit maximization.

2.1. The fundamental idea of road pricing

The idea of road pricing dates back to the 1920's (Pigou, 1920; Knight, 1924), and later seminal works on road pricing include Wardrop (1952), Walters (1961), Beckmann (1965) and Vickrey (1969). Essentially, the concept of road pricing is based on the fact that road use has negative externality. That is, an additional entry to a congested road not only introduces the private travel cost of the entering user, but also imposes a marginal travel cost on every existing user. To internalize this congestion externality and thereby maximize social welfare, the classic marginal-cost pricing principle states that road users of congested roads should pay a toll equal to the difference between the marginal social cost and the marginal private cost. With this marginal-cost pricing scheme, because each user will face the marginal social cost of road use other than the marginal private cost, the resulting user optimal traffic equilibrium will be exactly the socially optimal one.

The basic concept of road pricing can be easily explained by considering the elastic demand case. In the literature, the marginal-cost pricing principle is best illustrated by the demand-supply curves of the standard single-link case, i.e. a homogeneous traffic stream moving along a given uniform stretch of road (e.g. Walters, 1961; Evans, 1992; Hills, 1993). Specifically, when travel demand is elastic, the inverse demand function represents the marginal social benefit function, and at the "do-nothing" demand-supply equilibrium, it is easy to see that the marginal social cost is larger than the marginal social benefit, and thus a proper toll charge (the marginal-cost pricing scheme) should be introduced to reduce the travel demand to the socially optimal level.

This section is just a very brief review on the basic concept of road pricing, and the following sections are focused on several specific (and perhaps narrow) areas with direct relations to this thesis. Therefore, it should be mentioned here that more comprehensive reviews or monographs on road pricing studies include (certainly not limited to) Lo and Hickman (1997), Button and Verhoef (1998), Verhoef (1996), McDonald et al. (1999), Levinson (2002), Yang and Verhoef (2004), Santos (2004), and Yang and Huang (2005).

2.2. Bottleneck coarse tolling

Vickery's bottleneck model (Vickery, 1969) is widely used to study peak hour road congestion. One important property of the bottleneck model is that a continuously time-varying toll can completely eliminate queuing at the bottleneck and achieve the first-best social optimum without increasing anyone's cost. The first-best time-varying toll in the bottleneck model has been extensively studied (see, e.g., Arnott et al., 1994; Arnott and Kraus, 1995; Yang and Huang, 1997; Small and Verhoef, 2007; Van den Berg and Verhoef, 2011; Xiao et at., 2013). In spite of the appealing theoretical property, practical implementation of continuously time-varying tolls is difficult for various reasons such as high cost, high information requirement, and possible confusion caused to road users. In practice, step tolling is more widely adopted, particularly the single-step "coarse tolling", where a constant toll is implemented during part of the peak period.

There is a stream of literature on the bottleneck coarse tolling problem. Arnott et al. (1990, 1993) developed a coarse tolling model (hereafter the ADL model) where a mass of users arrive at the bottleneck when the toll ends. They pointed out that the optimal coarse tolling should be such that the queue at the bottleneck is eliminated at the starting and ending time points of the tolling period. Laih (1994, 2004) proposed a different model (hereafter the Laih model) which avoids mass arrival by allowing toll non-payers to wait for the toll to end on secondary lanes without blocking the toll payers. Xiao et al. (2011) extended the ADL model by providing details of how the queuing profile changes with respect to toll level under heterogeneous VOT assumption. They formulated a non-linear optimization problem to solve for the optimal tolling scheme. Another coarse tolling model is referred to as the "braking" model, independently developed by Lindsey et al. (2012) and Xiao et al. (2012). The braking model considers that, as the tolling period is about to end, users have an incentive to "brake" (stop or speed down) to wait until the toll ends. Van den Berg (2014) adopted all three (ADL, Laih and braking) coarse tolling models and considered three types of user heterogeneity to analyze the welfare and distributional effects of optimal coarse tolling. Chen et al. (2015) adopted the Laih model and proposed an algorithm to solve the optimal multi-step toll problem under general user heterogeneity where the ratios of schedule to queuing delay are different for any two groups of users. Nie (2015) considered all three coarse tolling models and designed different tradable credit schemes under different models. Jia et al. (2016) used the Laih model coarse tolling to manage the morning commute of household travels and analyzed the impact of the school-work schedule difference. Knockaert et al. (2016) adopted the ADL model and examined the welfare gain from differentiating the coarse tolling scheme for two group of users. Ren et al. (2016) proposed coarse tolling models where only part of the users have the braking behavior and the braking behavior causes dropped capacity. Li et al. (2017) investigated the step tolling problem under an activity-based bottleneck model setup where each commuter's travel decision is made through maximizing her own scheduling utility at home and work.

2.3. Toll road profit maximization

Privately operated toll roads are increasing around the world. Profit maximization is typically the goal of the private firm which operates the road. For the government, understanding the profit-oriented behavior of the firm is necessary for choosing suitable regulations. There is an extensive literature on modelling the profit-maximizing behavior of private toll roads (see, e.g., Lindsey and Verhoef, 2001; Yang and Huang, 2005; Small and Verhoef, 2007; Tsekeris and Voß, 2009). However, to the best of our knowledge, profit maximization by a private toll road that serves cars and trucks has not been studied (with the exception of Guo and Xu, 2016). The extension is important for two reasons. First, trucks differ from cars in three major ways: value of time (VOT), congestion externality as measured by passenger car equivalence (PCE), and pavement damage. Previous studies have considered either homogeneous road users or user heterogeneity in VOT only. Second, truck traffic could be an important source of revenue or profit for a toll road operator. For example, the Illinois Tollway implemented a 40% toll increase for trucks in 2015, which was expected to contribute about 60% of its nearly $154 million increase in toll revenue (Chicago Tribune, 2015).

In spite of the importance of trucks in profit generation for (private) toll roads, the literature on toll roads with trucks is mostly focused on system efficiency and policy, not on profit maximization. There is a stream of literature dedicated to truck-only toll lanes and tollways, which studied various aspects of truck toll lanes, including policy and implementation (Samuel et al., 2002), economic and financial feasibility (Holguín-Veras et al., 2003), selection of potential truck-only toll lanes (Chu and Meyer, 2008), and safety benefits (Chu and Meyer, 2010). Most of the research on truck use of toll roads consists of empirical studies and case studies. For example, Zhou et al. (2009) used interviews and survey data to understand truckers' use and non-use of toll roads, Swan and Belzer (2010) studied the truck traffic diversion from the tolled Ohio Turnpike, and a recent report (Geiselbrecht et al., 2015) reviewed many studies on truck use of toll roads in Texas.

A few papers have employed analytical models of toll roads with cars and trucks, all from a socially optimal perspective. Arnott et al. (1992) developed a model with two parallel routes and two user types that could be interpreted as cars and trucks, while they restricted their analyses to two group of car users with different VOTs and/or trip-timing preferences. De Palma et al. (2008) adopted the model to investigate the benefits of separating cars and trucks. They compared the effectiveness of different methods including lane access restrictions, differentiated car and truck tolls, and toll lanes for either cars or trucks. Holguín-Veras and Cetin (2009) used the multinomial logit model to formulate the discrete choice of time of travel for multi-class traffic (cars, small and large trucks) on a single corridor. They computed the socially optimal tolls and discussed the policy implications.

Recently, Guo and Xu (2016) examined thoroughly the profit-maximizing behavior of a private firm which operates a toll road competing against a free alternative in presence of cars and trucks. Trucks differ from cars in VOT, PCE, pavement damage, and link travel time function. They considered deterministic route choice for both cars and trucks, and found that the firm takes either a car-strategy or a truck-strategy for profit maximization. Their results suggest that the truck-to-car VOT ratio, the total traffic demand, and the difference in travel distance between the two roads are critical in shaping the firm's strategy. In particular, they found that attracting truck traffic is likely to reduce the toll road's profitability under practical conditions. This result could be used to explain the pricing strategies of real-world private toll roads such as Highway 407 in Toronto, which has long been criticized that it sets high truck tolls to discourage truck usage (e.g., Canadian Shipper, 2002).

CHAPTER 3

CONSTRAINED OPTIMIZATION FOR BOTTLENECK COARSE TOLLING

In this chapter, we study the optimal single-step coarse toll design problem for the bottleneck model where the toll level and toll window length have maximum acceptable upper bounds and the unconstrained optimal solution exceeds the upper bounds. We consider proportional user heterogeneity where users' values of time and schedule delay vary in fixed proportions. Three classic coarse tolling models are studied, the ADL, Laih and braking models. In the ADL model, toll non-payers form a mass arrival at the bottleneck following the last toll payer. In the Laih model, there is a separated waiting facility for toll non-payers to wait until the toll ends. In the braking model, toll non-payers can choose to defer their arrival at the bottleneck to avoid paying the toll.

3.1. Introduction

To the best of our knowledge, the previous works on bottleneck coarse tolling all adopted an unconstrained setup in the sense that the length of the tolling period (the toll window length) and the toll price level are unbounded. However, in practice, due to public acceptance reason, it is likely that the toll level and the toll window length have maximum allowable upper bounds, i.e., it may be politically unacceptable to charge a toll price that is too high or charge a toll for a period that is too long. Actually, it is very common to consider toll price caps in congestion pricing studies under static user equilibrium. For bottleneck dynamic user equilibrium, it is natural to also consider a toll window length cap. When the toll level and toll window length have upper bounds, the unconstrained optimal coarse tolling may be infeasible (i.e., the unconstrained optimal toll level and toll window length may exceed the upper bounds), and the traditional conclusions and insights regarding optimal bottleneck coarse tolling may not hold.

We consider that the coarse tolling scheme has a maximum acceptable toll level and a maximum acceptable toll window length, both exogenously given by public opinions towards peak hour congestion pricing. Under such a constrained optimization setup, we investigate the problem of total system cost minimization. We focus on the case that the unconstrained optimal toll level and toll window length exceed the upper bounds. We adopt the ADL, Laih and braking models of coarse tolling, and consider proportional user heterogeneity where users have different values of time (VOT) yet their values of schedule delay are proportional to their VOT. We find that, in the ADL and the Laih models, the constrained optimal coarse tolling chooses the maximum acceptable toll level and toll window length, which is consistent with the traditional insight that, because the toll replaces the queuing delay and thereby reduces the total system cost, increasing the toll in both toll price and tolling period will improve the system efficiency. While this is not surprising, an important new insight regarding the ADL model is established: under constrained optimization, because it is impossible to eliminate the queues at both the starting and the ending moments of the tolling period, in the ADL model the priority is to start the tolling period as late as possible to eliminate the queue at the toll ending moment only. This insight is never reported in traditional unconstrained coarse tolling studies. We find that, if the toll window length constraint is too stringent (the upper bound is too small), then any toll price will make the total system cost greater than the no-toll equilibrium, and thus no toll should be charged. When the toll window length constraint is not too stringent, the optimal solution chooses the maximum acceptable toll window length, while the optimal toll price may be an interior solution (i.e., less than the maximum acceptable level).

3.2. No-toll equilibrium

In this section we briefly review the no toll equilibrium of the bottleneck model. Consider there are N users commuting from home to work through a road containing a bottleneck. The bottleneck's capacity is s. When users' arrival rate at the bottleneck is higher than s, a queue will develop. When user's arrival rate is lower than s, the queue at the bottleneck will gradually dissipate. Free-flow travel time is normalized to zero without loss of generality. Users incur a unit cost of α from queuing delay (travel time), i.e., α is the VOT. Let t^* be the preferred arrival time at work of all users. If a user arrives at work before t^* (early arrival), she will incur a schedule early delay cost with a unit cost of β. If she arrives at work later than t^* (late arrival), she will incur a schedule late delay cost with a unit cost of γ. We use the term "travel cost" to represent the sum of queuing delay cost and schedule delay cost, and the term "travel price" to represent the sum of travel cost and toll.

Let $\alpha\left(x\right)$ be the VOT distribution function, which gives the xth user's VOT. Users are ordered in decreasing VOT, i.e., $\alpha\left(x\right)$ is decreasing in x. In this chapter we consider "proportional heterogeneity" such that $\beta\left(x\right)=\eta\alpha\left(x\right)$ and $\gamma\left(x\right)=\mu\alpha\left(x\right)$ for $x\in\left[0,N\right]$, where η and μ are constants satisfying $0<\eta<1<\mu$.

Under proportional heterogeneity, the profile of the no-toll equilibrium is the same as that under homogeneous VOT assumption. In equilibrium, no user can further reduce her travel cost (travel price equals travel cost under no toll) by adjusting her arrival time at the bottleneck. The travel cost of a user consists of two parts: queuing delay cost and schedule early/late delay cost. Let $q(t)$ denote the queue length at time t. Travel cost of the xth user arriving at time t (denoted by $C\left(x,t\right)$) can be given as

$$C(x,t)=\begin{cases}\alpha(x)\dfrac{q(t)}{s}+\beta(x)\left(t^*-t-\dfrac{q(t)}{s}\right), & t+\dfrac{q(t)}{s}\le t^* \text{ early arrival}\\ \alpha(x)\dfrac{q(t)}{s}+\gamma(x)\left(t+\dfrac{q(t)}{s}-t^*\right), & t+\dfrac{q(t)}{s}> t^* \text{ late arrival}\end{cases}$$

Let $\lambda(t)$ denote users' arrival rate at time t. With the fact of $q'(t)=\lambda(t)-s$, the arrival rate can be obtained by setting derivate of $C(x,t)$ with respect to t equal to zero, which gives us

$$q'(t)=\begin{cases}\dfrac{s\eta}{1-\eta}, & t+\dfrac{q(t)}{s}\le t^* \text{ early arrival}\\ -\dfrac{s\mu}{1+\mu}, & t+\dfrac{q(t)}{s}> t^* \text{ late arrival}\end{cases}$$

and

$$\lambda(t)=\begin{cases}\dfrac{s}{1-\eta}, & t+\dfrac{q(t)}{s}\le t^* \text{ early arrival}\\ \dfrac{s}{1+\mu}, & t+\dfrac{q(t)}{s}> t^* \text{ late arrival}\end{cases}$$

The arrival rates of users having schedule early and late delay are obtained as both constants, implying that a user's position in the queue is indeterminate. The first user arrives at the bottleneck at t_q and the last user arrives at $t_{q'}$. Considering the first user and the last user incurred only a schedule early delay cost and a schedule late delay cost, respectively, with the fact of $\alpha(x)(t^*-t_q)=\gamma(x)(t_{q'}-t^*)$ and $t_{q'}-t_q=N/s$, we can obtain $t^*-t_q=\mu N/(\eta+\mu)s$ and $t_{q'}-t^*=\eta N/(\eta+\mu)s$. Figure 3.1 shows the no-toll equilibrium profile.

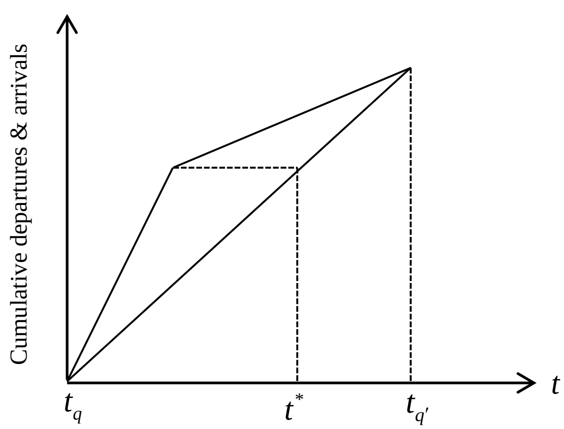

Figure 3.1. No-toll equilibrium profile

3.3. The ADL model

3.3.1 Equilibrium profile under coarse tolling in the ADL model

In this subsection we review the equilibrium under coarse tolling (the ADL model) and develop critical toll levels above which capacity waste happens. A coarse tolling scheme (t^{+}, t^{-}, ρ) is to impose a constant toll ρ from t^{+} to t^{-}. We assume $t^{+} \le t^{*} < t^{-}$ and $t^{-} - t^{+} < N/s$. A constant toll has no impact on users' arrival rates for periods with continuous arrivals. Thus the arrival rate of users having schedule early delay remains $s/(1-\eta)$, and the arrival rate of users having schedule late delay remains $s/(1+\mu)$ except for the mass arrival. In equilibrium, no user can further reduce her travel price (travel cost plus toll) by adjusting her arrival time at the bottleneck. Because of the discontinuity of toll at t^{+} and t^{-}, to satisfy the equilibrium condition, there must be a period that no one arrives at the bottleneck before the first toll-payer arrives, and there must be a mass of individuals arriving at the bottleneck after the last toll-payer arrives.

When the toll window length is not too long and toll level is not too high, the bottleneck can be fully utilized with no capacity waste. Here "capacity waste" means that, within the morning peak period $[t_q, t_{q'}]$, there is a period during which no one uses the bottleneck. Figure 3.2 shows the equilibrium profile under coarse tolling without capacity waste, where $\tilde{t}$ is the arrival time of the last toll non-payer arriving before t^{+}, t_y is the arrival time of the first toll payer, and t_m is the mass arrival time.

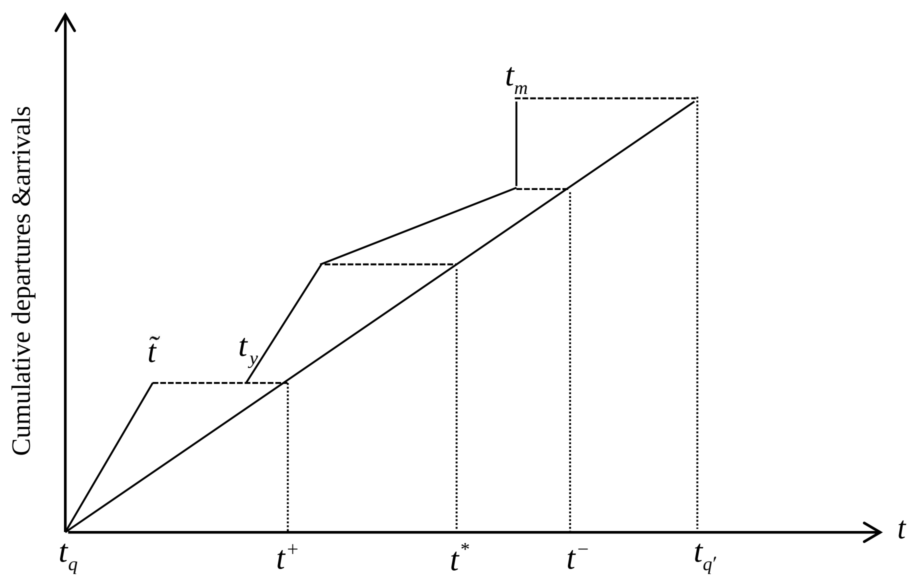

Figure 3.2. Equilibrium profile under coarse tolling: the ADL model

In equilibrium, if the xth user is a toll non-payer, her travel cost is given by (consider the first toll non-payer, arriving at t_q)

$$C^{\text{non}}(x) = \beta(x)\left(t^* - t_q\right) \tag{3.1}$$

and her travel price is $P(x) = C^{\text{non}}(x)$ as she does not pay the toll. Note that while Eq. (3.1) is obtained by considering the first toll non-payer who arrives at t_q, it applies to all non-payers because the arrival time of a user is indeterminate in equilibrium (any toll non-payer could arrive at t_q). If the xth user is a toll payer, her travel cost is given by (consider the first toll payer, arriving at t_y)

$$C^{\text{pay}}(x) = \alpha(x)\left(t^+ - t_y\right) + \beta(x)\left(t^* - t^+\right) \tag{3.2}$$

and her travel price is $P(x) = C^{\text{non}}(x) + \rho$ as she has to pay the toll. Similarly, Eq. (3.2) applies to all toll payers because any toll payer could arrive at t_y in equilibrium.

Under the condition of no capacity waste, the mass arrival time t_m is also the arrival time of the last toll payer, i.e., mass arrival happens immediately following the arrival of the last toll payer. Every user in the mass arrival is assumed to experience an average queuing delay and schedule late delay of the total mass. If the xth user is in the mass arrival, her travel price is given by

$$P(x) = C(x) = \alpha(x)\left(\frac{t^- + t_{q'}}{2} - t_m\right) + \gamma(x)\left(\frac{t^- + t_{q'}}{2} - t^*\right)$$

In equilibrium, there is an indifferent user who can arrive at any time as she always incurs identical travel price. For those who have higher VOT than the indifferent user, they will pay the toll to pass the bottleneck. For those who have lower VOT than the indifferent user, they will avoid the toll by coming earlier or later. Let $V = s\left(t^- - t^+\right)$, which is the number of toll payers under the condition of no capacity waste. Then the indifferent user is the Vth user with VOT $\alpha(V)$.

Capacity waste happens at the beginning of the toll window if $t_y > t^+$, i.e., if the first toll payer arrives later than the toll starting time (no one uses the bottleneck from t^+ to t_y). Thus, $t_y = t^+$ is a critical condition representing that the first toll payer arrives exactly at the toll starting time (capacity waste would happen if the toll is any higher). Similarly, $t_m = t^-$ is a critical condition representing that the last toll payer arrives exactly at the toll ending time (capacity waste would happen at the ending of the toll window if the toll is any higher). In the following we will derive the expressions of t_y and t_m, based on which we can obtain critical toll levels above which capacity waste happens.

We start by deriving $t_{q'}$ and t_q. The indifferent user has equal travel price either being the last toll payer (i.e., arrives at t_m and pays the toll) or joining the mass arrival, thus we have

$$\alpha(V)(t^- - t_m) + \gamma(V)(t^- - t^*) + \rho = \alpha(V)\left(\frac{t^- + t_{q'}}{2} - t_m\right) + \gamma(V)\left(\frac{t^- + t_{q'}}{2} - t^*\right)$$

which gives

$$t_{q'} = \frac{2}{1+\mu}\frac{\rho}{\alpha(V)} + t^- \tag{3.3}$$

In view of $t_{q'} - t_q = N/s$, Eq. (3.3) readily gives

$$t_q = \frac{2}{1+\mu}\frac{\rho}{\alpha(V)} + t^- - \frac{N}{s} \tag{3.4}$$

The indifferent user has equal travel price either being the first toll payer (arrives at t_y) or being the toll non-payer that arrives at $\tilde{t}$, thus we have

$$\alpha(V)(t^+ - t_y) + \beta(V)(t^* - t^+) + \rho = \alpha(V)(t^+ - \tilde{t}) + \beta(V)(t^* - t^+)$$

which gives

$$t_y = \frac{\rho}{\alpha(V)} + \tilde{t} \tag{3.5}$$

Making use of (3.4)-(3.5) and the following relationships between the cumulative arrival and departure

$$\frac{s}{1-\eta}(\tilde{t} - t_q) = s(t^+ - t_q)$$

$$\frac{s}{1-\eta}(t_z - t_y) = s(t^* - t^+)$$

$$\frac{s}{1+\mu}(t_m - t_z) = s(t^- - t^*)$$

we can obtain

$$t_y = \frac{1+\mu+2\eta}{1+\mu}\frac{\rho}{\alpha(V)} + (1-\eta)t^+ + \eta t^- - \eta\frac{N}{s} \tag{3.6}$$

$$t_m = \frac{1+\mu+2\eta}{1+\mu}\frac{\rho}{\alpha(V)} + (1+\eta+\mu)t^+ - (\eta+\mu)t^* - \eta\frac{N}{s} \tag{3.7}$$

We can see from (3.3)-(3.7) that, for a given toll window position (t^+, t^-), all the critical arrival time instants increase with the toll ρ, i.e., all users will postpone their arrival time when the toll increases (the equilibrium profile moves rightward when the toll increases). When the toll is increased to a certain level, the first toll payer will arrive exactly at t^+ ($t_y = t^+$) or the last toll payer will arrive exactly at t^- ($t_m = t^-$). At this moment, if we keep increasing the toll, capacity waste will occur at t^+ or t^-. By setting $t_y = t^+$ in (3.6), we can obtain the critical toll level

$$\rho^{\mathrm{crit1}}\left(t^{+},t^{-}\right)=\frac{\eta\left(1+\mu\right)}{1+\mu+2\eta}\alpha\left(V\right)\left[\frac{N}{s}-\left(t^{-}-t^{+}\right)\right] \tag{3.8}$$

By setting $t_m = t^-$ in (3.7), we can obtain the critical toll level

$$\rho^{\mathrm{crit2}}\left(t^{+},t^{-}\right)=\frac{\eta\left(1+\mu\right)}{1+\mu+2\eta}\alpha\left(V\right)\left[\frac{N}{s}-\frac{\eta+\mu}{\eta}\left(t^{-}-t^{*}\right)\right] \tag{3.9}$$

Comparing (3.8) and (3.9), if the toll window $\left(t^{+},t^{-}\right)$ satisfies $\eta\left(t^{*}-t^{+}\right)>\mu\left(t^{-}-t^{*}\right)$, then $\rho^{\mathrm{crit1}}<\rho^{\mathrm{crit2}}$, which means ρ^{crit1} is more critical and thus capacity waste first happens at t^{+} if we keep increasing the toll level. Similarly, if $\eta\left(t^{*}-t^{+}\right)<\mu\left(t^{-}-t^{*}\right)$, then $\rho^{\mathrm{crit1}}>\rho^{\mathrm{crit2}}$, which means ρ^{crit2} is more critical and thus capacity waste first happens at t^{-} as the toll level increases. If $\eta\left(t^{*}-t^{+}\right)=\mu\left(t^{-}-t^{*}\right)$, then $\rho^{\mathrm{crit1}}=\rho^{\mathrm{crit2}}$, which means capacity waste happens simultaneously at both t^{+} and t^{-} as the toll level increases. Note that η and μ represent the unit schedule early and late delay costs, respectively. Thus, $\eta\left(t^{*}-t^{+}\right)=\mu\left(t^{-}-t^{*}\right)$ means that the part of the toll window before the work start time, $t^{*}-t^{+}$, and the part after, $t^{-}-t^{*}$, have equal monetary value in terms of schedule delay cost.

Definition 3.1. A toll window $\left(t^{+},t^{-}\right)$ is said to be *balanced* if $\eta\left(t^{*}-t^{+}\right)=\mu\left(t^{-}-t^{*}\right)$, *early* if $\eta\left(t^{*}-t^{+}\right)>\mu\left(t^{-}-t^{*}\right)$, and *late* if $\eta\left(t^{*}-t^{+}\right)<\mu\left(t^{-}-t^{*}\right)$.

As will be shown in the next section, a balanced toll window position is optimal for a given toll window length. For an early toll window, there will be no capacity waste as long as $\rho\le\rho^{\mathrm{crit1}}$. For a late toll window, there will be no capacity waste as long as $\rho\le\rho^{\mathrm{crit2}}$. For ease of exposition, we define $\rho^{\mathrm{crit}}\left(t^{+},t^{-}\right)=\min\left\{\rho^{\mathrm{crit1}}\left(t^{+},t^{-}\right),\rho^{\mathrm{crit2}}\left(t^{+},t^{-}\right)\right\}$.

3.3.2. Constrained optimization of coarse tolling in the ADL model

In this subsection, we will look into the constrained optimization problem of coarse tolling, which is to minimize the total system cost subject to constraints on toll window length and toll level. Let $\bar{l}$ and $\bar{\rho}$ be the maximum acceptable toll window length and toll level, respectively. Then the total system cost minimization problem is

$$\min_{t^{+},t^{-},\rho} TC\left(t^{+},t^{-},\rho\right)=\int_{V}^{N}C^{\mathrm{non}}\left(x\right)dx+\int_{0}^{V}C^{\mathrm{pay}}\left(x\right)dx \tag{3.10}$$

subject to

$$t^{-}-t^{+}\le\bar{l} \tag{3.11}$$

$$\rho \leq \bar{\rho} \tag{3.12}$$

In objective function (3.10), $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ are the equilibrium travel costs of toll non-payers and toll payers, respectively, given by (3.1) and (3.2) in the ADL model under the condition of no capacity waste. For ease of exposition, we define $l = t^- - t^+$, the toll window length of tolling scheme (t^+, t^-, ρ). Then $V = ls$. Denote l^{uncon} and ρ^{uncon} the l and ρ of the unconstrained optimal solution to objective function (3.10). To make our analysis and discussion focused, throughout this chapter in all three models, we consider $\bar{l} < l^{\text{uncon}}$ and $\bar{\rho} < \rho^{\text{uncon}}$, i.e., the maximum acceptable toll window length and toll level are less than the first-best values.

We first introduce the following lemma, which allows our analyses to exclude scenarios with capacity waste.

Lemma 3.1. In the ADL model, for any tolling scheme with capacity waste, there exists a tolling scheme with no capacity waste, shorter toll window length, lower toll level, and lower total system cost.

Proof: We give the proof for the case that toll (t^+, t^-, ρ) has capacity waste only at t^+. The proof for the case that capacity waste happens at t^- or at both t^+ and t^- can be done in the same way and thus omitted here. In this case the toll level can be obtained as the indifferent user's travel price difference between arriving at t_q and t_y, which gives us

$$\rho = \alpha(V)\eta(t_y - t_q)$$

By arriving at t_q or joining the mass arrival, the indifferent user has equal travel price, which leads to

$$\eta(t^* - t_q) = \left(\frac{t^- + t_{q'}}{2} - t_m\right) + \mu\left(\frac{t^- + t_{q'}}{2} - t^*\right) \tag{3.13}$$

Let us consider tolling scheme $(t^+_{new}, t^-, \rho_{new})$, where $t^+_{new} = t_y$ and $\rho_{new} = \rho^{\text{crit1}}(t_y, t^-)$. Since ρ_{new} is the critical toll level, there is no capacity waste under $(t^+_{new}, t^-, \rho_{new})$. Besides, no queue exists at t^+_{new} and the toll payers' profile should be exactly same as that under (t^+, t^-, ρ). Considering the indifferent user's travel price, ρ_{new} can be further expressed as

$$\rho_{new} = \alpha(V)\eta(t^+_{new} - t_{qnew})$$

and it is easy to see that

$$\eta(t^* - t_{qnew}) = \left(\frac{t^- + t_{q'new}}{2} - t_m\right) + \mu\left(\frac{t^- + t_{q'new}}{2} - t^*\right) \tag{3.14}$$

Comparing (3.13) and (3.14), if $t_{qnew} \le t_q$, in order to let (3.13) and (3.14) hold, we must have $t_{q'new} \ge t_{q'}$, which leads to

$$s\left(t_{q'new} - t_{qnew}\right) > N$$

This violates the no capacity waste condition, so it must holds that $t_{qnew} > t_q$, which readily shows that $\rho_{new} < \rho$, $C_{new}^{non}\left(x\right) < C^{non}\left(x\right)$, $C_{new}^{pay}\left(x\right) = C^{pay}\left(x\right)$ and $P_{new}^{pay}\left(x\right) < P^{pay}\left(x\right)$. This completes the proof.

The proof of Lemma 3.1 follows an intuitive line. For a tolling scheme with capacity waste, we can shorten the toll window to the clearing period of toll payers (from the first toll payer's clearing time point to the last toll payer's clearing time point), and reduce the toll level to the corresponding critical level. This new tolling scheme has no capacity waste, shorter toll window length, and lower toll level. Note that this new tolling scheme does not change the amount of toll payers or their arrival pattern. Thus V and $C^{\text{pay}}\left(x\right)$ do not change. In the meanwhile, with the shorter toll window length, the first toll non-payer arrives at the bottleneck later and incurs less schedule early delay cost. Thus $C^{\text{non}}\left(x\right)$ is lower, and the total system cost is lower.

From Lemma 3.1, if a feasible tolling scheme of problem (3.10)-(3.12) has capacity waste, then there exists another feasible tolling scheme which does not have capacity waste and improves the objective function value. Therefore, to solve problem (3.10)-(3.12), we can focus our analyses on scenarios without capacity waste, where the equilibrium conditions are given by (3.1)-(3.7) and the toll level satisfies $\rho \le \rho^{\text{crit}}\left(t^+, t^-\right)$. As such, in our subsequent analyses we will regard $\rho^{\text{crit}}\left(t^+, t^-\right)$ as the highest feasible toll level for any toll window $\left(t^+, t^-\right)$.

In the following we will establish a few important properties of objective function (3.10) without considering constraints (3.11)-(3.12). These properties will then be applied to constraints (3.11)-(3.12) to solve the constrained optimization problem.

We first express $C^{\text{non}}\left(x\right)$ and $C^{\text{pay}}\left(x\right)$ in terms of $\left(t^+, t^-, \rho\right)$. Substituting (3.4) into (3.1), and (3.6) into (3.2), we have

$$C^{\text{non}}\left(x\right) = \beta\left(x\right)\left[\frac{N}{s} - \left(t^- - t^*\right) - \frac{2}{1+\mu}\frac{\rho}{\alpha\left(V\right)}\right] \tag{3.15}$$

$$C^{\text{pay}}\left(x\right) = \beta\left(x\right)\left[\frac{N}{s} - \left(t^- - t^*\right) - \left(\frac{1}{\eta} + \frac{2}{1+\mu}\right)\frac{\rho}{\alpha\left(V\right)}\right] \tag{3.16}$$

Comparing (3.15) and (3.16), we have $C^{\text{non}}(x) - C^{\text{pay}}(x) = \rho\alpha(x)/\alpha(V)$. That is, being a toll payer reduces the xth user's travel cost by $\rho\alpha(x)/\alpha(V)$. Thus, the xth user would choose to pay the toll ρ if $\alpha(x) > \alpha(V)$, which confirms that the Vth user is the indifferent user.

Because $\alpha(V) = \alpha(ls)$ decreases with l, from (3.15) and (3.16) it is easy to see that $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ both decrease with t^-, ρ and l, which leads to the following lemma.

Lemma 3.2. In the ADL model, consider two tolling schemes (t_1^+, t_1^-, ρ_1) and (t_2^+, t_2^-, ρ_2), both without capacity waste, i.e., $\rho_i \le \rho^{\text{crit}}(t_i^+, t_i^-)$, $i = 1, 2$. If $t_1^- \ge t_2^-$, $\rho_1 \ge \rho_2$, and $l_1 \ge l_2$, with at least one strict inequality, then $TC(t_1^+, t_1^-, \rho_1) < TC(t_2^+, t_2^-, \rho_2)$.

Proof. Let $C_i^{\text{non}}(x)$ and $C_i^{\text{pay}}(x)$ be the specifications of $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ under (t_i^+, t_i^-, ρ_i), $i = 1, 2$. It readily holds $C_1^{\text{non}}(x) < C_2^{\text{non}}(x)$ and $C_1^{\text{pay}}(x) < C_2^{\text{pay}}(x)$. Because $C^{\text{pay}}(x) < C^{\text{non}}(x)$ always holds, we also have $C_1^{\text{pay}}(x) < C_1^{\text{non}}(x) < C_2^{\text{non}}(x)$. From (3.10), we have

$$
\begin{aligned}
& TC(t_1^+, t_1^-, \rho_1) - TC(t_2^+, t_2^-, \rho_2) \\
&= \int_{V_1}^{N} \left(C_1^{\text{non}}(x) - C_2^{\text{non}}(x)\right) dx + \int_{V_2}^{V_1} \left(C_1^{\text{pay}}(x) - C_2^{\text{non}}(x)\right) dx + \int_0^{V_2} \left(C_1^{\text{pay}}(x) - C_2^{\text{pay}}(x)\right) dx \\
&< 0
\end{aligned}
$$

This completes the proof.

Lemma 3.2 states that increasing either one of t^-, ρ and l without decreasing the other two will reduce the total system cost. In words, each of the three strategies, postponing the toll ending time, increasing the toll level, and stretching the toll window length, can reduce the total system cost if not conflicting with the other two strategies.

To proceed, we introduce an important benchmark tolling scheme. For a given l, we use $\{l, \rho^{\text{crit}}, \text{balanc}\}$ to denote the tolling scheme satisfying $t^- - t^+ = l$ and

$$\eta(t^* - t^+) = \mu(t^- - t^*) \tag{3.17}$$

$$\rho = \rho^{\text{crit}}(t^+, t^-) \tag{3.18}$$

In words, for a given toll window length l, tolling scheme $\{l, \rho^{\text{crit}}, \text{balanc}\}$ is to position the toll window balanced and charge the corresponding critical toll level. Denote $\rho^{\text{crit}}(l, \text{balanc})$ the toll level under tolling scheme $\{l, \rho^{\text{crit}}, \text{balanc}\}$.

Lemma 3.3. In the ADL model,

(a). for a given l, $\rho = \rho^{\text{crit}}(l, \text{balanc})$ is the highest toll level subject to $\rho \le \rho^{\text{crit}}(t^+, t^-)$ and $t^- - t^+ = l$;

(b). $\rho^{\text{crit}}(l, \text{balanc})$ decreases with l.

Proof. (a). It suffices to prove that, subject to $t^- - t^+ = l$ for a given l, $\rho^{\text{crit}}(t^+, t^-)$ is maximized if (t^+, t^-) is balanced. Because $\rho^{\text{crit}}(t^+, t^-) = \min\{\rho^{\text{crit1}}(t^+, t^-), \rho^{\text{crit2}}(t^+, t^-)\}$ and $\rho^{\text{crit1}}(t^+, t^-)$ is a constant for a given l from (3.8), it suffices to show that $\rho^{\text{crit2}}(t^+, t^-) = \rho^{\text{crit1}}(t^+, t^-)$ if (t^+, t^-) is balanced, which follows readily from (3.8) and (3.9).

(b). From the proof of (a), $\rho^{\text{crit}}(l, \text{balanc}) = \rho^{\text{crit1}}(t^+, t^-)$, which decreases with l from (3.8). This completes the proof.

Lemma 3.4. In the ADL model, for a given l, the minimum $TC(t^+, t^-, \rho)$ is attained by $\{l, \rho^{\text{crit}}, \text{balanc}\}$.

Proof. From Lemma 3.2, the highest toll level $\rho = \rho^{\text{crit}}(t^+, t^-)$ minimizes $TC(t^+, t^-, \rho)$ for any (t^+, t^-). Thus, for a given l, the optimal (t^+, t^-, ρ) must satisfy (3.18). Because t^- alone determines (t^+, t^-, ρ) under conditions $t^- - t^+ = l$ and (3.18), it suffices to prove that $TC(t^+, t^-, \rho)$ decreases with t^- within the range $\eta(t^* - t^+) \ge \mu(t^- - t^*)$, and increases with t^- within the range $\eta(t^* - t^+) \le \mu(t^- - t^*)$.

We first consider the range $\eta(t^* - t^+) \ge \mu(t^- - t^*)$, where $\rho^{\text{crit}}(t^+, t^-) = \rho^{\text{crit1}}(t^+, t^-)$. From (3.8), for a given l, $\rho^{\text{crit1}}(t^+, t^-)$ is a constant. Thus, according to Lemma 3.2, increasing t^- (without changing l and ρ) will decrease $TC(t^+, t^-, \rho)$, i.e., $TC(t^+, t^-, \rho)$ decreases with t^-.

We then consider the range $\eta(t^* - t^+) \le \mu(t^- - t^*)$, where $\rho^{\text{crit}}(t^+, t^-) = \rho^{\text{crit2}}(t^+, t^-)$. Substituting $\rho = \rho^{\text{crit2}}(t^+, t^-)$ given by (3.9) into (3.15) and (3.16), we have

$$C^{\text{non}}(x) = \frac{1+\mu}{1+\mu+2\eta}\beta(x)\left[\frac{N}{s} + \frac{\mu-1}{1+\mu}(t^- - t^*)\right] \tag{3.19}$$

$$C^{\text{pay}}(x) = \mu\alpha(x)(t^- - t^*) \tag{3.20}$$

From (3.19)-(3.20), $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ both increase with t^-. Thus $TC(t^+,t^-,\rho)$ increases with t^-, which completes the proof.

From Lemma 3.4, if l and ρ are unconstrained, the unconstrained optimal tolling scheme must be $\{l,\rho^{\text{crit}},\text{balanc}\}$ for some l, i.e., it must hold $\rho^{\text{uncon}}=\rho^{\text{crit}}(l^{\text{uncon}},\text{balanc})$. This result is consistent with Arnott et al. (1990) and Xiao et al. (2011): under the unconstrained optimal tolling scheme, no queue exists at t^+ or t^-, which can only be achieved by charging the critical toll price and setting the toll window balanced.

Now we introduce the second important tolling scheme. For a given combination (l,ρ) such that $\rho<\rho^{\text{crit}}(l,\text{balanc})$, we use $\{l,\rho,\text{latest}\}$ to denote the tolling scheme satisfying $t^- - t^+ = l$ and

$$(t^+,t^-)=(t^*,t^*+l),\ \text{if}\ \rho<\rho^{\text{crit2}}(t^*,t^*+l) \tag{3.21}$$

$$\rho^{\text{crit2}}(t^+,t^-)=\rho,\ \text{if}\ \rho\geq\rho^{\text{crit2}}(t^*,t^*+l) \tag{3.22}$$

In words, for a given combination of toll window length and toll level (l,ρ), tolling scheme $\{l,\rho,\text{latest}\}$ is to position the toll window as late as possible until either $t^+=t^*$ or $\rho^{\text{crit2}}(t^+,t^-)=\rho$ is reached, whichever comes first. Note that, for a given l, Eq. (3.9) shows that $\rho^{\text{crit2}}(t^+,t^-)$ decreases with t^-, thus $\rho^{\text{crit2}}(t^+,t^-)=\rho$ indeed gives the latest toll window position under given ρ (any later position would make $\rho>\rho^{\text{crit2}}(t^+,t^-)$). If $\rho<\rho^{\text{crit2}}(t^*,t^*+l)$, then $\rho^{\text{crit2}}(t^+,t^-)=\rho$ does not have a solution because we do not allow $t^+>t^*$, in which case (t^*,t^*+l) is the latest toll window position.

From Lemma 3.2, for given l and ρ, the minimum $TC(t^+,t^-,\rho)$ is attained by the maximum t^-, i.e., the latest toll window position, which readily gives the following lemma.

In the ADL model, for a given combination (l,ρ) such that $\rho<\rho^{\text{crit}}(l,\text{balanc})$, the minimum $TC(t^+,t^-,\rho)$ is attained by $\{l,\rho,\text{latest}\}$.

Denote $TC(l,\rho,\text{latest})$ the total system cost under tolling scheme $\{l,\rho,\text{latest}\}$. We have the following important result.

Lemma 3.5. In the ADL model, consider (l_1, ρ_1) and (l_2, ρ_2) satisfying $\rho_i < \rho^{\text{crit}}(l_i, \text{balanc})$, $i =$ 1, 2. If $\rho_1 \geq \rho_2$ and $l_1 \geq l_2$, with at least one strict inequality, then $TC(l_1, \rho_1, \text{latest}) < TC(l_2, \rho_2, \text{latest})$.

Proof. $\{l_i, \rho_i, \text{latest}\}$ have two forms depending on whether $\rho_i \geq \rho^{\text{crit2}}(t^*, t^* + l_i)$ holds, given by (3.21)-(3.22). From (3.9), $\rho^{\text{crit2}}(t^*, t^* + l)$ decreases with l, thus $\rho^{\text{crit2}}(t^*, t^* + l_2) \geq \rho^{\text{crit2}}(t^*, t^* + l_1)$. Then, if $\rho_2 \geq \rho^{\text{crit2}}(t^*, t^* + l_2)$, in view of $\rho_1 \geq \rho_2 \geq \rho^{\text{crit2}}(t^*, t^* + l_2) \geq \rho^{\text{crit2}}(t^*, t^* + l_1)$, it must also hold $\rho_1 \geq \rho^{\text{crit2}}(t^*, t^* + l_1)$. Therefore, there are three cases: (a) $\rho_i < \rho^{\text{crit2}}(t^*, t^* + l_i)$, $i =$ 1, 2; (b) $\rho_i \geq \rho^{\text{crit2}}(t^*, t^* + l_i)$, $i =$ 1, 2; (c) $\rho_1 \geq \rho^{\text{crit2}}(t^*, t^* + l_1)$ and $\rho_2 < \rho^{\text{crit2}}(t^*, t^* + l_2)$.

We first consider $\rho_i < \rho^{\text{crit2}}(t^*, t^* + l_i)$, $i =$ 1, 2, in which case $\{l_i, \rho_i, \text{latest}\}$ reduces to $(t^*, t^* + l_i, \rho_i)$, $i = 1$, 2. From Lemma 3.2, we readily have $TC(t^*, t^* + l_1, \rho_1) < TC(t^*, t^* + l_2, \rho_2)$.

We then consider $\rho_i \geq \rho^{\text{crit2}}(t^*, t^* + l_i)$, $i = 1$, 2, in which case $\{l_i, \rho_i, \text{latest}\}$ is determined by $t_i^- - t_i^+ = l_i$ and $\rho^{\text{crit2}}(t_i^+, t_i^-) = \rho_i$, $i = 1$, 2. From the expression of $\rho^{\text{crit2}}(t^+, t^-)$ given by (3.9), because $\alpha(V) = \alpha(ls)$ decreases with l, it must hold $t_1^- < t_2^-$. Under condition $\rho = \rho^{\text{crit2}}(t^+, t^-)$, $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ are given by (3.19)-(3.20), both increasing with t^-. Thus, we have $C_1^{\text{non}}(x) < C_2^{\text{non}}(x)$ and $C_1^{\text{pay}}(x) < C_2^{\text{pay}}(x)$. Then, following the proof of Lemma 3.2, it can be easily shown that $TC(t_1^+, t_1^-, \rho_1) - TC(t_2^+, t_2^-, \rho_2) < 0$.

We finally consider $\rho_1 \geq \rho^{\text{crit2}}(t^*, t^* + l_1)$ and $\rho_2 < \rho^{\text{crit2}}(t^*, t^* + l_2)$, in which case $\{l_2, \rho_2, \text{latest}\}$ reduces to $(t^*, t^* + l_2, \rho_2)$, and $\{l_1, \rho_1, \text{latest}\}$ is determined by $t_1^- - t_1^+ = l_1$ and $\rho^{\text{crit2}}(t_1^+, t_1^-) = \rho_1$. We consider two cases, $\rho_1 > \rho^{\text{crit2}}(t^*, t^* + l_2)$ and $\rho_1 \leq \rho^{\text{crit2}}(t^*, t^* + l_2)$. If $\rho_1 > \rho^{\text{crit2}}(t^*, t^* + l_2)$, then let $\rho_3 = \rho^{\text{crit2}}(t^*, t^* + l_2)$, and consider the auxiliary tolling scheme $(t^*, t^* + l_2, \rho_3)$. It holds $TC(t^*, t^* + l_2, \rho_3) < TC(t^*, t^* + l_2, \rho_2)$ from Lemma 3.1, and $TC(t_1^+, t_1^-, \rho_1) < TC(t^*, t^* + l_2, \rho_3)$ from the previous case of "$\rho_i \geq \rho^{\text{crit2}}(t^*, t^* + l_i)$, $i = 1, 2$". Thus $TC(t_1^+, t_1^-, \rho_1) < TC(t^*, t^* + l_2, \rho_2)$. Now we consider $\rho_1 \leq \rho^{\text{crit2}}(t^*, t^* + l_2)$. Consider the auxiliary tolling scheme $(t^*, t^* + l_2, \rho_1)$.

From Lemma 3.1, it holds $TC\left(t^*, t^*+l_2, \rho_1\right) \le TC\left(t^*, t^*+l_2, \rho_2\right)$, where "=" can hold only if $\rho_1 = \rho_2$. Let l_4 be such that $\rho^{\text{crit2}}\left(t^*, t^*+l_4\right) = \rho_1$. From the expression of $\rho^{\text{crit2}}\left(t^+, t^-\right)$ given by (3.9), because $\alpha(V) = \alpha(ls)$ decreases with l, it must hold $l_1 \ge l_4 \ge l_2$. Consider the auxiliary tolling scheme $\left(t^*, t^*+l_4, \rho_1\right)$. It holds $TC\left(t^*, t^*+l_4, \rho_1\right) \le TC\left(t^*, t^*+l_2, \rho_1\right)$ from Lemma 3.1, where "=" can hold only if $l_4 = l_2$. It holds $TC\left(t_1^+, t_1^-, \rho_1\right) < TC\left(t^*, t^*+l_4, \rho_1\right)$ from the previous case of "$\rho_i \ge \rho^{\text{crit2}}\left(t^*, t^*+l_i\right)$, $i = 1,\ 2$", unless $l_1 = l_4$ which gives $\left(t_1^+, t_1^-, \rho_1\right) = \left(t^*, t^*+l_4, \rho_1\right)$. Therefore, we have $TC\left(t_1^+, t_1^-, \rho_1\right) \le TC\left(t^*, t^*+l_4, \rho_1\right) \le TC\left(t^*, t^*+l_2, \rho_1\right) \le TC\left(t^*, t^*+l_2, \rho_2\right)$, which gives $TC\left(t_1^+, t_1^-, \rho_1\right) \le TC\left(t^*, t^*+l_2, \rho_2\right)$, where "=" can hold only if $\rho_1 = \rho_2$ and $l_1 = l_2$. This gives $TC\left(t_1^+, t_1^-, \rho_1\right) < TC\left(t^*, t^*+l_2, \rho_2\right)$, and completes the proof.

The proof of Lemma 3.5 relies on (3.19) and (3.20), which show that users' travel costs increase with t^- under $\rho = \rho^{\text{crit2}}\left(t^+, t^-\right)$. That is, among the three strategies suggested by Lemma 3.2 to reduce the total system cost (i.e., increasing t^-, ρ and l), when there is a trade-off between increasing t^- and increasing ρ and l, (3.19) and (3.20) demonstrate that the net effect of reducing t^- while increasing ρ and l improves the total system cost. In short, combining Lemma 3.2 and Lemma 3.5, to reduce the total system cost, we should first increase ρ and l to their upper bounds and then increase t^- to the latest position. This immediately leads to the following major result.

Proposition 3.1. In the ADL model, for the constrained optimization problem (3.10)-(3.12) with $\bar{l} < l^{\text{uncon}}$ and $\bar{\rho} < \rho^{\text{uncon}}$, the optimal solution is $\left\{\bar{l}, \bar{\rho}, \text{latest}\right\}$.

Proof. Because $\rho^{\text{crit}}(l, \text{balanc})$ decreases with l, $\rho < \rho^{\text{crit}}(l, \text{balanc})$ holds for any feasible (l, ρ) in view of $\rho \le \bar{\rho} < \rho^{\text{uncon}} = \rho^{\text{crit}}\left(l^{\text{uncon}}, \text{balanc}\right) < \rho^{\text{crit}}\left(\bar{l}, \text{balanc}\right) \le \rho^{\text{crit}}(l, \text{balanc})$. Then, from Lemma 3.4, $TC(l, \rho, \text{latest})$ is the minimum objective function value for any feasible (l, ρ). Thus it suffices to prove that $(l, \rho) = \left(\bar{l}, \bar{\rho}\right)$ gives the minimum $TC(l, \rho, \text{latest})$, which follows readily from Lemma 3.5. This completes the proof.

Proposition 3.1 states that, in the ADL model, the constrained optimal tolling scheme is to choose the maximum acceptable toll level and toll window length, and start the tolling period as late as possible without causing capacity waste. This result suggests that, because under constrained optimization it is impossible to eliminate the queues at both the starting and the ending moments of the tolling period, the priority is to start the tolling period late to eliminate the queue at the toll ending moment only. This insight is never reported in traditional unconstrained coarse tolling studies. More interpretations of Proposition 3.1 will be provided in the next section when we compare the results of the ADL and the Laih models.

3.4. The Laih model

In this section we study the constrained optimization of bottleneck coarse tolling in the Laih model. The Laih model is characterized by that the mass arrival is eliminated through allowing toll non-payers to wait for the toll to end on secondary lanes without blocking the toll payers. The toll non-payers experience longer queuing delay than the toll payers. In the Laih model, the position of the peak hour is exactly same as that of no-toll equilibrium. The arrival rate of users having schedule early delay remains $s/(1-\eta)$, and the arrival rate of users having schedule late delay remains $s/(1+\mu)$.

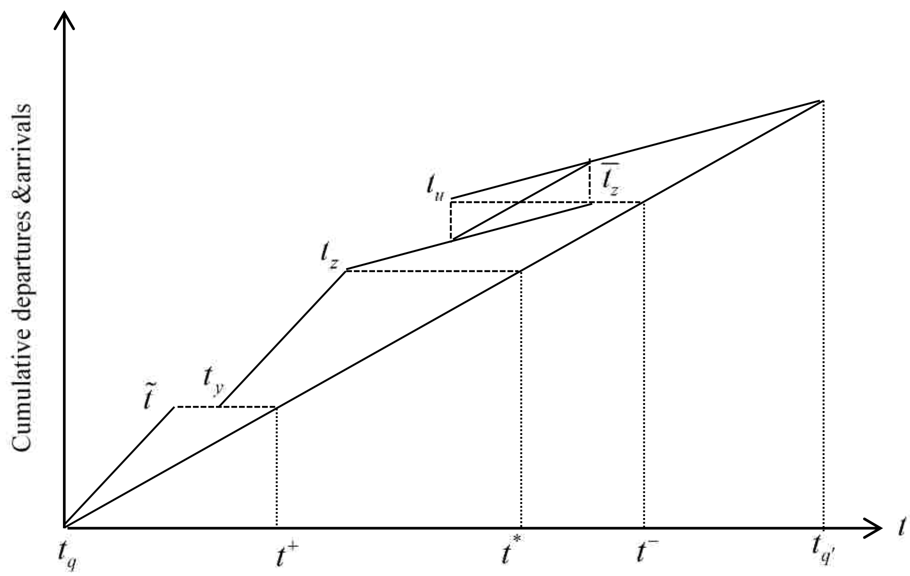

Figure 3.3. Equilibrium profile under coarse tolling: the Laih model

Figure 3.3 shows the equilibrium profile under coarse tolling of Laih model without capacity waste. t_u is the arrival time of the first toll non-payer entering the separated waiting lane. The toll payer that has no schedule delay arrives at t_z. $\overline{t_z}$ is the last toll payer's arrival time. The last toll payer is cleared at t^-. After the toll is cancelled at t^-, the toll non-payers in the separated waiting lane start to leave the bottleneck. The first toll non-payer that uses the separated waiting lane has the same schedule late delay as the last toll payer.

Using techniques similar to the derivation of (3.1)-(3.7) in the ADL model, i.e., making use of $C^{\text{non}}(V) = C^{\text{pay}}(V) + \rho$ for the indifferent user and comparing the travel costs and travel prices of toll payers and non-payers arriving at the bottleneck at different time instants, we can derive all the critical arrival time instants as well as the equilibrium travel costs of users. Specifically, we have

$$C^{\text{non}}\left(x\right)=\alpha\left(x\right)\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{3.23}$$

$$C^{\text{pay}}\left(x\right)=\alpha\left(x\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\frac{\rho}{\alpha\left(V\right)}\right] \tag{3.24}$$

Comparing (3.23) and (3.24), it is easy to see $C^{\text{non}}\left(x\right)-C^{\text{pay}}\left(x\right)=\rho\alpha\left(x\right)/\alpha\left(V\right)$, which, similar to the ADL model, confirms that the V th user is the indifferent user.

Using techniques similar to the derivation of (3.8) and (3.9) in the ADL model, we can derive the critical toll levels in the Laih model. Specifically, by setting $t_y=t^+$, we can obtain the critical toll level

$$\rho^{\text{crit1}}\left(t^+,t^-\right)=\alpha\left(V\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\eta\left(t^*-t^+\right)\right] \tag{3.25}$$

By setting $\bar{t}_z=t^-$, we can obtain the critical toll level

$$\rho^{\text{crit2}}\left(t^+,t^-\right)=\alpha\left(V\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\mu\left(t^--t^*\right)\right] \tag{3.26}$$

Comparing (3.25) and (3.26) here gives exactly the same results as comparing (3.8) and (3.9) in the ADL model: $\rho^{\text{crit1}}<,\ >$ and $=\rho^{\text{crit2}}$, respectively, for $\eta\left(t^*-t^+\right)>,\ <$ and $=\mu\left(t^--t^*\right)$. Thus the interpretations of ρ^{crit1} and ρ^{crit2} are exactly the same as those in the ADL model, we still have Definition 1 on balanced, early and late toll windows, and we still define $\rho^{\text{crit}}\left(t^+,t^-\right)=\min\left\{\rho^{\text{crit1}}\left(t^+,t^-\right),\rho^{\text{crit2}}\left(t^+,t^-\right)\right\}$. Similar to the ADL model, the scenarios with capacity waste deserves no consideration in the Laih model, and we focus our analyses on scenarios without capacity waste, where the toll level satisfies $\rho\le\rho^{\text{crit}}\left(t^+,t^-\right)$.

Note that the constrained optimization problem is still given by (3.10)-(3.12), with $C^{\text{non}}\left(x\right)$ and $C^{\text{pay}}\left(x\right)$ given by (3.23) and (3.24). From (3.23) $C^{\text{non}}\left(x\right)$ is independent of the tolling scheme and is equal to the travel cost of the no-toll equilibrium. This is because the Laih model preserves the peak period t_q and $t_{q'}$ of the no-toll equilibrium. From (3.24) $C^{\text{pay}}\left(x\right)$ decreases with ρ and l. Because $\left(l,\rho\right)$ determines $C^{\text{pay}}\left(x\right)$ and $V=ls$, and $C^{\text{non}}\left(x\right)$ is independent of the tolling scheme, $\left(l,\rho\right)$ determines the total system cost $TC=\int_V^N C^{\text{non}}\left(x\right)dx+\int_0^V C^{\text{pay}}\left(x\right)dx$. That is, in the Laih model TC is determined by $\left(l,\rho\right)$, the toll window length and toll level, rather than the specific position of $\left(t^+,t^-\right)$. Specifically, from (3.23) and (3.24) we can obtain

$$TC(l,\rho)=\frac{\eta\mu}{\eta+\mu}\frac{N}{s}\int_0^N \alpha(x)\,dx-\frac{\rho}{\alpha(V)}\int_0^V \alpha(x)dx$$

Because $C^{\text{non}}(x)$ is independent of the tolling scheme and $C^{\text{pay}}(x)$ decreases with ρ and l, using similar techniques in the proof of Lemma 3.2, we can prove the following lemma.

Lemma 3.6. In the Laih model, consider two tolling schemes $\left(t_1^+,t_1^-,\rho_1\right)$ and $\left(t_2^+,t_2^-,\rho_2\right)$, both without capacity waste, i.e., $\rho_i\le\rho^{\text{crit}}\left(t_i^+,t_i^-\right)$, $i=1,2$. If $l_1\ge l_2$ and $\rho_1\ge\rho_2$, with at least one strict inequality, then $TC(l_1,\rho_1)<TC(l_2,\rho_2)$.

Lemma 3.6 states that $TC(l,\rho)$ strictly decreases with l and ρ, which suggests that the constrained optimization problem simply has $\left(\bar{l},\bar{\rho}\right)$ as the optimal solution. While this is true, we still need to show that there exists a toll window positon $\left(t^+,t^-\right)$ that can realize $\left(\bar{l},\bar{\rho}\right)$ without causing capacity waste. To do so, following the analyses in the ADL model, we still define $\left\{l,\rho^{\text{crit}},\text{balanc}\right\}$ and $\rho^{\text{crit}}(l,\text{balanc})$ using (3.17) and (3.18). We then have the following intermediate results.

Lemma 3.7. In the Laih model,

(a). for a given l, $\rho=\rho^{\text{crit}}(l,\text{balanc})$ is the highest toll level subject to $\rho\le\rho^{\text{crit}}\left(t^+,t^-\right)$ and $t^- - t^+ = l$, and is attainable only when $\eta\left(t^*-t^+\right)=\mu\left(t^- - t^*\right)$;

(b). $\rho^{\text{crit}}(l,\text{balanc})$ decreases with l;

(c). for a given l, the minimum $TC(l,\rho)$ is attained by $\left\{l,\rho^{\text{crit}},\text{balanc}\right\}$;

(d). the unconstrained optimal solution satisfies $\rho^{\text{uncon}}=\rho^{\text{crit}}\left(l^{\text{uncon}},\text{balanc}\right)$.

Proof. It is easy to see that (d) follows readily from (c), and (c) follows readily from (a) and Lemma 3.6. Thus in the following we only prove (a) and (b).

(a) It suffices to prove that, subject to $t^- - t^+ = l$ for a given l, $\rho^{\text{crit}}\left(t^+,t^-\right)$ is maximized if and only if $\left(t^+,t^-\right)$ satisfies $\eta\left(t^*-t^+\right)=\mu\left(t^- - t^*\right)$, which follows readily from (3.25), (3.26) and $\rho^{\text{crit}}\left(t^+,t^-\right)=\min\left\{\rho^{\text{crit1}}\left(t^+,t^-\right),\rho^{\text{crit2}}\left(t^+,t^-\right)\right\}$.

(b). Substituting $t^- - t^+ = l$ and $\eta\left(t^* - t^+\right) = \mu\left(t^- - t^*\right)$ into either (3.25) or (3.26), it is easy to obtain $\rho^{\text{crit}}\left(l,\text{balanc}\right) = \alpha\left(V\right)\dfrac{\eta\mu}{\eta+\mu}\left(\dfrac{N}{s} - l\right)$, which decreases with l. This completes the proof.

Lemma 3.7(a) is the Laih-model counterpart of Lemma 3.3(a). The difference is that, in the ADL model, $\rho = \rho^{\text{crit}}\left(l,\text{balanc}\right)$ can be attained by both early and balanced toll windows because $\rho^{\text{crit1}}\left(t^+, t^-\right)$ given by (3.8) is a constant for a given l. In the Laih model, $\rho = \rho^{\text{crit}}\left(l,\text{balanc}\right)$ can be attained only by the balanced toll window because $\rho^{\text{crit1}}\left(t^+, t^-\right)$ given by (3.25) increases with $\left(t^+, t^-\right)$ and $\rho^{\text{crit2}}\left(t^+, t^-\right)$ given by (3.26) decreases with $\left(t^+, t^-\right)$.

Lemma 3.7(c) is the Laih-model counterpart of Lemma 3.4. While they have exactly the same result in two different models, the proofs are quite different. In the Laih model, Lemma 3.7(c) holds simply because $\rho = \rho^{\text{crit}}\left(l,\text{balanc}\right)$ is the highest toll for a given l and $TC\left(l,\rho\right)$ strictly decreases with ρ. In the ADL model, there is a trade-off between increasing ρ and increasing t^-, and we had to use (3.19) and (3.20) to demonstrate that the net effect of increasing ρ while reducing t^- improves TC.

Because $\rho = \rho^{\text{crit}}\left(l,\text{balanc}\right)$ can be attained only at the balanced position, for $\left(l,\rho\right)$ such that $\rho < \rho^{\text{crit}}\left(l,\text{balanc}\right)$, there is a range of $\left(t^+, t^-\right)$ that can realize $\left(l,\rho\right)$. That is, if the toll level is less than the highest level for a given toll window length, then the toll window position can be moved earlier or later than the balanced position while maintaining the same $\left(l,\rho\right)$ without causing capacity waste. Specifically, for a given combination $\left(l,\rho\right)$ such that $\rho < \rho^{\text{crit}}\left(l,\text{balanc}\right)$, the earliest and latest $\left(t^+, t^-\right)$ positions are given by setting t^- to be, respectively,

$$t^-_{\text{earliest}}\left(l,\rho\right) = t^* + \max\left\{ l - \frac{1}{\eta}\left(\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\rho}{\alpha\left(ls\right)}\right), 0 \right\} \tag{3.27}$$

$$t^-_{\text{latest}}\left(l,\rho\right) = t^* + \min\left\{ \frac{1}{\mu}\left(\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\rho}{\alpha\left(ls\right)}\right), l \right\} \tag{3.28}$$

The earliest position $t^-_{\text{earliest}}\left(l,\rho\right)$ is obtained by moving the toll window as early as possible until either $t^- = t^*$ or $\rho^{\text{crit1}}\left(t^+, t^-\right) = \rho$ is reached, whichever comes first. The latest position $t^-_{\text{latest}}\left(l,\rho\right)$ is obtained by moving the toll window as late as possible until either $t^- = t^* + l$ or $\rho^{\text{crit2}}\left(t^+, t^-\right) = \rho$ is reached, whichever comes first.

Now we are ready to give our major result for the Laih model.

Proposition 3.2. In the Laih model, for the constrained optimization problem (3.10)-(3.12) with $\overline{l} < l^{\text{uncon}}$ and $\overline{\rho} < \rho^{\text{uncon}}$, the optimal solution is

$$\begin{cases} t^- - t^+ = \overline{l} \\ \rho = \overline{\rho} \\ t^- \in \left[t^-_{\text{earliest}}\left(\overline{l}, \overline{\rho}\right), t^-_{\text{latest}}\left(\overline{l}, \overline{\rho}\right) \right] \end{cases}$$

Proof: From Lemma 3.6 and the definition of $t^-_{\text{earliest}}\left(l, \rho\right)$ and $t^-_{\text{latest}}\left(l, \rho\right)$ given by (25)-(26), it suffices to prove $\overline{\rho} < \rho^{\text{crit}}\left(\overline{l}, \text{balanc}\right)$, which holds because $\rho^{\text{crit}}\left(l, \text{balanc}\right)$ decreases with l and thereby $\overline{\rho} < \rho^{\text{uncon}} = \rho^{\text{crit}}\left(l^{\text{uncon}}, \text{balanc}\right) < \rho^{\text{crit}}\left(\overline{l}, \text{balanc}\right)$. This completes the proof.

Proposition 3.2 states that, in the Laih model, the constrained optimal tolling scheme is to choose the maximum acceptable toll level and toll window length, and position the toll window anywhere in the range that would not cause capacity.

Comparing Proposition 3.1 in the ADL model and Proposition 3.2 in the Laih model, the common part is that both models choose the maximum acceptable toll level and toll window length. This can be explained by the fact that, in both models, the toll replaces the queuing delay of toll payers without increasing the cost of the toll non-payers, and thus increasing the toll in both toll price and tolling period will increase the benefit of imposing a toll.

The different part is that, the Laih model has no requirement on the specific toll window position, while the ADL model requires the toll window positioned as late as possible. This can be explained as follows. In the ADL model, the time interval between the mass arrival time and the toll ending time could be viewed as a period during which a mass of users waiting for the toll to end, which represents a pure system cost. Moving the toll window later will make the mass arrival time closer to the toll ending time, i.e., attract the mass of users to arrive at the bottleneck closer to the toll ending time and wait a shorter period for the toll to end. This reduces the mass of users' queuing delay and improves the total system cost, and thus the ADL model requires the latest toll window position. In the Laih model, without mass arrival, changing the toll window position has no impact on users' travel costs and thus no requirement is imposed on the toll window position.

3.5. The braking model

In this section, we study the constrained optimization of bottleneck coarse tolling in the braking model. The braking model, or tactical waiting model, is developed by Lindsey et al. (2012) and Xiao et al. (2012). Different from the ADL and the Laih models, the braking model eliminates the mass arrival phenomenon by assuming that travelers can delay reaching the bottleneck to avoid the toll by slowing down their cars, although there is still capacity at the bottleneck to serve them without delay. Such a braking behavior can cause higher queuing and schedule delay cost for the toll non-payers, but is justified by lower travel price compared with paying the toll.

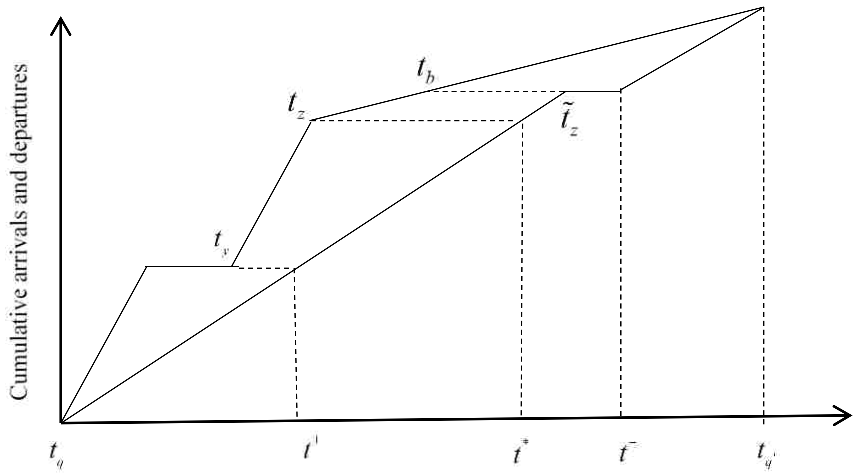

Figure 3.4. Equilibrium profile under coarse tolling: the braking model

Depending on the specific toll window position and toll level, there are many possible equilibrium profiles in the braking model as shown in Xiao et al. (2012). Among these profiles, profile 1 in Xiao et al. (2012) has already been proved to contain the unconstrained optimal tolling scheme. This profile is characterized by that no capacity waste exists at t^+, braking starts before all toll payers are cleared, and the last toll payer arrives at work late. For our constrained optimization, it can also be shown that the optimal solution falls under this profile. Therefore, to keep our analysis focused, and also to be consistent with the ADL and the Laih models (i.e., the profile under study contains the unconstrained optimal solution, and the last toll payer arrives at work late), we only consider profile 1 in Xiao et al. (2012). Figure 3.4 shows this equilibrium profile. In Figure 3.4, t_b denotes the moment users start the braking behavior. It is also the moment when the last toll payer arrives at the bottleneck. $\tilde{t}_z$ denotes the last toll payer's clearing time at the bottleneck. As can be seen from the equilibrium profile, no mass arrival exists in the braking model. After the toll is cancelled at t^-, those braking travelers start to leave the bottleneck. They choose to delay their arrival at the bottleneck to avoid paying the toll. The boundary conditions of such an equilibrium profile are

$$t_y \le t^+ \tag{3.29}$$

$$t_b \le \tilde{t}_z \tag{3.30}$$

$$t^* \le \tilde{t}_z \tag{3.31}$$

Condition (3.29) requires that the first toll payer arrives at the bottleneck no later than the start of the tolling period (no capacity waste exists at t^+). Condition (3.30) states that the first braking traveler's arrival time cannot be later than the last toll payer's clearing time (braking starts before all toll-payers are cleared). Condition (3.31) requires that the last toll payer arrives at work late. Note that in the braking model there is a "wasted" tolling period from $\tilde{t}_z$ to t^-, during which the bottleneck is not utilized while users wait for the toll to end. The "effective" tolling period is from t^+ to $\tilde{t}_z$.

Using techniques similar to the derivation of (3.1)-(3.7) in the ADL model, i.e., making use of $C^{\text{non}}(V) = C^{\text{pay}}(V) + \rho$ for the indifferent user and comparing the travel costs and travel prices of toll payers and non-payers arriving at the bottleneck at different time instants, we can derive all the critical arrival time instants as well as the equilibrium travel costs of users. Specifically, we have

$$\rho = \alpha(V)(1+\mu)(t^- - \tilde{t}_z), \text{ where } V = s(\tilde{t}_z - t^+) \tag{3.32}$$

$$C^{\text{non}}(x) = \alpha(x)\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s} + t^- - \tilde{t}_z\right) \tag{3.33}$$

$$C^{\text{pay}}(x) = \alpha(x)\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \alpha(x)\left(\mu + 1 - \frac{\eta\mu}{\eta+\mu}\right)(t^- - \tilde{t}_z) \tag{3.34}$$

From (3.32)-(3.34), it is easy to see $C^{\text{non}}(x) - C^{\text{pay}}(x) = \rho\alpha(x)/\alpha(V)$, which, similar to the ADL and the Laih models, confirms that the V th user is the indifferent user. It is not surprising that in the braking model, with a "wasted" tolling period $(t^- - \tilde{t}_z)$, the toll non-payer's system cost is made higher than that in the no-toll equilibrium. This is different from the ADL and the Laih models where the toll non-payer's system cost is not higher than that in the no-toll equilibrium.

Observe from (3.33) and (3.34) that, other than the basic bottleneck parameters, $C^{\text{non}}(x)$ and $C^{\text{pay}}(x)$ solely depend on the "wasted" tolling period $(t^- - \tilde{t}_z)$. For ease of exposition, denote $l_w = t^- - \tilde{t}_z$, the "wasted" tolling period, then $l - l_w = \tilde{t}_z - t^+$ is the "effective" tolling period, $V = s(l - l_w)$, and (3.32) can be rewritten as

$$\rho = \alpha(s(l - l_w))(1+\mu)l_w \tag{3.35}$$

Because $\alpha(x)$ is decreasing, the right-hand side of (3.35) is strictly increasing in l_w. Therefore, l_w is unique for a given combination (l,ρ). We use $l_w(l,\rho)$ to highlight l_w is a function of (l,ρ) as determined by (3.35). Because (l,ρ) determines l_w and thereby determines $C^{\text{non}}(x)$, $C^{\text{pay}}(x)$, and V, it also determines the total system cost $TC = \int_V^N C^{\text{non}}(x)dx + \int_0^V C^{\text{pay}}(x)dx$. That is, similar to the Laih model, here in the braking model TC is determined by (l,ρ), the toll window length and toll level, rather than the specific position of (t^+, t^-). Specifically, from (3.33) and (3.34) we can obtain

$$TC(l,\rho) = \frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s} + l_w\right)\int_0^N \alpha(x)dx - (1+\mu)l_w\int_0^V \alpha(x)dx \tag{3.36}$$

3.5.1. General properties of $l_w(l,\rho)$ and $TC(l,\rho)$

In this subsection we will establish a few properties of $l_w(l,\rho)$ and $TC(l,\rho)$, which will be used later for constrained optimization analyses. We first examine the properties of $l_w = l_w(l,\rho)$ based on (3.35). From (3.35) we can obtain

$$\frac{\partial l_w}{\partial l} = \frac{1}{1 - \dfrac{\alpha(V)}{\alpha'(V) s l_w}} \tag{3.37}$$

$$\frac{\partial l_w}{\partial \rho} = \frac{1}{(1+\mu)\left[\alpha(V) - \alpha'(V) s l_w\right]} \tag{3.38}$$

In view of $\alpha'(V) \le 0$, we readily have $\partial l_w / \partial \rho > 0$ and $0 \le \partial l_w / \partial l < 1$ (note $\partial l_w / \partial l = 0$ if and only if $\alpha'(V) = 0$), i.e., l_w increases with ρ and l.

We then derive the first order derivatives of $TC(l,\rho)$ based on (3.36). In view of $V = s(l - l_w)$ and $l_w = l_w(l,\rho)$, from (3.36) we have

$$\frac{\partial TC}{\partial l} = A(V)\frac{\partial l_w}{\partial l} - (1+\mu)\alpha(V) s l_w \tag{3.39}$$

$$\frac{\partial TC}{\partial \rho} = A(V)\frac{\partial l_w}{\partial \rho} \tag{3.40}$$

where

$$A(V) = \frac{\eta\mu}{\eta+\mu}\int_0^N \alpha(x)\,dx + (1+\mu)\alpha(V) s l_w - (1+\mu)\int_0^V \alpha(x)dx \tag{3.41}$$

To proceed, we denote $l^{\text{crit}}(\rho)$ the longest toll window length under given ρ that satisfies the equilibrium profile conditions (3.29)-(3.31). It can be verified that $l^{\text{crit}}(\rho)$ is determined by setting (3.29) and (3.30) to be equality (i.e., by setting $t_y = t^+$ and $t_b = \tilde{t}_z$), and $l^{\text{crit}}(\rho)$ decreases with ρ. Note that $l = l^{\text{crit}}(\rho)$ also means that ρ is the highest toll level under l. In other words, if $l < l^{\text{crit}}(\rho)$, then, within the range specified by conditions (3.29)-(3.31), ρ can be increased under fixed l up to ρ' such that $l = l^{\text{crit}}(\rho')$, and l can be increased under fixed ρ up to $l' = l^{\text{crit}}(\rho)$.

Consider the unconstrained optimal tolling scheme $\left(l^{\text{uncon}}, \rho^{\text{uncon}}\right)$, which minimizes objective function (3.36) subject to (3.29)-(3.31) only (i.e., without constraints $l \le \bar{l}$ and $\rho \le \bar{\rho}$), we have the following lemma.

Lemma 3.8. In the braking model, if $\rho^{\text{uncon}} > 0$, then $l^{\text{uncon}} = l^{\text{crit}}\left(\rho^{\text{uncon}}\right)$.

Proof. If $l^{\text{uncon}} < l^{\text{crit}}\left(\rho^{\text{uncon}}\right)$, then ρ^{uncon} can be increased under l^{uncon}, which means $\partial TC/\partial\rho = 0$ must hold for optimality. From (3.40) and in view of $\partial l_w/\partial\rho > 0$, $\partial TC/\partial\rho = 0$ gives $A(V) = 0$, which in turn gives $\partial TC/\partial l < 0$ from (3.39). $\partial TC/\partial l < 0$ requires $l^{\text{uncon}} = l^{\text{crit}}\left(\rho^{\text{uncon}}\right)$ for optimality, which completes the proof.

Lemma 3.8 states that the unconstrained optimal tolling scheme always gives $t_y = t^+$ and $t_b = \tilde{t}_z$ as long as coarse tolling can reduce the total system cost. Xiao et al. (2012) proved this result for the homogeneous user case without requiring $\rho^{\text{uncon}} > 0$, because $\rho^{\text{uncon}} > 0$ is trivial with homogeneous users. In our analyses with heterogeneous users, $\rho^{\text{uncon}} > 0$ is not trivial as it might depend on the specific form of the VOT distribution $\alpha(x)$. At least it cannot be easily seen from (3.36)-(3.41). However, $\rho^{\text{uncon}} = 0$ is not interesting (and intuitively unlikely), and thus we limit our attention to the case that $\rho^{\text{uncon}} > 0$.

As in the ADL and the Laih models, in our analyses of constrained optimization, we consider constraints $l \le \bar{l}$ and $\rho \le \bar{\rho}$, where $\bar{l} < l^{\text{uncon}}$ and $\bar{\rho} < \rho^{\text{uncon}}$.

Lemma 3.9. In the braking model, consider $l \le \bar{l} < l^{\text{uncon}}$ and $\rho \le \bar{\rho} < \rho^{\text{uncon}}$, then $\bar{l} < l^{\text{crit}}(\rho)$.

Proof. Because $l^{\text{crit}}(\rho)$ decreases with ρ, it holds $l^{\text{crit}}(\rho) > l^{\text{crit}}\left(\rho^{\text{uncon}}\right) = l^{\text{uncon}} > \bar{l}$.

Lemma 3.9 states that constraint $l \le \bar{l}$ is more critical than constraint $l \le l^{\text{crit}}(\rho)$, which means that a tolling scheme satisfying $l \le \bar{l} < l^{\text{uncon}}$ and $\rho \le \bar{\rho} < \rho^{\text{uncon}}$ is always feasible within the range specified by conditions (3.29)-(3.31). Therefore, in our subsequent analyses, we focus on minimizing objective function (3.36) subject to $l \le \bar{l}$ and $\rho \le \bar{\rho}$, ignoring conditions (3.29)-(3.31) and their implied constraint $l \le l^{\text{crit}}(\rho)$. Conditions (3.29)-(3.31) will be used only when we specify the toll window position $\left(t^+, t^-\right)$.

3.5.2. Constrained optimization with homogeneous users

In this subsection we start with the homogenous user case to obtain some insights into constrained optimization. We will discuss extending the insights into the heterogeneous user case in the next subsection. When users are homogeneous, i.e., $\alpha(x) = \alpha, \quad x \in [0, N]$, from (3.35) we have $l_w = \rho/(1+\mu)\alpha$. Then (3.36) is can be specified as

$$TC(l,\rho) = \frac{s}{(1+\mu)\alpha}\rho^2 + \left(\frac{\eta\mu}{(\eta+\mu)(1+\mu)}N - sl\right)\rho + \alpha\frac{\eta\mu}{\eta+\mu}\frac{N^2}{s} \tag{3.42}$$

It is clear from (3.42) that $TC(l,\rho)$ is linear in l (decreases with l) and is quadratic in ρ. Therefore, under constraints $l \le \bar{l}$ and $\rho \le \bar{\rho}$, the optimal l is always $l = \bar{l}$, while the optimal ρ may be an interior solution. We denote $\rho^{\text{glob}}(\bar{l})$ the globally optimal ρ under $l = \bar{l}$, then from (3.42) we have

$$\rho^{\text{glob}}(\bar{l}) = \frac{1}{2}(1+\mu)\alpha\bar{l} - \frac{1}{2}\frac{\eta\mu}{\eta+\mu}\alpha\frac{N}{s} \tag{3.43}$$

It is then very clear that the constrained optimal ρ is $\rho = 0$ if $\rho^{\text{glob}}(\bar{l}) \le 0$ ($\rho \ge 0$ is an implicit constraint throughout this chapter), $\rho = \bar{\rho}$ if $\bar{\rho} \le \rho^{\text{glob}}(\bar{l})$, and $\rho = \rho^{\text{glob}}(\bar{l})$ if $\bar{\rho} > \rho^{\text{glob}}(\bar{l}) > 0$.

We summarize these results into the following proposition.

Proposition 3.3. In the braking model with homogenous users, for the constrained optimization problem (3.10)-(3.12) with $\bar{l} < l^{\text{uncon}}$ and $\bar{\rho} < \rho^{\text{uncon}}$,

(a). if $\bar{l} \le \eta\mu N/(\eta+\mu)(1+\mu)s$, the optimal toll level is $\rho = 0$, i.e., no toll should be charged.

(b). if $\bar{l} > \eta\mu N/(\eta+\mu)(1+\mu)s$ and $\bar{\rho} \le \rho^{\text{glob}}(\bar{l})$, the optimal solution is

$$\begin{cases} t^- - t^+ = \bar{l} \\ \rho = \bar{\rho} \\ t^- - t^* \in \left[\Phi_1(\bar{l},\bar{\rho}), \Phi_2(\bar{l},\bar{\rho})\right] \end{cases}$$

(c). if $\bar{l} > \eta\mu N/(\eta+\mu)(1+\mu)s$ and $\bar{\rho} > \rho^{\text{glob}}(\bar{l})$, the optimal solution is

$$\begin{cases} t^- - t^+ = \bar{l} \\ \rho = \rho^{\text{glob}}(\bar{l}) \\ t^- - t^* \in \left[\varphi_1(\bar{l}), \varphi_2(\bar{l})\right] \end{cases}$$

Similar to the Laih model, because the total cost depends on the toll window length and toll level, rather than the specific position of (t^+, t^-), the constrained optimal tolling scheme has a range of (t^+, t^-) positions specified by $\Phi_1(\bar{l},\bar{\rho})$, $\Phi_2(\bar{l},\bar{\rho})$, $\varphi_1(\bar{l})$ and $\varphi_2(\bar{l})$ in Proposition 3.3(b) and 3.3(c).

Proof. To show the derivation of these earliest and latest toll window positions in Proposition 3.3, a $(t^* - t^+, t^- - t^*)$ two-dimensional space for braking model is first introduced.

The earliest and latest toll window positions are attained when the optimal toll level (either $\bar{\rho}$ or $\rho^{\text{glob}}\left(\bar{l}\right)$) is equal to the critical toll levels at the positions. The critical toll levels are determined by setting the equilibrium profile conditions (3.29), (3.30) and (3.31) to be equality. The criticalness of each of these conditions is determined by the given toll window's length and position. To compare the criticalness of these conditions, a two-dimensional space of $\left(t^* - t^+, t^- - t^*\right)$ is used (shown by Figure 3.5). The lines in Figure 3.5 will be explained as follows.

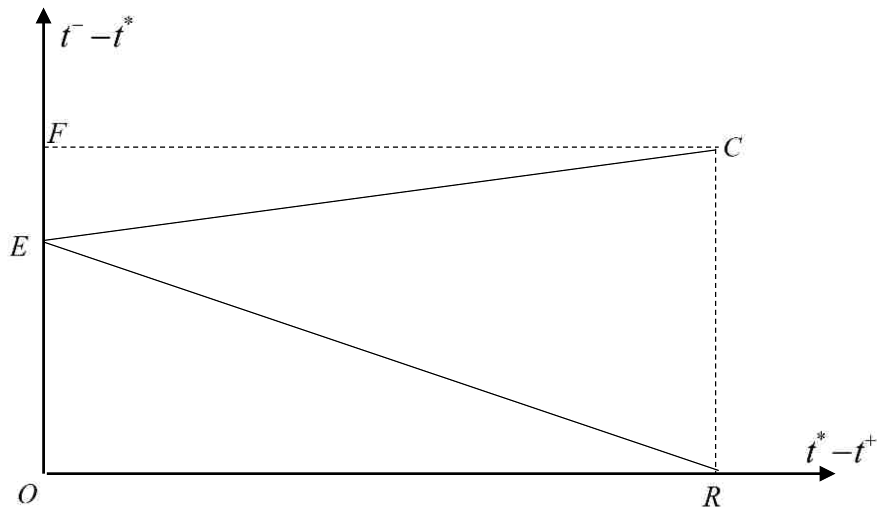

Figure 3.5. Zones with different critical toll levels

Setting (3.29) to be equality gives a lower bound of $\tilde{t}_z$, denoted by

$$\tilde{t}_z^{*y} = t^- - \frac{\eta+\mu}{\eta+\mu+\mu^2}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \eta\left(t^* - t^+\right)\right]$$

At $\tilde{t}_z = \tilde{t}_z^{*y}$, the toll is pushed to the critical level that eliminates queue at t^+ (i.e., $t_y = t^+$). Any higher toll will cause capacity waste at t^+. Such a critical toll is denoted by

$$\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) = \alpha\left(1+\mu\right)\frac{\eta+\mu}{\eta+\mu+\mu^2}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \eta\left(t^* - t^+\right)\right]$$

Setting (3.30) to be equality gives a lower bound of $\tilde{t}_z$, denoted by

$$\tilde{t}_z^{*b} = t^- - \frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right]$$

At $\tilde{t}_z = \tilde{t}_z^{*b}$, the toll is pushed to the critical level that eliminates queue at $\tilde{t}_z$ (i.e., $\tilde{t}_z = t_b$). Such a critical toll is denoted by

$$\rho_{EFC}^{\text{crit}}\left(t^- - t^*\right) = \alpha\left(1+\mu\right)\frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right]$$

Similarly, from (3.31), at $\tilde{t}_z = t^*$, the toll is pushed to the critical level that makes the last toll payer cleared at t^*. Any higher toll will make all toll payers arrive at work early. Such a critical toll is denoted by

$$\rho_{EOR}^{\text{crit}}\left(t^- - t^*\right) = \alpha\left(1+\mu\right)\left(t^- - t^*\right)$$

To compare $\tilde{t}_z^{*y}$ and $\tilde{t}_z^{*b}$, by setting $\tilde{t}_z^{*y} = \tilde{t}_z^{*b}$, it can be obtained that

$$\frac{\mu\left(\eta+\mu+\mu^2\right)}{\eta+\mu-\eta\mu}\left(t^- - t^*\right) - \eta\left(t^* - t^+\right) = \frac{\eta\mu^2}{\eta+\mu-\eta\mu}\frac{N}{s} \tag{3.44}$$

In Figure 3.5, line EC denotes Eq. (3.44). On line EC, it holds $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) = \rho_{EFC}^{\text{crit}}\left(t^- - t^*\right)$. When the toll is pushed to this critical level, no queue exists at t^+ and $\tilde{t}_z$. To compare $\tilde{t}_z^{*y}$ and t^*, by setting $\tilde{t}_z^{*y} = t^*$, it can be obtained that

$$t^- - t^* + \frac{\eta\left(\eta+\mu\right)}{\eta+\mu+\mu^2}\left(t^* - t^+\right) = \frac{\eta\mu}{\eta+\mu+\mu^2}\frac{N}{s} \tag{3.45}$$

In Figure 3.5, line ER denotes Eq. (3.45). On line ER, it holds $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) = \rho_{EOR}^{\text{crit}}\left(t^- - t^*\right)$. When the toll is pushed to this critical level, no queue exists at t^+ and the last toll payer is cleared at t^*.

In Figure 3.5, the coordinate of C is $\left(\mu N/(\eta+\mu)s, \eta N/(\eta+\mu)s\right)$. Coordinate of E is $\left(0, \eta\mu N/\left(\eta+\mu+\mu^2\right)s\right)$. Coordinate of R is $\left(\mu N/(\eta+\mu)s, 0\right)$. In EFC, it holds that $\tilde{t}_z^{*b} > \tilde{t}_z^{*y} > t^*$, so that $\rho_{EFC}^{\text{crit}}\left(t^- - t^*\right) < \rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) < \rho_{EOR}^{\text{crit}}\left(t^- - t^*\right)$. In ECR, it holds that $\tilde{t}_z^{*y} > \tilde{t}_z^{*b}$ and $\tilde{t}_z^{*y} > t^*$, so that $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) < \rho_{EFC}^{\text{crit}}\left(t^- - t^*\right)$ and $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) < \rho_{EOR}^{\text{crit}}\left(t^- - t^*\right)$. In EOR, it holds that $t^* > \tilde{t}_z^{*y} > \tilde{t}_z^{*b}$, so that $\rho_{EOR}^{\text{crit}}\left(t^- - t^*\right) < \rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) < \rho_{EFC}^{\text{crit}}\left(t^- - t^*\right)$. Based on equilibrium profile conditions (3.29), (3.30) and (3.31), in EFC, $\rho_{EFC}^{\text{crit}}\left(t^- - t^*\right)$ is the most critical toll level, so the last toll payer's clearing time satisfies $\tilde{t}_z \geq \tilde{t}_z^{*b}$. In ECR, $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right)$ is the most critical toll level, so the last toll payer's clearing time satisfies that $\tilde{t}_z \geq \tilde{t}_z^{*y}$. In EOR, $\rho_{EOR}^{\text{crit}}\left(t^- - t^*\right)$ is the most critical toll level, so the last toll payer's clearing time satisfies that $\tilde{t}_z \geq t^*$.

By comparing $\rho_{EFC}^{\text{crit}}\left(t^- - t^*\right)$, $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right)$, $\rho_{EOR}^{\text{crit}}\left(t^- - t^*\right)$, $\rho^{\text{glob}}\left(\bar{l}\right)$ and $\bar{\rho}$ at different $\bar{l}$, the earliest and latest positions of the optimal toll window are obtained as follows.

Earliest and latest toll window positions in Proposition 3.3(b)

In Proposition 3.3(b), the optimal toll window length and toll level is $\left(\bar{l}, \bar{\rho}\right)$ and the earliest toll window position is

$$t^{-}-t^{*}=\Phi_{1}\left(\bar{l},\bar{\rho}\right)=\begin{cases}\bar{l}-\dfrac{\mu}{\eta+\mu}\dfrac{N}{s}+\dfrac{\bar{\rho}\left(\eta+\mu+\mu^{2}\right)}{\alpha\eta\left(\eta+\mu\right)\left(\mu+1\right)},\ \text{if } \bar{l}\geq f\left(\bar{\rho}\right)\\ \dfrac{\bar{\rho}}{\alpha\left(1+\mu\right)},\ \text{if } \bar{l}<f\left(\bar{\rho}\right)\end{cases}$$

where

$$f\left(\bar{\rho}\right)=\frac{\mu}{\eta+\mu}\frac{N}{s}-\frac{\eta+\mu+\mu^{2}-\eta^{2}-\eta\mu}{\eta\left(\eta+\mu\right)}\frac{\bar{\rho}}{\alpha\left(1+\mu\right)}$$

The latest toll window position is

$$t^{-}-t^{*}=\Phi_{2}\left(\bar{l},\bar{\rho}\right)=\begin{cases}g\left(\bar{\rho}\right),\ \text{if } \bar{l}\geq g\left(\bar{\rho}\right)\\ \bar{l},\ \text{if } \bar{l}<g\left(\bar{\rho}\right)\end{cases}$$

where

$$g\left(\bar{\rho}\right)=\frac{\eta}{\eta+\mu}\frac{N}{s}-\frac{\eta+\mu-\eta\mu}{\mu\left(\eta+\mu\right)}\frac{\bar{\rho}}{\alpha\left(1+\mu\right)}$$

$f\left(\bar{\rho}\right)$ and $g\left(\bar{\rho}\right)$ denotes two critical toll window lengths. If $\bar{l}\geq f\left(\bar{\rho}\right)$, the toll window can be moved to the earliest position that eliminates the queue at t^{+} (i.e., $t_{y}=t^{+}$). If $\bar{l}<f\left(\bar{\rho}\right)$, the toll window can be moved to the earliest position that makes all toll payers arrive at work early (i.e., $\tilde{t}_{z}=t^{*}$). If $\bar{l}\geq g\left(\bar{\rho}\right)$, the toll window can be moved to the latest position that eliminates the queue at $\tilde{t}_{z}$ (i.e., $\tilde{t}_{z}=t_{b}$). If $\bar{l}<g\left(\bar{\rho}\right)$, the toll window can be moved to the latest position that makes all toll payers arrive at work late (i.e., $t^{+}=t^{*}$). At $t^{-}-t^{*}=\Phi_{1}\left(\bar{l},\bar{\rho}\right)$, it holds either $t_{y}=t^{+}$ or $\tilde{t}_{z}=t^{*}$ (i.e. no queue exists at t^{+} or the last toll payer is cleared at t^{*}). Any earlier position will cause capacity waste at t^{+} or make the last toll payer cleared before t^{*}. At $t^{-}-t^{*}=\Phi_{2}\left(\bar{l},\bar{\rho}\right)$, it holds either $\tilde{t}_{z}=t_{b}$ or $t^{-}-t^{*}=\bar{l}$ (i.e. no queue exists at $\tilde{t}_{z}$ or $t^{+}=t^{*}$). Any later position will cause capacity waste at $\tilde{t}_{z}$ or make $t^{+}>t^{*}$.

Mathematically, with $t^{-}-t^{+}=\bar{l}$, $\Phi_{1}\left(\bar{l},\bar{\rho}\right)$ is obtained by setting $\rho_{ECR}^{\text{crit}}\left(t^{*}-t^{+}\right)=\bar{\rho}$ for $\bar{l}\geq f\left(\bar{\rho}\right)$ and $\rho_{EOR}^{\text{crit}}\left(t^{-}-t^{*}\right)=\bar{\rho}$ for $\bar{l}<f\left(\bar{\rho}\right)$, and $\Phi_{2}\left(\bar{l},\bar{\rho}\right)$ is obtained by setting $\rho_{EFC}^{\text{crit}}\left(t^{-}-t^{*}\right)=\bar{\rho}$ for $\bar{l}\geq g\left(\bar{\rho}\right)$ and $t^{*}-t^{+}=0$ for $\bar{l}<g\left(\bar{\rho}\right)$.

Earliest and latest toll window positions in Proposition 3.3(c)

In Proposition 3.3(c), the optimal toll window length and toll level is $\left(\bar{l},\rho^{\text{glob}}\left(\bar{l}\right)\right)$ and the earliest toll window position is

$$t^- - t^* = \varphi_1\left(\bar{l}\right) = \begin{cases} \bar{l} - \dfrac{\mu\left(\eta+\mu+\mu^2\right)}{2(\eta+\mu)^2(\mu+1)}\dfrac{N}{s} + \dfrac{\eta+\mu+\mu^2}{2\eta(\eta+\mu)}\left(\bar{l} - \dfrac{2\eta\mu}{\eta+\mu+\mu^2}\dfrac{N}{s}\right), \text{ if } \bar{l} \geq l_1 \\ \bar{l} - \dfrac{1}{2}\left(\bar{l} + \dfrac{\eta\mu}{(\eta+\mu)(\mu+1)}\dfrac{N}{s}\right), \text{ if } \bar{l} < l_1 \end{cases}$$

where

$$l_1 = \frac{\eta\mu\left(2\eta+2\mu+\eta\mu+2\mu^2\right)+(\eta+\mu)\left(\eta\mu^2+\eta\mu-\eta^2\mu\right)}{(\eta+\mu)(\mu+1)\left(\eta^2+\eta\mu+\eta+\mu+\mu^2\right)}\frac{N}{s}$$

The latest toll window position is

$$t^- - t^* = \varphi_2\left(\bar{l}\right) = \begin{cases} \dfrac{\eta+\mu-\eta\mu}{2(\eta+\mu)\mu}\left[\dfrac{\eta\mu}{(\eta+\mu)(\mu+1)}\dfrac{N}{s} + \dfrac{2\eta\mu}{(\eta+\mu-\eta\mu)}\dfrac{N}{s} - \bar{l}\right], \text{ if } \bar{l} \geq l_2 \\ \bar{l}, \text{ if } \bar{l} < l_2 \end{cases}$$

where

$$l_2 = \left[\frac{\eta+\mu-\eta\mu}{(\eta+\mu)(\mu+1)} + 2\right]\frac{\eta\mu}{\eta\mu+2\mu^2+\eta+\mu}\frac{N}{s}$$

l_1 and l_2 denotes two critical toll window lengths. If $\bar{l} \geq l_1$, the toll window can be moved to the earliest position that that eliminates the queue at t^+ (i.e., $t_y = t^+$). If $\bar{l} < l_1$, the toll window can be moved to the earliest position that makes all toll payers arrive at work early (i.e., $\tilde{t}_z = t^*$). If $\bar{l} \geq l_2$, the toll window can be moved to the latest position that eliminates the queue at $\tilde{t}_z$ (i.e., $\tilde{t}_z = t_b$). If $\bar{l} < l_2$, the toll window can be moved to the latest position that makes all toll payers arrive at work late (i.e., $t^+ = t^*$). At $t^- - t^* = \varphi_1\left(\bar{l}\right)$, it holds either $t_y = t^+$ or $\tilde{t}_z = t^*$ (i.e. no queue exists at t^+ or the last toll payer is cleared at t^*). Any earlier position will cause capacity waste at t^+ or the last toll payer will be cleared before t^*. At $t^- - t^* = \varphi_2\left(\bar{l}\right)$, it holds either $\tilde{t}_z = t_b$ or $t^- - t^* = \bar{l}$ (i.e. no queue exists at $\tilde{t}_z$ or $t^+ = t^*$). Any later position will cause capacity waste at $\tilde{t}_z$ or make $t^+ > t^*$.

Mathematically, with $t^- - t^+ = \bar{l}$, $\varphi_1\left(\bar{l}\right)$ is obtained by setting $\rho_{ECR}^{\text{crit}}\left(t^* - t^+\right) = \rho^{\text{glob}}\left(\bar{l}\right)$ for $\bar{l} \geq l_1$ and $\rho_{EOR}^{\text{crit}}\left(t^- - t^*\right) = \rho^{\text{glob}}\left(\bar{l}\right)$ for $\bar{l} < l_1$, and $\varphi_2\left(\bar{l}\right)$ is obtained by setting $\rho_{EFC}^{\text{crit}}\left(t^- - t^*\right) = \rho^{\text{glob}}\left(\bar{l}\right)$ for $\bar{l} \geq l_2$ and $t^* - t^+ = 0$ for $\bar{l} < l_2$. This completes the proof.

In Proposition 3.3(a), $\bar{l} \leq \eta\mu N/(\eta+\mu)(1+\mu)s$ gives $\rho^{\text{glob}}\left(\bar{l}\right) \leq 0$ and thus the optimal toll is $\rho = 0$. Proposition 3.3 (a) states that, if the toll window length constraint is too stringent (toll window not allowed to be longer than a threshold), then any toll price will make the total system cost greater than the no-toll equilibrium, and thus no toll should be charged. In Proposition 3.3(b) and 3.3(c), when the toll window is allowed to be longer than the threshold, the optimal toll price chooses the smaller one between the maximum acceptable toll price and the globally optimal toll price.

These results regarding the toll price can all be explained by the fact that, in the braking model, with a "wasted" tolling period, tolling reduces the cost of toll payers but increases the cost of toll non-payers, and thus there is always a trade-off between the two groups. When the toll window is too short as in Proposition 3.3(a), there are too few toll payers and thus it is not worthwhile to impose any toll. When the toll window is allowed to be long enough as in Proposition 3.3(b) and 3.3(c), it is worthwhile to impose a toll because the effect of reducing the costs of toll payers outweighs the effect of increasing the costs of toll non-payers. In this case, the trade-off between toll payers and non-payers determines a best toll price, which, depending on the toll window length, may be higher or lower than the maximum acceptable toll price. When the maximum acceptable toll price is not greater than the best toll price as in Proposition 3.3(b), we should set the toll price as high as possible. When the maximum acceptable toll price exceeds the best toll price as in Proposition 3.3(c), we should simply choose the best toll price.

In Proposition 3.3(b) and 3.3(c), the optimal toll window length is set as the maximum acceptable toll window length. This can be explained by the fact that, with homogeneous user, the "wasted" tolling period depends on the toll price only, i.e., the toll window length has no impact on the "wasted" tolling period and thus does not affect the costs of toll payers and non-payers. In the meanwhile, increasing the toll window length increases the number of toll payers, and changing a user from a toll non-payer to a toll payer always reduces her travel cost. Therefore, increasing the toll window length always improves the total system cost.

3.5.3. Constrained optimization with heterogeneous users

In this subsection we discuss which parts of Proposition 3.3 and the associated insights can be generalized into the heterogeneous user case. First of all, because the total cost depends on the toll window length and toll level, rather than the specific position of $\left(t^{+},t^{-}\right)$, the constrained optimal tolling scheme has a range of $\left(t^{+},t^{-}\right)$ positions as in the Laih model. This property still holds with heterogeneous users. Regarding toll price and toll window length, we have the following result.

Proposition 3.4. In the braking model, for the constrained optimization problem (3.10)-(3.12) with $\bar{l} < l^{\text{uncon}}$ and $\bar{\rho} < \rho^{\text{uncon}}$,

(a). if $\bar{l}$ is sufficiently small, the optimal toll level is $\rho = 0$, i.e., no toll should be charged.

(b). if the optimal toll level is $\rho > 0$, then the optimal toll window length is $l = \bar{l}$.

Proof.

(a). From (3.40) and in view of $\partial l_w/\partial \rho > 0$, it suffices to prove $A(V) > 0$ for a sufficiently small l. From $V = s(l - l_w)$ and $\partial l_w/\partial l < 1$, a sufficiently small l gives a sufficiently small V, which in turn makes the third term of the RHS of (3.41) sufficiently small and thereby $A(V) > 0$. This completes the proof.

(b). If the optimal toll level is $\rho > 0$, from (3.40) and in view of $\partial l_w/\partial \rho > 0$, it must hold $A(V) \le 0$. Then from (3.39) and in view of $\partial l_w/\partial l \ge 0$, we have $\partial TC/\partial l < 0$, which requires the optimal toll window length $l = \bar{l}$ under constraints $l \le \bar{l}$. This completes the proof.

Proposition 3.4 suggests that two important insights from the homogenous user case carries on to the heterogeneous user case. Specifically, Proposition 3.4(a) states that, when the toll window length constraint is too stringent, then no toll should be charged, which generalizes Proposition 3.3(a) into the heterogeneous case. Proposition 3.4(b) states that, if coarse tolling can reduce the total system cost, then the toll window length should be set as the maximum acceptable toll window length, which generalizes the results on toll window length in Proposition 3.3(b) and 3.3(c) into the heterogeneous case.

While the interpretation of Proposition 3.4 is generally the same as that of Proposition 3.3 (i.e., all the results originate from the trade-off between toll payers and non-payers), the reasoning of Proposition 3.4(b) is actually much more complicated than that of the homogeneous case. Specifically, when users are heterogeneous, the "wasted" tolling period increases with both the toll price and the toll window length, which means that the toll window length plays a similar role as the toll price in the trade-off between toll payers and non-payers. In other words, unlike in the homogeneous case, the total system cost is generally nonlinear in toll window length (the specific form depends on the VOT distribution) and may increase or decrease with toll window length. Therefore, the optimal toll window length might have a similar structure as the optimal toll price in Proposition 3.3(b) and 3.3(c), i.e., an interior optimal solution is possible. However, by examining the properties of $l_w(l, \rho)$ and $TC(l, \rho)$ given by (3.37)-(3.41), we find that, when toll price is charged at the optimal level, the net effect of increasing the toll window length is positive in reducing the total system cost, and thus we have Proposition 3.4(b).

3.6. Conclusions

Due to public acceptance reason, for peak hour congestion pricing, it may be politically unacceptable to charge a toll price that is too high or charge a toll for a period that is too long. Motivated by this, we study bottleneck coarse tolling in a constrained optimization setup, where there is a maximum acceptable toll level and a maximum acceptable toll window length. Three widely used coarse tolling models are studied, the ADL, Laih and braking models. The basic user behavioral difference between these three models are: in the ADL model, toll non-payers form a mass arrival at the bottleneck following the last toll payer's arrival at the bottleneck; in the Laih model, a separated waiting facility is built aside of the bottleneck for toll non-payers to wait until the toll ends; in the braking model, toll non-payers can choose to defer their arrival at the bottleneck to avoid paying the toll. In all three models, we consider proportional user heterogeneity, and focus on the case that the unconstrained optimal toll level and toll window length exceed the maximum acceptable upper bounds.

We find that, in the ADL and the Laih models, the constrained optimal coarse tolling chooses the maximum acceptable toll level and toll window length, which is consistent with the traditional insight that, because the toll replaces the queuing delay and thereby reduces the total system cost, increasing the toll in both toll price and tolling period will improve the system efficiency. While this is not surprising, an important new insight regarding the ADL model is established: under constrained optimization, because it is impossible to eliminate the queues at both the starting and the ending moments of the tolling period, in the ADL model the priority is to start the tolling period as late as possible to eliminate the queue at the toll ending moment only. This insight is never reported in traditional unconstrained coarse tolling studies. We find that, if the toll window length constraint is too stringent (the upper bound is too small), then any toll price will make the total system cost greater than the no-toll equilibrium, and thus no toll should be charged. When the toll window length constraint is not too stringent, the optimal solution chooses the maximum acceptable toll window length, while the optimal toll price may be an interior solution (i.e., less than the maximum acceptable level).

Comparing the three models, one common result is that the constrained optimal tolling scheme in all three models chooses the maximum acceptable toll window length. It should be noted that this result is not trivial for the braking model with heterogeneous users, where the trade-off between toll payers and non-payers makes the total system cost nonlinear and non-monotonic in toll window length. This result holds in the braking model with heterogeneous users because we proved that, when toll price is charged at the optimal level, the net effect of increasing the toll window length is positive in improving the total system cost. The Laih model and the braking model have a common feature that the total system cost depends on the toll level and toll window length rather than the specific toll window position. Therefore, in these two models the constrained optimal tolling scheme has a range of toll window positions. By contrast, the ADL model requires the toll window to be positioned as late as possible to minimize the queuing delay of the mass arrival users. A unique feature of the braking model is that the constrained optimal toll price may be less than the maximum acceptable level. This is because, unlike the ADL and Laih models, in the braking model there is a trade-off between toll payers and non-payers, which, depending on the toll window length, may give an interior optimal toll price.

In summary of all three models, in designing bottleneck coarse tolling, when the unconstrained optimal solution exceeds the toll level and toll window length upper bounds, it is generally safe to push the toll window length to its upper bound. When the mass arrival behavior has to be considered, the specific position of the toll window matters, i.e., the later the better. When the braking behavior has to be considered, it may not be optimal to charge the maximum acceptable toll level.

CHAPTER 4

BOTTLENECK COARSE TOLLING UNDER THE EXISTENCE OF OVERTAKING BEHAVIOR

In the previous chapter, we studied the constrained optimization for bottleneck coarse tolling. Three bottleneck models are studied: the ADL model, the Laih model and the braking model. In this chapter, we develop a new coarse tolling model that considers commuters' overtaking behavior: toll payers can overtake those braking commuters (toll non-payers) to take advantage of the tolling period to pay toll to pass the bottleneck. The overtaking model systematically combines Laih model and braking model's together by capturing both model's properties and provides more realistic insights on how to achieve system cost minimization for the coarse tolling problem during morning peak hour.

4.1. Introduction

Bottleneck coarse tolling has been studied for decades. Three bottleneck models are mainly used to conduct the study: the ADL model, the Laih model and the braking model. The ADL model is depicted by that a mass of toll non-payers arrive at the bottleneck right after the last toll payer's arrival, so that they can avoid paying the toll. Such behavior causes a long queue at the ending time of the tolling period. Since these commuters all want to seize the opportunity to avoid the toll, the position of each commuter in the queue is random. The Laih model is characterized by that a separated waiting lane or facility is built aside of the bottleneck for toll non-payers to wait until toll is cancelled. The braking model is featured by that the toll non-payer can choose to defer their arrival at the bottleneck by braking her car to avoid paying the toll. The invention of Laih model and braking model is to eliminate or replace the mass arrival behavior at the bottleneck. Although all three models can reflect some important properties of morning commute behavior respectively, no model has been invented to capture all the properties reflected by these three models. The reason is that the rule that each model is based on is contradicting with each other. For the ADL model, the rule is first in first out: if a commuter arrives at the bottleneck within the tolling period, she must pay the toll. For the Laih model, due to the existence of a separated waiting lane, the first in first out rule is violated: even a commuter arrives within the tolling period, she does not need to pay the toll if she choose to wait in the separated waiting lane and the rule stipulates that, after the toll is cancelled, the commuters waiting in the separated waiting lane has the road right to leave the bottleneck first. The rule of the braking model is that even a commuter arrives within the tolling period, she does not have to pay the toll if she chooses to wait at the bottleneck (not to leave the bottleneck). Since the rule still follows first in and first out, the braking commuters can block the road for those willing to pay the toll. As the result, all toll payers have to arrive at the bottleneck before the first braking commuter's arrival. The rule difference of three models are summarized in the following flow chart.

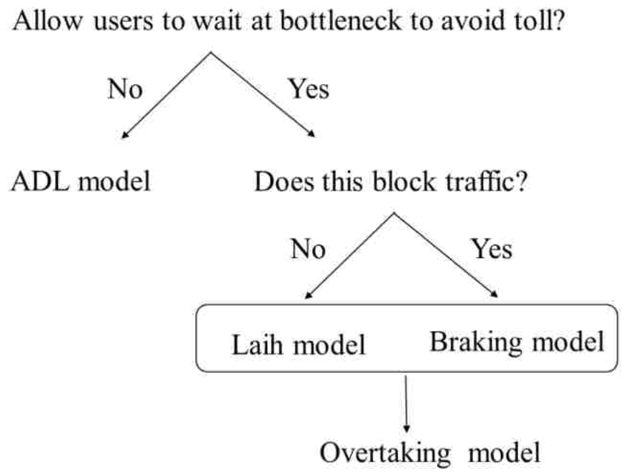

Chart 4.1. Rule difference of three models

Motivated by the rule difference of three models, we introduce a new "overtaking model" that can capture both the property of Laih model and braking model. In reality, toll payers can overtake those braking commuters (toll non-payers) to take advantage of the tolling period to pay toll to pass the bottleneck in order to reduce travel price. Such overtaking behavior can easily be observed in the morning commute period. The overtaking behavior is incurred a constant unit cost of k . k is the cost to overtake one vehicle or the impedance caused by one vehicle. If k is sufficiently large, the proposed overtaking model will become the braking model (i.e. no commuter has incentive to overtake). If $k=0$, the overtaking model can be treated as Laih model. As soon as the braking behavior starts, the overtaking behavior happens. In this research, the toll payer arriving right before the first braking commuter starts to brake is assumed to have schedule late delay cost, so that all overtaking commuters have schedule late delay cost.

4.2. The overtaking model

In this section, the equilibrium profile of overtaking model is analyzed. Figure 4.1 shows the equilibrium profile of the overtaking model, where there is no capacity waste at t^+ and t^-, and all overtaking commuters have schedule late delay. For model completeness, the individual travel price is given in the following

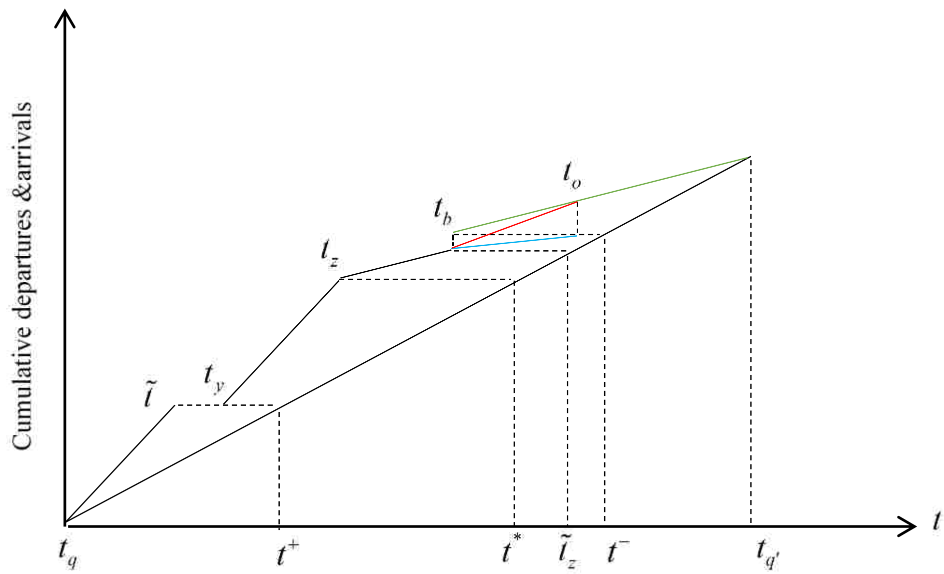

Figure 4.1. Equilibrium profile of overtaking model

For toll non-payers arriving before t^+, let $\lambda_1(\omega)$ denote their arrival rate, and their travel price is given as

$$P^{\text{non}}(t)=\alpha\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}+\beta\left(t^*-t-\frac{\int_{t_q}^{t}\lambda_1(\omega)d\omega-s(t-t_q)}{s}\right) \tag{4.1}$$

It can be obtained that

$$\frac{dP^{\text{non}}(t)}{dt}=\alpha\frac{\lambda_1(t)-s}{s}+\beta\left(-1-\frac{\lambda_1(t)-s}{s}\right)$$

Based on the definition of equilibrium, by setting $dP^{\text{non}}(t)/dt=0$, it gives $\lambda_1(t)=s/(1-\eta)$. Substituting $\lambda_1(t)$ into (4.1), the travel price or cost of a toll non-payer is given as

$$P^{\text{non}}(t)=C^{\text{non}}(t)=\beta(t^*-t_q) \tag{4.2}$$

For toll payers arriving before t_z (the arrival time when toll payer has no schedule delay cost), let $\lambda_2(\omega)$ denote their arrival rate, they have schedule early delay, and their travel price is given as

$$P^{\text{pay}}(t)=\alpha\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_q)}{s}+\beta\left(t^*-t-\frac{\int_{t_q}^{\tilde{t}}\lambda_1(\omega)d\omega+\int_{t_y}^{t}\lambda_2(\omega)d\omega-s(t-t_q)}{s}\right)+\rho$$

where $\tilde{t}$ is the last toll non-payer's arrival time before t^+. It can be obtained that

$$\frac{dP^{\text{pay}}(t)}{dt}=(\alpha-\beta)\frac{\lambda_2(t)}{s}-\alpha$$

By setting $dP^{\text{pay}}(t)/dt = 0$, it gives $\lambda_2(t) = s/(1-\eta)$. With the fact of $\int_{t_q}^{\tilde{t}} \lambda_1(\omega)\,d\omega = s(t^+ - t_q)$, it can be obtained that

$$P^{\text{pay}}(t) = \alpha(t^+ - t_y) + \beta(t^* - t^+) + \rho \tag{4.3}$$

For toll payers arriving after t_z and before any braking or overtaking happens, let $\lambda_3(\omega)$ denote their arrival rate, they have schedule late delay, and the travel price is given as

$$P^{\text{pay}}(t) = \alpha \frac{\int_{t_q}^{\tilde{t}} \lambda_1(\omega)\,d\omega + \int_{t_y}^{t_z} \lambda_2(\omega)\,d\omega + \int_{t_z}^{t} \lambda_3(\omega)\,d\omega - s(t-t_q)}{s} +$$
$$\gamma\left(t + \frac{\int_{t_q}^{\tilde{t}} \lambda_1(\omega)\,d\omega + \int_{t_y}^{t_z} \lambda_2(\omega)\,d\omega + \int_{t_z}^{t} \lambda_3(\omega)\,d\omega - s(t-t_q)}{s} - t^*\right) + \rho$$

It can be obtained that

$$\frac{dP^{\text{pay}}(t)}{dt} = (\alpha+\gamma)\frac{\lambda_3(t) - s}{s} + \gamma$$

By setting $dP^{\text{pay}}(t)/dt = 0$, it gives $\lambda_3(t) = s/(1+\mu)$. With the fact of $\int_{t_y}^{t_z} \lambda_2(\omega)\,d\omega = s(t^* - t^+)$, it can be obtained that

$$P^{\text{pay}}(t) = \alpha(t^* - t_z) + \rho \tag{4.4}$$

The first braking commuter is assumed to start braking at t_b. The toll payer that arrives right before t_b is assumed to leave the bottleneck at $\tilde{t}_z$. If she has schedule late delay (i.e. $\tilde{t}_z \geq t^*$), at equilibrium, it holds that

$$\frac{s}{1+\mu}(t_b - t_z) = s(\tilde{t}_z - t^*) \tag{4.5}$$

(4.4) and (4.5) gives

$$P^{\text{pay}}(t) = \alpha(\tilde{t}_z - t_b) + \gamma(\tilde{t}_z - t^*) + \rho \tag{4.6}$$

Let $\lambda_b(\omega)$ denote the arrival rate of the braking commuters. For a braking commuter arriving before t^-, her travel price is given as

$$P^{\text{non}}(t) = \alpha\left(\frac{\int_{t_b}^{t} \lambda_b(\omega)\,d\omega}{s} + t^- - t\right) + \gamma\left(t^- + \frac{\int_{t_b}^{t} \lambda_b(\omega)\,d\omega}{s} - t^*\right) \tag{4.7}$$

It can be obtained that

$$\frac{dP^{\text{non}}(t)}{dt} = (\alpha+\gamma)\frac{\lambda_b(t)}{s} - \alpha$$

By setting $dP^{\text{non}}(t)/dt = 0$, it can be obtained that $\lambda_b(t) = s/(1+\mu)$. From (4.7), it gives that

$$P^{\text{non}}(t) = \alpha(t^- - t_b) + \gamma(t^- - t^*) \tag{4.8}$$

For braking commuters arriving after t^{-}, their arrival rate can also be obtained as $s/(1+\mu)$. Their travel price is same as equation (4.8).

In the braking model, a braking commuter blocks those who arrive after her, so toll payers should arrive before t_b. While in the overtaking model, toll payers can overtake those braking commuters, so after t_b, there are still toll payers arriving at the bottleneck. The overtaking cost is assumed to be proportionate to the number of braking commuters by the time the toll payer arrives at the bottleneck. The overtaking cost is given as $k\int_{t_b}^{t}\lambda_b(\omega)d\omega$, which captures the property that if the toll payers decides to arrive later, she will face a longer queue ahead of her or her travel will be impeded by more braking commuters, so that her overtaking cost will be higher. In reality, k does not have to be the cost of overtaking one vehicle. It can also be the risk cost incurred by impedance or blocking of the braking commuters. Let $\lambda_o(\omega)$ denote the arrival rate of the overtaking commuters. If arriving before $\tilde{t}_z$, an overtaking commuter's travel price of is given by

$$P^{\text{pay}}(t)=\alpha\left(\frac{\int_{t_b}^{t}\lambda_o(\omega)d\omega}{s}+\tilde{t}_z-t\right)+\gamma\left(\tilde{t}_z+\frac{\int_{t_b}^{t}\lambda_o(\omega)d\omega}{s}-t^{*}\right)+k\int_{t_b}^{t}\lambda_b(\omega)d\omega+\rho \tag{4.9}$$

It can be obtained that

$$\frac{dP^{\text{pay}}(t)}{dt}=(\alpha+\gamma)\frac{\lambda_o(t)}{s}-\alpha+k\lambda_b(t)$$

By setting $dP^{\text{pay}}(t)/dt=0$, it gives

$$\lambda_o(t)=\frac{s}{1+\mu}\left(1-\frac{sk}{\alpha(1+\mu)}\right) \tag{4.10}$$

The condition that overtaking behavior exists is

$$1-\frac{sk}{\alpha(1+\mu)}>0,$$

which is equivalent to $k<\alpha(1+\mu)/s$, i.e., the unit braking cost has to be less than a threshold for overtaking behaviors to happen. For ease of exposition, denote

$$\theta=\frac{sk}{\alpha(1+\mu)}$$

Then (4.10) can be rewritten as $\lambda_o(t)=s(1-\theta)/(1+\mu)$, and condition $k<\alpha(1+\mu)/s$ is equivalent to $\theta<1$. Throughout this chapter, $k<\alpha(1+\mu)/s$ and $\theta<1$ are both used and considered interchangeable. In other words, θ represents the unit overtaking cost k.

Making use of (4.9) and (4.10), it gives that

$$P^{\text{pay}}(t)=\alpha(\tilde{t}_z-t_b)+\gamma(\tilde{t}_z-t^{*})+\rho \tag{4.11}$$

At equilibrium, $P^{\text{pay}}(t) = P^{\text{non}}(t)$, from (4.8) and (4.11), the toll price satisfies

$$\rho = \alpha(1+\mu)\left(t^{-} - \tilde{t}_z\right) \tag{4.12}$$

For overtaking commuters arriving after $\tilde{t}_z$, their arrival rate and travel price is same as that of equation (4.10) and (4.11). The last overtaking commuter is assumed to arrive at t_o. If there is no capacity waste at t^{-}, at equilibrium, it holds that

$$\frac{s}{1+\mu}(1-\theta)(t_o - t_b) = s\left(t^{-} - \tilde{t}_z\right) \tag{4.13}$$

(4.13) gives

$$t_o = t_b + \frac{(1+\mu)\left(t^{-} - \tilde{t}_z\right)}{1-\theta} \tag{4.14}$$

Since $\tilde{t}_z < t^{-}$, from (4.14), it gives $t_o > t_b$. (4.12) and (4.14) show that, when $\rho \to 0$, it holds $\tilde{t}_z \to t^{-}$ and $t_o \to t_b$, meaning that when toll level is sufficiently low, commuters will not have much incentive to brake or overtake.

Equation (4.9) shows that the first overtaking commuter arrives right after t_b. That is because if she arrives any moment after t_b, she could be incurred a higher overtaking cost. Mathematically, it can be proved like this: suppose the first overtaking commuter arrives at $t_{o'}$, compared with the toll payer arriving right before t_b, at equilibrium, it holds that $\alpha\left(t_{o'} - t_b\right) = k\int_{t_b}^{t_{o'}} \lambda_b(\omega)\,d\omega$, which gives $\alpha\left(t_{o'} - t_b\right) = k\left(t_{o'} - t_b\right)s/(1+\mu)$, or equivalently, $\left(t_{o'} - t_b\right) = \theta\left(t_{o'} - t_b\right)$. This readily gives $t_{o'} = t_b$ in view of $\theta < 1$.

The first commuter and last commuter has no queuing cost, so it holds that $\beta\left(t^{*} - t_q\right) = \gamma\left(t_{q'} - t^{*}\right)$. If there is no capacity waste, with the fact of $t_{q'} - t_q = N/s$, it can be obtained that

$$t^{*} - t_q = \frac{\mu}{\eta+\mu}\frac{N}{s} \text{ and } t_{q'} - t^{*} = \frac{\eta}{\eta+\mu}\frac{N}{s}$$

The toll non-payer's travel price can now be given as

$$P^{\text{non}}(t) = \alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.15}$$

From (4.15) and (4.8), it gives that

$$t_b = t^{-} + \mu\left(t^{-} - t^{*}\right) - \frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.16}$$

(4.16) implies that at equilibrium, the time braking behavior happens is not affected by toll level. It is only determined by the toll ending time. From (4.3) and (4.11), it can be obtained that

$$t_y = t^{+} - \tilde{t}_z + t_b - \mu\left(\tilde{t}_z - t^{*}\right) + \eta\left(t^{*} - t^{+}\right) \tag{4.17}$$

The conditions that no capacity waste exists at t^+ and t^-, and all overtaking commuters have schedule late delay are given by

$$t_y \le t^+ \tag{4.18}$$

$$t_o \le t^- \tag{4.19}$$

$$\tilde{t}_z \ge t^* \tag{4.20}$$

(4.18) states that, the condition there is no capacity waste at t^+ is that the first toll payer arrives before t^+. (4.19) states that, the condition there is no capacity waste at t^- is that the last overtaking commuter arrives at the bottleneck before t^-. (4.20) states that all overtaking commuters have schedule late delay cost. By setting $t_y = t^+$ and $t_o = t^-$, from (4.14), (4.16) and (4.17) , it gives two critical $\tilde{t}_z$ values (representing two critical toll levels), denoted by $\tilde{t}_z^y$ and $\tilde{t}_z^o$, respectively, and given by

$$\tilde{t}_z^y = t^- + \frac{\eta}{1+\mu}\left(t^* - t^+\right) - \frac{\eta\mu}{(1+\mu)(\eta+\mu)}\frac{N}{s}$$

$$\tilde{t}_z^o = t^- - \frac{1}{1+\mu}(1-\theta)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right]$$

Conditions (4.18) and (4.19) are equivalent to (4.21) and (4.22), respectively

$$\tilde{t}_z \ge \tilde{t}_z^y \tag{4.21}$$

$$\tilde{t}_z \ge \tilde{t}_z^o \tag{4.22}$$

From (4.12), the lower bounds on $\tilde{t}_z$ given by conditions (4.20)-(4.22) represent three upper bounds on the toll level. Since $\tilde{t}_z < t^-$, from (4.21) and (4.22), it can be obtained that

$$t^* - t^+ < \frac{\mu}{\eta+\mu}\frac{N}{s} \text{ and } t^- - t^* < \frac{\eta}{\eta+\mu}\frac{N}{s}$$

The equilibrium profile of the overtaking model is stable and cannot be replaced by the braking model for $\theta < 1$, which is explained in the following proposition.

Proposition 4.1. If the unit overtaking cost is low ($\theta < 1$), the equilibrium profile of the braking model will collapse and becomes the equilibrium profile of the overtaking model.

Proof. Figure 4.2 shows the equilibrium profile of the braking model.

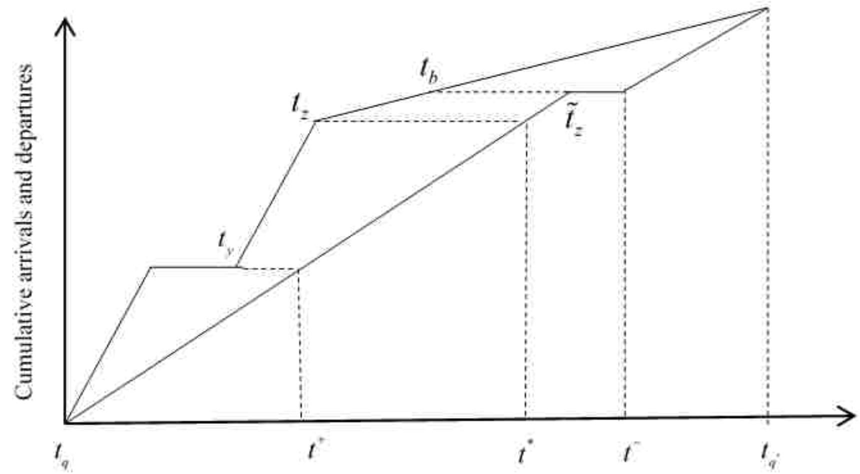

Figure 4.2. Equilibrium profile without capacity waste in the braking model

For $\theta < 1$, under the equilibrium profile of the braking model, as a commuter is allowed to overtake, she could overtake at $\tilde{t}_z$, so that she does not have queuing cost. Her travel price is given as

$$P^{\text{pay}} = \gamma\left(\tilde{t}_z - t^*\right) + k\int_{t_b}^{\tilde{t}_z} \lambda_b\left(\omega\right) d\omega + \rho \tag{4.23}$$

The toll payer arriving just before t_b has a travel price of

$$P^{\text{pay}} = \alpha\left(\tilde{t}_z - t_b\right) + \gamma\left(\tilde{t}_z - t^*\right) + \rho \tag{4.24}$$

Since $\lambda_b\left(\omega\right) = s/\left(1+\mu\right)$ and $\theta < 1$, from (4.23) and (4.24) it holds that

$$\gamma\left(\tilde{t}_z - t^*\right) + k\frac{s}{1+\mu}\left(\tilde{t}_z - t_b\right) + \rho < \alpha\left(\tilde{t}_z - t_b\right) + \gamma\left(\tilde{t}_z - t^*\right) + \rho$$

which means that, if overtaking behavior is allowed, the commuter would have incentive to overtake to reduce travel price and the equilibrium profile of braking model would collapse. So, for $\theta < 1$, only the equilibrium profile of overtaking model exists. This also shows that, if $\theta \geq 1$, overtaking behavior does not exist and the equilibrium profile will be the braking model.

4.3. Optimal tolling scheme of equilibrium profile without capacity waste

Equation (4.12) shows that ρ is a function of $\tilde{t}_z$ and $t^- - t^*$. A tolling scheme is determined by ρ, $t^* - t^+$ and $t^- - t^*$. So, the total system cost can be treated as a function of $\tilde{t}_z$, $t^* - t^+$ and $t^- - t^*$. For homogeneous users, the total system cost is given as

$$TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right) = \alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N - \alpha\left(1+\mu\right)\left(t^- - \tilde{t}_z\right)s\left(t^- - t^+\right) \tag{4.25}$$

In (4.25), the first term is the total travel price. Second term is the total toll revenue. It can be acquired that

$$\frac{\partial TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right)}{\partial \tilde{t}_z} = \alpha\left(1+\mu\right)s\left(t^- - t^+\right) > 0$$

so, for a given toll window $\left(t^* - t^+, t^- - t^*\right)$, $TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right)$ increases with $\tilde{t}_z$. In the following we will compare the three lower bounds on $\tilde{t}_z$ given by conditions (4.20)-(4.22). Comparing the three lower bounds ($\tilde{t}_z^y$, $\tilde{t}_z^o$ and t^*) results in three lines in the $\left(t^* - t^+, t^- - t^*\right)$ two-dimensional space as shown in Figure 4.3.

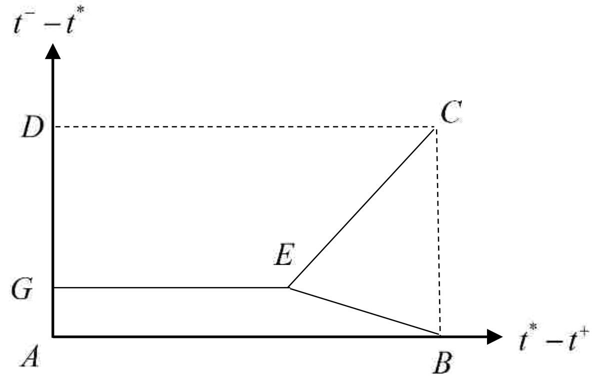

Figure 4.3. The $\left(t^* - t^+, t^- - t^*\right)$ two-dimensional space

To compare which one of $\tilde{t}_z^y$ and $\tilde{t}_z^o$ is more critical, set $\tilde{t}_z^y = \tilde{t}_z^o$, which is shown as EC in Figure 4.3. The equation of EC is given as

$$\eta\left(t^* - t^+\right) - \mu\left(1-\theta\right)\left(t^- - t^*\right) = \theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.26}$$

The coordinate of C is $\left(\mu N/\left(\eta+\mu\right)s, \eta N/\left(\eta+\mu\right)s\right)$. To compare which one of $\tilde{t}_z^y$ and t^* is more critical, set $\tilde{t}_z^y = t^*$, which is shown as EB in Figure 4.3. The equation of EB is given as

$$t^- - t^* + \frac{\eta}{1+\mu}\left(t^* - t^+\right) = \frac{\eta\mu}{\left(1+\mu\right)\left(\eta+\mu\right)}\frac{N}{s} \tag{4.27}$$

To compare which one of $\tilde{t}_z^o$ and t^* is more critical, set $\tilde{t}_z^o = t^*$, which is shown as EG in Figure 4.3. The equation of EG is given as

$$t^- - t^* = \frac{1-\theta}{1+\mu\left(2-\theta\right)}\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.28}$$

To show the derivation of the optimal tolling scheme, the following lemma is introduced.

Lemma 4.1. For the equilibrium profile that no capacity waste exists at t^+ and t^-, and all overtaking commuters have schedule late delay, the optimal toll window is on EC in Figure 4.3.

Proof. In Figure 4.3, a toll window on EC is featured by $t_y = t^+$ and $t_o = t^-$ under the critical toll. In $DGEC$, it holds that $\tilde{t}_z^o > \tilde{t}_z^y$ and $\tilde{t}_z^o > t^*$. Based on equilibrium condition (4.18), (4.19) and

(4.20), it holds that $\tilde{t}_z \geq \tilde{t}_z^o$. As $\partial TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right) / \partial \tilde{t}_z > 0$, in $DGEC$, the minimum total cost of a given toll window is obtained by setting $\tilde{t}_z = \tilde{t}_z^o$. From (4.25), it is given by

$$TC^{DGEC}\left(t^* - t^+, t^- - t^*\right) = \alpha \frac{\eta\mu}{\eta + \mu} \frac{N}{s} N - \alpha(1-\theta)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right] s\left(t^- - t^+\right)$$

It can be obtained that

$$\frac{\partial TC^{DGEC}}{\partial\left(t^* - t^+\right)} = -\alpha(1-\theta)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right] s < 0$$

So, given $t^- - t^*$, TC^{DGEC} decreases with $t^* - t^+$. The optimal toll window is on EC.

In ECB, it holds that $\tilde{t}_z^y > \tilde{t}_z^o$ and $\tilde{t}_z^y > t^*$. Based on equilibrium condition, it holds that $\tilde{t}_z \geq \tilde{t}_z^y$. As $\partial TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right) / \partial \tilde{t}_z > 0$, in ECB, the minimum total cost of a given toll window is obtained by setting $\tilde{t}_z = \tilde{t}_z^y$. From (4.25), it is given by

$$TC^{ECB}\left(t^* - t^+, t^- - t^*\right) = \alpha \frac{\eta\mu}{\eta+\mu}\frac{N}{s} N - \alpha(1+\mu)\left[\frac{\eta\mu}{(1+\mu)(\eta+\mu)}\frac{N}{s} - \frac{\eta}{1+\mu}\left(t^* - t^+\right)\right] s\left(t^- - t^+\right)$$

It can be obtained that

$$\frac{\partial TC^{ECB}}{\partial\left(t^- - t^*\right)} = -\alpha(1+\mu)\left[\frac{\eta\mu}{(1+\mu)(\eta+\mu)}\frac{N}{s} - \frac{\eta}{1+\mu}\left(t^* - t^+\right)\right] s < 0$$

So, given $t^* - t^+$, TC^{ECB} decreases with $t^- - t^*$. The optimal toll window is on EC.

In $GEBA$, it holds that $t^* > \tilde{t}_z^o$ and $t^* > \tilde{t}_z^y$. Based on equilibrium condition, it holds that $\tilde{t}_z \geq t^*$. As $\partial TC\left(\tilde{t}_z, t^* - t^+, t^- - t^*\right) / \partial \tilde{t}_z > 0$, in $GEBA$, the minimum total cost of a given toll window is obtained by setting $\tilde{t}_z = t^*$. From (4.25), it is given by

$$TC^{GEBA}\left(t^* - t^+, t^- - t^*\right) = \alpha \frac{\eta\mu}{\eta+\mu}\frac{N}{s} N - \alpha(1+\mu)\left(t^- - t^*\right) s\left(t^- - t^+\right)$$

It can be obtained that

$$\frac{\partial TC^{GEBA}}{\partial\left(t^* - t^+\right)} = -\alpha(1+\mu)\left(t^- - t^*\right) s < 0$$

So, given $t^- - t^*$, TC^{GEBA} decreases with $t^* - t^+$. The optimal toll window is on EB. Since on EB it holds that $\tilde{t}_z^y = t^*$ and $TC^{ECB} = TC^{GEBA}$, it readily gives that in $ABCD$ the optimal toll window is on EC. This completes the proof.

Let $\left(t^* - t^+\right)^{\text{op}}$ and $\left(t^- - t^*\right)^{\text{op}}$ denote the optimal toll window, and $\left(t^- - t^+\right)^{\text{op}}$ denote the optimal toll window length . Let $\left(t^* - t^+\right)^E$ and $\left(t^- - t^*\right)^E$ denote the toll window, and $\left(t^- - t^+\right)^E$ denote the toll window length corresponding to E in Figure 4.3. Let $\left(t^*-t^+\right)_{globe}^{EC}$ and $\left(t^--t^*\right)_{globe}^{EC}$ denote the global optimal toll window on EC, where

$$\left(t^*-t^+\right)_{globe}^{EC}=\frac{\frac{\eta\mu}{\eta+\mu}\theta+\frac{1}{2}\mu\left(1-\theta\right)}{\eta+\mu\left(1-\theta\right)}\frac{N}{s}$$

$$\left(t^--t^*\right)_{globe}^{EC}=\frac{\frac{1}{2}\eta-\frac{\eta\mu}{\eta+\mu}\theta}{\eta+\mu\left(1-\theta\right)}\frac{N}{s}$$

Based on Lemma 4.1, we have the following proposition.

Proposition 4.2. For the equilibrium profile that no capacity waste exists at t^+ and t^-, and all overtaking commuters have schedule late delay, there exists a critical unit overtaking cost represented by

$$\theta_E=\frac{\eta+\mu}{\mu\left(2+\mu-\eta\right)}$$

such that the optimal toll window is given as

(a). If $\theta\in\left[0,\theta_E\right]$, it holds $\left(t^*-t^+\right)^{\text{op}}=\left(t^*-t^+\right)_{globe}^{EC}$, $\left(t^--t^*\right)^{\text{op}}=\left(t^--t^*\right)_{globe}^{EC}$, and $\left(t^--t^+\right)^{\text{op}}=N/2s$.

(b). If $\theta\in\left(\theta_E,1\right)$, it holds $\left(t^*-t^+\right)^{\text{op}}=\left(t^*-t^+\right)^{E}$, $\left(t^--t^*\right)^{\text{op}}=\left(t^--t^*\right)^{E}$, and $\left(t^--t^+\right)^{\text{op}}>N/2s$.

Proof. On EC, the minimum total cost of a given toll window can be obtained as

$$TC^{EC}\left(t^--t^*\right)=\alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N-\alpha s\left(1-\theta\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\mu\left(t^--t^*\right)\right]\times$$
$$\left[t^--t^*+\theta\frac{\mu}{\eta+\mu}\frac{N}{s}+\frac{\mu}{\eta}\left(1-\theta\right)\left(t^--t^*\right)\right]$$

It can be obtained that

$$\frac{dTC^{EC}\left(t^--t^*\right)}{d\left(t^--t^*\right)}=-\alpha s\left(1-\theta\right)\left\{-\mu\left[t^--t^*+\theta\frac{\mu}{\eta+\mu}\frac{N}{s}+\frac{\mu}{\eta}\left(1-\theta\right)\left(t^--t^*\right)\right]+\right.$$
$$\left.\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\mu\left(t^--t^*\right)\right]\left[1+\frac{\mu}{\eta}\left(1-\theta\right)\right]\right\}$$

$$\frac{d^2TC^{EC}\left(t^--t^*\right)}{d\left(t^--t^*\right)^2}=2\alpha s\left(1-\theta\right)\mu\left[1+\frac{\mu}{\eta}\left(1-\theta\right)\right]>0$$

So, $TC^{EC}\left(t^- - t^*\right)$ is strictly convex for $t^- - t^* \in \left(-\infty, +\infty\right)$. By setting $dTC^{EC}\left(t^- - t^*\right) \big/ d\left(t^- - t^*\right) = 0$, it gives that

$$\left(t^- - t^*\right)^{\text{op}} = \left(t^- - t^*\right)^{EC}_{globe} \tag{4.29}$$

From (4.26), the optimal $t^* - t^+$ can be obtained as

$$\left(t^* - t^+\right)^{\text{op}} = \left(t^* - t^+\right)^{EC}_{globe} \tag{4.30}$$

From (4.29) and (4.30), the optimal toll window length can be obtained as

$$\left(t^- - t^+\right)^{\text{op}} = \frac{1}{2}\frac{N}{s}$$

The optimal total cost can be acquired as

$$TC^{\text{op}} = \alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N - \alpha s\left(1-\theta\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)^{EC}_{globe}\right]\frac{1}{2}\frac{N}{s} \tag{4.31}$$

The optimal toll level is obtained as

$$\rho^{\text{op}} = \alpha\left(1-\theta\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)^{EC}_{globe}\right] \tag{4.32}$$

As the optimal toll window is on EC, it must hold that $\left(t^- - t^*\right)^{\text{op}} \geq \left(t^- - t^*\right)^{E}$, which readily gives $0 \leq \theta \leq \theta_E$. This completes the proof of Proposition 4.2(a).

For $\theta \in \left(\theta_E, 1\right)$, it holds $\left(t^- - t^*\right)^{EC}_{globe} < \left(t^- - t^*\right)^{E}$. As $d^2TC^{EC}\left(t^- - t^*\right) \big/ d\left(t^- - t^*\right)^2 > 0$, it holds that $dTC^{EC}\left(t^- - t^*\right) \big/ d\left(t^- - t^*\right) > 0$ for $t^- - t^* > \left(t^- - t^*\right)^{EC}_{globe}$. Since E has the minimum $t^- - t^*$ on EC, the optimal toll window is given by

$$\left(t^- - t^*\right)^{\text{op}} = \left(t^- - t^*\right)^{E} = \frac{1-\theta}{1+\mu\left(2-\theta\right)}\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.33}$$

$$\left(t^* - t^+\right)^{\text{op}} = \left(t^* - t^+\right)^{E} = \frac{\theta+\mu}{1+\mu\left(2-\theta\right)}\frac{\mu}{\eta+\mu}\frac{N}{s} \tag{4.34}$$

The optimal toll window length can be obtained as

$$\left(t^- - t^+\right)^{\text{op}} = \left(t^- - t^+\right)^{E} = \frac{\left(1-\eta\right)\theta+\eta+\mu}{1+\mu\left(2-\theta\right)}\frac{\mu}{\eta+\mu}\frac{N}{s} > \frac{1}{2}\frac{N}{s} \tag{4.35}$$

The optimal total cost can be acquired as

$$TC^{\text{op}} = \alpha N\frac{\eta\mu}{\eta+\mu}\left[\frac{N}{s} - \frac{\left(1+\mu\right)\left(1-\theta\right)}{1+\mu\left(2-\theta\right)}\left(t^- - t^+\right)^{E}\right] \tag{4.36}$$

The optimal toll level is obtained as

$$\rho^{\text{op}} = \alpha\left(1-\theta\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)^{E}\right] \tag{4.37}$$

This completes the proof of Proposition 4.2(b).

Proposition 4.2 states that, for equilibrium profile that no capacity waste exists and all overtaking commuters have schedule late delay, the optimal tolling scheme is featured by that the first toll payer arrives at t^+ and the last overtaking commuter arrives at t^-, both of them experiencing no queuing delay. For $\theta \le \theta_E$ and $\theta > \theta_E$, although the forms of solutions are different, the properties of them are identical. For $\theta \le \theta_E$, the optimal toll window is obtained by using the first order condition. For $\theta > \theta_E$, as the solution obtained from first order condition is below E and thus not feasible, the optimal toll window is given by E.

4.4. Impact of k

In this section, the impact of the unit overtaking cost on the optimal tolling scheme is investigated and discussed. Lemma 4.2 is used to summarize the impact of k.

Lemma 4.2. For the equilibrium profile that no capacity waste exists at t^+ and t^-, and all overtaking commuters have schedule late delay, the optimal tolling scheme has the following properties:

(a). If $\theta \in [0, \theta_E]$, TC^{op} increases with k, ρ^{op} decreases with k, $\left(t^- - t^+\right)^{\text{op}} = N/2s$ and the optimal toll window position becomes earlier as k increases.

(b). If $\theta \in (\theta_E, 1)$, TC^{op} increases with k, ρ^{op} decreases with k, $\left(t^- - t^+\right)^{\text{op}}$ increases with k, $\left(t^- - t^+\right)^{\text{op}} > N/2s$ and the optimal toll window position becomes earlier as k increases.

Proof. For $\theta \in [0, \theta_E]$, from (4.29), it can be obtained that

$$\frac{d\left(t^- - t^*\right)^{\text{op}}}{dk} = \frac{-\frac{1}{2}\frac{\mu}{\eta}\frac{s}{\alpha(1+\mu)}}{\left[1+\frac{\mu}{\eta}(1-\theta)\right]^2}\frac{N}{s} < 0$$

so $\left(t^- - t^*\right)^{\text{op}}$ decreases with k. As $\left(t^- - t^+\right)^{\text{op}} = N/2s$, $\left(t^* - t^+\right)^{\text{op}}$ increases with k. So, the toll window position becomes earlier with k. From (4.31), it can be obtained that

$$\frac{dTC^{\text{op}}}{dk} = \alpha s \frac{1}{2}\frac{N}{s}\frac{s}{\alpha(1+\mu)}\mu \left\{ \frac{\eta}{\eta+\mu}\frac{N}{s} - \frac{\left(t^- - t^*\right)^{EC}_{globe}}{1+\frac{\mu}{\eta}(1-\theta)} + \frac{\frac{\mu^2}{\eta(\eta+\mu)}\theta(1-\theta)}{\left(1+\frac{\mu}{\eta}(1-\theta)\right)^2}\frac{N}{s} \right\}$$

Since $\frac{\eta}{\eta+\mu}\frac{N}{s} > \left(t^- - t^*\right)^{EC}_{globe}$ for $\theta \in [0, \theta_E]$, it holds that $\frac{dTC^{\text{op}}}{dk} > 0$. So, the optimal total cost increases with k. From (4.32), it can be obtained that

$$\frac{d\rho^{\text{op}}}{dk}=-\alpha\frac{s}{\alpha(1+\mu)}\mu\left\{\frac{\eta}{\eta+\mu}\frac{N}{s}-\frac{\left(t^{-}-t^{*}\right)_{globe}^{EC}}{1+\frac{\mu}{\eta}(1-\theta)}+\frac{\frac{\mu^{2}}{\eta(\eta+\mu)}\theta(1-\theta)}{\left(1+\frac{\mu}{\eta}(1-\theta)\right)^{2}}\frac{N}{s}\right\}<0$$

So, the optimal toll level decreases with k.

For $\theta\in(\theta_E,1)$, from (4.33), it can be obtained that

$$\frac{d\left(t^{-}-t^{*}\right)^{\text{op}}}{dk}=\frac{\eta}{\eta+\mu}\frac{N}{s}\frac{-\frac{\mu}{1+\mu}\frac{s}{\alpha(1+\mu)}}{\left(1+\frac{\mu}{1+\mu}(1-\theta)\right)^{2}}<0$$

So, $\left(t^{-}-t^{*}\right)^{\text{op}}$ decreases with k. From (4.34), it is straight forward that $\left(t^{*}-t^{+}\right)^{\text{op}}$ increases with k. From (4.35), it is straight forward that $\left(t^{-}-t^{+}\right)^{\text{op}}$ increases with k. So, the toll window position becomes earlier. From (4.37), it can be obtained that

$$\frac{d\rho^{\text{op}}}{dk}=\alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}\frac{-\frac{s}{\alpha(1+\mu)}}{\left(1+\frac{\mu}{1+\mu}(1-\theta)\right)^{2}}<0$$

So, ρ^{op} decreases with k. From (4.36), it can be obtained that

$$\frac{dTC^{\text{op}}}{dk}=-\alpha s\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}\right)^{2}\frac{\frac{s}{\alpha(1+\mu)}\frac{\mu}{1+\mu}\frac{1}{\eta+\mu}\left[1-2\eta-\theta(2-2\eta+\mu)\right]}{\left(1+\frac{\mu}{1+\mu}(1-\theta)\right)^{3}}$$

Since $\theta>\theta_E$, it holds that $\theta(2-2\eta+\mu)>1$, which readily gives $\frac{dTC^{\text{op}}}{dk}>0$. So, the optimal total cost increases with k. This completes the proof.

From (4.31), if $\theta=0$ (i.e. overtaking does not incur cost), it can be obtained that

$$\lambda_o(t)=\frac{s}{1+\mu}$$

$$TC^{\text{op}}=\frac{3}{4}\alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N$$

which are the arrival rate of toll payers having schedule late delay and the optimal total cost in the Laih model. This implies that Laih model can be treated as a special case of the overtaking model when overtaking incurs cost. As TC^{op} increases with k, the optimal total cost of the overtaking model is higher than that of Laih model.

From (4.36) and (4.37), if $\theta \to 1$, it holds that $\rho^{\text{op}} \to 0$, $\lambda_o(t) \to 0$ and $TC^{\text{op}} \to \alpha \frac{\eta\mu}{\eta+\mu}\frac{N}{s}N$. This implies that, when unit overtaking cost is sufficiently large, the optimal toll level approaches zero and the optimal total cost approaches that of no toll equilibrium. Since the optimal total cost of braking model is lower than that of no toll equilibrium, this indicates that when unit overtaking cost is sufficiently large, the braking model is better than the overtaking model in terms of total system cost. When the toll level is sufficiently high, there will be capacity waste in the overtaking model and the overtaking model can reduce to the braking model. This indicates that, in the overtaking model, the equilibrium profile with capacity waste can have a lower total cost than that without capacity waste. Thus, in the next section, we will investigate the equilibrium profile of the overtaking model with capacity waste.

4.5. Equilibrium profile with capacity waste

Figure 4.1 shows the equilibrium profile of overtaking model without capacity waste. For a given toll window, $\tilde{t}_z^y$, $\tilde{t}_z^o$ and t^* corresponds to three different critical toll level. If toll price exceeds the critical toll level, there will be capacity waste at t^+ or t^-, or make $\tilde{t}_z < t^*$ (i.e. some overtaking commuters have schedule early delay). In this study, we only need to consider the capacity waste scenario where capacity waste appears at t^-, no capacity waste exists at t^+ and all overtaking commuters have schedule late delay. This requires the toll window in $DGEC$ in Figure 4.3. Toll windows in ECB and $GEBA$ in Figure 4.3 do not need to be considered for the following reasons.

For a toll window in ECB in Figure 4.3, if $\rho > \alpha(1+\mu)\left(t^- - \tilde{t}_z^y\right)$, capacity waste will appear at t^+ and no capacity waste exists at t^-. This equilibrium profile need not be considered because capacity waste at t^+ can only cause a pure system waste from t^+ to t_y (i.e. a tolling period no one comes or leaves the bottleneck). Such waste increases the length of peak hour and the number of toll non-payers. The decrement of toll payers' system cost can not offset the increment of toll non-payers' system cost, thus making total cost higher than the no waste profile.

In Figure 4.3, for a toll window in $GEBA$, if $\rho > \alpha(1+\mu)\left(t^- - t^*\right)$, it holds that $\tilde{t}_z < t^*$ and no capacity waste exists at t^+ or t^-. In $GEBA$, when toll level is sufficiently large, the equilibrium profile of the overtaking model will reduce to that of the braking model where all toll payers have schedule early delay. Such profile has been proved by Xiao et al. (2012) to be worse than the profile where some toll payers have schedule late delay (which is achieved by toll windows in $DGEC$). Thus, toll windows in $GEBA$ need not be considered.

In summary, only toll windows in $DGEC$ in Figure 4.3 need to be considered for the case with capacity waste. For a toll window in $DGEC$, if $\rho > \alpha(1+\mu)(t^- - \tilde{t}_z^o)$, capacity waste will appear at t^-, no capacity waste exists at t^+ and all overtaking commuters have schedule late delay. Figure 4.4 shows such an equilibrium profile. As can be seen in Figure 4.4, since toll level is high, the last overtaking commuter has no queuing delay and leaves the bottleneck before t^- (i.e. $t_o \le t^-$). No commuter would have incentive to overtake between t_o and t^-.

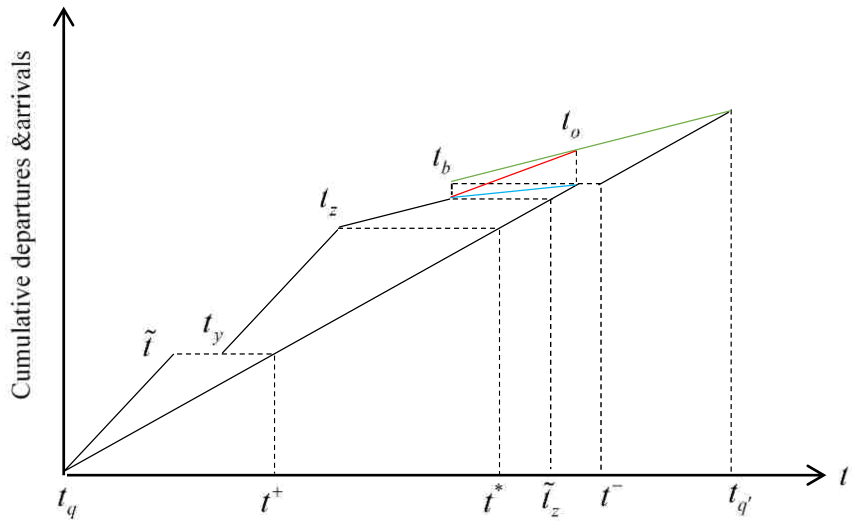

Figure 4.4. Equilibrium profile with capacity waste at only t^-

For the profile with capacity waste at t^-, as toll only causes some overtaking commuters to join braking (i.e. partially overtaking), equation (4.1) to (4.12) still hold. Since no commuters overtake between t_o and t^-, the length of the peak hour is given as

$$t_{q'} - t_q = \frac{N}{s} + t^- - t_o \tag{4.38}$$

With the fact of $\beta(t^* - t_q) = \gamma(t_{q'} - t^*)$, it holds that

$$t^* - t_q = \frac{\mu}{\eta + \mu}\left(\frac{N}{s} + t^- - t_o\right) \tag{4.39}$$

The last overtaking commuter arrives at t_o and experiences no queuing delay. Her travel price is given as

$$P^{\text{pay}}(t) = \gamma(t_o - t^*) + k\frac{s}{1+\mu}(t_o - t_b) + \rho \tag{4.40}$$

Making use of (4.2), (4.8) and (4.39), it can be obtained that

$$t_b = t^- + \mu(t^- - t^*) - \frac{\eta\mu}{\eta + \mu}\left(\frac{N}{s} + t^- - t_o\right) \tag{4.41}$$

Making use of (4.11), (4.40) and (4.41), it gives

$$\tilde{t}_z = \frac{1}{1+\mu}(1-\theta)\left[t^- + \mu\left(t^- - t^*\right) - \frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}t^-\right] + \frac{1}{1+\mu}\left[\mu + \theta + \frac{\eta\mu}{\eta+\mu}(1-\theta)\right]t_o \tag{4.42}$$

The equilibrium condition that capacity waste only exists at t^- and all overtaking commuters have schedule late delay is given by

$$\tilde{t}_z \geq t_b \tag{4.43}$$

$$t_y \leq t^+ \tag{4.44}$$

$$\tilde{t}_z \geq t^* \tag{4.45}$$

(4.43) gives the condition that overtaking behavior exists. If $\tilde{t}_z = t_b$, no commuter would overtake, and the overtaking model just reduces to the braking model. From (4.11) and (4.40), condition (4.43) is equivalent to $t_o \geq t_b$, which also means that a sufficiently high toll would make $t_o = t_b$ (i.e. overtaking behavior does not exist). (4.44) gives the condition that there is no capacity waste at t^+. (4.45) simply states that all overtaking commuters have schedule late delay. By setting conditions (4.43)-(4.45) to equality, i.e., setting $\tilde{t}_z = t_b$, $t_y = t^+$ and $\tilde{t}_z = t^*$, from (4.17), (4.41) and (4.42), it gives three critical t_o values (representing three critical toll levels), denoted by t_o^{w1}, t_o^{w2} and t_o^{w3}, respectively, and given by

$$t_o^{w1} = t^- - \frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right] \tag{4.46}$$

$$t_o^{w2} = \frac{\mu t^* + \eta\left(t^* - t^+\right) + \theta\left[t^- + \mu\left(t^- - t^*\right) - \frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}t^-\right]}{\mu + \frac{\eta+\mu-\eta\mu}{\eta+\mu}\theta} \tag{4.47}$$

$$t_o^{w3} = \frac{(1+\mu)t^* - (1-\theta)\left[t^- + \mu\left(t^- - t^*\right) - \frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}t^-\right]}{\mu + \theta + \frac{\eta\mu}{\eta+\mu}(1-\theta)}$$

Conditions (4.43)-(4.45) are equivalent to (4.48)-(4.50), respectively

$$t_o \geq t_o^{w1} \tag{4.48}$$

$$t_o \geq t_o^{w2} \tag{4.49}$$

$$t_o \geq t_o^{w3} \tag{4.50}$$

From (4.12) and (4.42), the lower bounds on t_o given by conditions (4.48)-(4.50) represent three upper bounds on the toll level. Since $t_o \leq t^-$, from (4.48)-(4.50), it can be obtained that

$$t^- - t^* \leq \frac{\eta}{\eta+\mu}\frac{N}{s} \tag{4.51}$$

$$(1-\theta)\mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right) \geq -\theta \frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.52}$$

$$t^- - t^* \geq \frac{1-\theta}{1+\mu(2-\theta)}\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.53}$$

(4.51)-(4.53) corresponds to region $DGEC$ in Figure 4.3. Specifically, setting (4.51)-(4.53) to equality represents CD, EC and EG, respectively. For a given toll window, t_o^{w1}, t_o^{w2} and t_o^{w3} corresponds to three different critical toll levels which can be obtained by making use of (4.12) and (4.42). At $t_o = t_o^{w1}$, no commuters would overtake. At $t_o = t_o^{w2}$, the first toll payer arrives at t^+. At $t_o = t_o^{w3}$, any higher toll will make some overtaking commuters arrive at work early.

Comparing the three lower bounds (t_o^{w1}, t_o^{w2} and t_o^{w3}) results in two lines in the $\left(t^* - t^+, t^- - t^*\right)$ two-dimensional space as shown in Figure 4.5. To compare which one of t_o^{w1} and t_o^{w2} is more critical, set $t_o^{w1} = t_o^{w2}$, which is shown as CF in Figure 4.5. The equation of CF is given as

$$\left(\eta+\mu+\mu^2\right)\mu\left(t^- - t^*\right) - \eta\left(\eta+\mu-\eta\mu\right)\left(t^* - t^+\right) = \eta\mu^2\frac{N}{s} \tag{4.54}$$

To compare which one of t_o^{w2} and t_o^{w3} is more critical, set $t_o^{w2} = t_o^{w3}$, which is shown as EF in Figure 4.5. The equation of EF is given as

$$\left[\theta\left(\eta+\mu-\eta\mu\right)+\mu^2+2\eta\mu\right]\eta\left(t^* - t^+\right) + \left(\eta+\mu+\mu^2\right)\left(\mu+\theta\right)\left(t^- - t^*\right) = \left(\mu+\theta\right)\eta\mu\frac{N}{s} \tag{4.55}$$

In CFD, it holds that $t_o^{w1} > t_o^{w2} > t_o^{w3}$. In CFE, it holds that $t_o^{w2} > t_o^{w1}$, $t_o^{w2} > t_o^{w3}$. In FEG, it holds that $t_o^{w3} > t_o^{w2} > t_o^{w1}$. HI, HJ, JK and KL will be explained in next section.

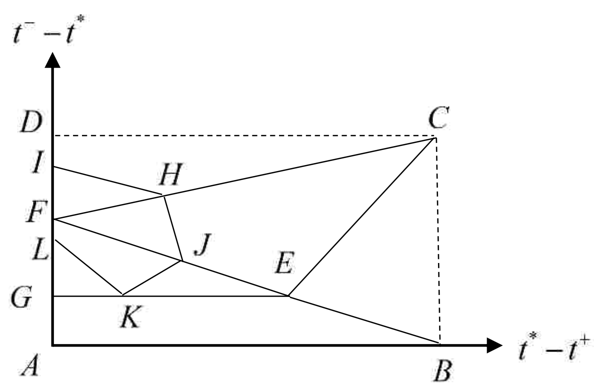

Figure 4.5. $\left(t^* - t^+, t^- - t^*\right)$ two-dimensional space for profile with capacity waste at t^-

4.6. Optimal tolling scheme of equilibrium profile with capacity waste

In this section, we investigate the optimal tolling scheme for profile with capacity waste. ρ is a function of t_o and $t^- - t^*$. A tolling scheme is determined by ρ, $t^* - t^+$ and $t^- - t^*$. So the total system cost can be treated as a function of t_o, $t^* - t^+$ and $t^- - t^*$. The total system cost is given as

$$TC\left(t_o, t^* - t^+, t^- - t^*\right) = \alpha \frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s} + t^- - t_o\right)N - \alpha\left(1+\mu\right)\left(t^- - \tilde{t}_z\right)s\left(t_o - t^+\right) \qquad (4.56)$$

In (4.56), the first term is the total travel price. The second term is the total toll revenue. It can be obtained that

$$\frac{\partial TC}{\partial t_o} = -\alpha\frac{\eta\mu}{\eta+\mu}N - \alpha\left(1+\mu\right)s\left\{-\frac{1}{1+\mu}\left[\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right]\left(t_o - t^+\right) + \left(t^- - \tilde{t}_z\right)\right\}$$

$$\frac{\partial^2 TC}{\partial t_o^{\,2}} = 2\alpha s\left[\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right]$$

It is straight forward that $\partial^2 TC / \partial t_o^{\,2} > 0$, so for a given toll window TC is strictly convex with respect to t_o at $t_o \in \left(-\infty, +\infty\right)$. By setting $\partial TC / \partial t_o = 0$, it gives

$$t_o^{globe} = \frac{\frac{\eta\mu}{\eta+\mu}\frac{N}{s} + \left[\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right]t^+ + \left(1+\mu\right)t^- - \left(1-\theta\right)\left[t^- + \mu\left(t^- - t^*\right) - \frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}t^-\right]}{2\left[\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right]}$$

Making use of t_o^{globe} and (4.56), at $t_o = t_o^{globe}$, it can be obtained that

$$TC^{globe}\left(t^* - t^+, t^- - t^*\right) = \alpha\frac{\eta\mu}{\eta+\mu}N\left(\frac{N}{s} + t^- - t^+\right) - \alpha s\left[\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right]\left(t_o^{globe} - t^+\right)^2$$

and

$$\begin{aligned}\frac{\partial TC^{globe}}{\partial\left(t^* - t^+\right)} = {} & \frac{1}{2}\alpha\frac{\eta\mu}{\eta+\mu}N - \frac{1}{2}\alpha s\left(\theta\frac{\eta+\mu+\mu^2}{\eta+\mu} + \frac{\eta\mu}{\eta+\mu}\right)\left(t^- - t^*\right) \\ & - \frac{1}{2}\alpha s\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}\left(1-\theta\right)\right)\left(t^* - t^+\right) - \frac{1}{2}\alpha s\left(1-\theta\right)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}\end{aligned}$$

For a given toll window, under equilibrium constraints, to see if t_o can take the value of t_o^{globe}, we need to compare t_o^{globe} with each of t_o^{w1}, t_o^{w2} t_o^{w3} and t^-. Comparing t_o^{globe} with each of t_o^{w1}, t_o^{w2} t_o^{w3} and t^- results in four lines HI, HJ, JK and KL in the $\left(t^* - t^+, t^- - t^*\right)$ two-dimensional space as shown in Figure 4.5. To compare which one of t_o^{w1} and t_o^{globe} is more critical, set $t_o^{w1} = t_o^{globe}$, which is shown as HI in Figure 4.5. The equation of HI is given as

$$\left[\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\frac{\eta+\mu+\eta\mu+2\mu^2}{\eta+\mu-\eta\mu}+(1-\theta)\mu\right](t^- - t^*) + \left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)(t^* - t^+) = \frac{2\eta\mu(1+\mu)}{\eta+\mu-\eta\mu}\frac{N}{s}+\theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.57}$$

To compare which one of t_o^{w2} and t_o^{globe} is more critical, set $t_o^{w2} = t_o^{globe}$, which is shown as HJ in Figure 4.5. The equation of HJ is given by

$$\begin{aligned}&\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\left(\mu+2\eta+\frac{\eta+\mu-\eta\mu}{\eta+\mu}\theta\right)(t^* - t^+)+\\&\left[\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\frac{\eta+\mu+\mu^2}{\eta+\mu}\theta-\frac{\eta\mu^2}{\eta+\mu}(1-\theta)\right](t^- - t^*)\\&=\left(\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta\right)(2+\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}+2\left(\frac{\eta\mu}{\eta+\mu}\right)^2\theta\frac{N}{s}\end{aligned} \tag{4.58}$$

To compare which one of t_o^{w3} and t_o^{globe} is more critical, set $t_o^{w3} = t_o^{globe}$, which is denoted by JK in Figure 4.5.

To compare which one of t_o^{globe} and t^- is more critical, set $t_o^{globe} = t^-$, which is shown as KL in Figure 4.5. The equation of KL is given by

$$\begin{aligned}&\left(2\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)-\mu\theta\right)(t^- - t^*)+\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)(t^* - t^+)\\&=2\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s}\end{aligned} \tag{4.59}$$

In $DLKEC$, it holds that $t^- \ge t_o^{globe}$ and

$$\begin{aligned}\frac{\partial TC^{globe}}{\partial(t^* - t^+)} &\le \frac{1}{2}\alpha\frac{\eta\mu}{\eta+\mu}N-\frac{1}{2}\alpha s(1-\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\frac{1}{2}\alpha s\left(\theta\frac{\eta+\mu+\mu^2}{\eta+\mu}+\frac{\eta\mu}{\eta+\mu}\right)\times\\&\frac{2-\theta}{\mu(2-\theta)+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)}\frac{\eta\mu}{\eta+\mu}\frac{N}{s}<0\end{aligned}$$

so, $\partial TC^{globe}/\partial(t^* - t^+) < 0$ if $t^- \ge t_o^{globe}$.

Before showing the derivation of the optimal tolling scheme, the following lemma is introduced.

Lemma 4.3. For the equilibrium profile that capacity waste only exists at t^- and all overtaking commuters have schedule late delay (i.e. toll windows in $DGEC$ of Figure 4.5), the optimal toll window is in $CHJE$ in Figure 4.5.

Proof. We first prove that the optimal toll window in LGK is on KL, then we prove that the optimal toll window in $IHJKL$ is on HI, HJ or JK, then we prove that the optimal toll window in $CDIH$ is on HC, then we prove that the optimal toll window in JKE is on JE.

In LGK, for a given toll window, as $t^- < t_o^{globe}$ and $\partial^2 TC/\partial t_o^{\,2} > 0$, for $t_o \le t^-$, it holds that $\partial TC/\partial t_o < 0$. So TC decreases with t_o. The minimum total cost of a given toll window in LGK is obtained by setting $t_o = t^-$, which gives us

$$TC^{LGK}\left(t^* - t^+, t^- - t^*\right) = \alpha \frac{\eta\mu}{\eta+\mu}\frac{N}{s}N - \alpha\left(1-\theta\right)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right]s\left(t^- - t^+\right)$$

It is straight forward that $\partial TC^{LGK}/\partial\left(t^* - t^+\right) < 0$, so the optimal toll window in LGK is on KL. On KL, it is trivial that $TC^{LGK} = TC^{globe}$.

In IHF, it holds that $t_o^{w1} < t_o^{globe}$. In HFJ, it holds that $t_o^{w2} < t_o^{globe}$. In $FJKL$, it holds that $t_o^{w3} < t_o^{globe}$. So, in $IHJKL$, for a given toll window, the minimum total cost is obtained by setting $t_o = t_o^{globe}$ and TC^{globe} is the minimum total cost. Since $\partial TC^{globe}/\partial\left(t^* - t^+\right) < 0$ for $t^- \ge t_o^{globe}$, the optimal toll window in $IHJKL$ is on HI, HJ or JK.

In $CDIH$, for a given toll window, as $t_o^{w1} > t_o^{globe}$ and $\partial^2 TC/\partial t_o^{\,2} > 0$, for $t_o \ge t_o^{w1}$, it holds that $\partial TC/\partial t_o > 0$. So TC increases with t_o. The minimum total cost of a given toll window in $CDIH$ is obtained by setting $t_o = t_o^{w1}$, which gives us

$$\begin{aligned}TC^{CDIH}\left(t^* - t^+, t^- - t^*\right) &= \alpha \frac{\eta\mu}{\eta+\mu}\frac{N^2}{s} + \alpha\frac{\eta\mu}{\eta+\mu}N\frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right] \\ &-\alpha\left(1+\mu\right)\frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right]s\left[t^- - \frac{\eta+\mu}{\eta+\mu-\eta\mu}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)\right] - t^+\right]\end{aligned}$$

It is straight forward that $\partial TC^{CDIH}/\partial\left(t^* - t^+\right) < 0$, so the optimal toll window in $CDIH$ is on HC. On HI, it holds that $TC^{CDIH} = TC^{globe}$. In JKE, it is trivial that the optimal toll window is on JE. This completes the proof.

With Lemma 4.3, we can focus on the analysis of toll windows in $CHJE$. In $CHJE$, for a given toll window, as $t_o^{w2} > t_o^{globe}$ and $\partial^2 TC/\partial t_o^{\,2} > 0$, for $t_o \ge t_o^{w2}$, it holds that $\partial TC/\partial t_o > 0$. So TC increases with t_o. The minimum total cost of a given toll window in $CHJE$ is obtained by setting $t_o = t_o^{w2}$. When $t_o = t_o^{w2}$, it holds that

$$\left(1+\mu\right)\left(t^- - \tilde{t}_z\right) = \frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s} + t^- - t_o^{w2}\right) - \eta\left(t^* - t^+\right)$$

So, based on (4.56), the minimum total cost of a given toll window in $CHJE$ is obtained as

$$TC^{CHJE}\left(t^*-t^+,t^--t^*\right)=\alpha\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}+t^--t_o^{w2}\right)N-\alpha\left[\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}+t^--t_o^{w2}\right)-\eta\left(t^*-t^+\right)\right]s\left(t_o^{w2}-t^+\right)$$

It can be obtained that

$$\begin{aligned}\frac{\partial TC^{CHJE}}{\partial\left(t^*-t^+\right)}&=-\alpha s\frac{\partial t_o^{w2}}{\partial\left(t^*-t^+\right)}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\frac{\eta\mu}{\eta+\mu}\left(t_o^{w2}-t^+\right)+\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}+t^--t_o^{w2}\right)-\eta\left(t^*-t^+\right)\right]\\&\quad+\alpha s\eta\left(t_o^{w2}-t^+\right)-\alpha s\left[\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}+t^--t_o^{w2}\right)-\eta\left(t^*-t^+\right)\right]\end{aligned}\tag{4.60}$$

where

$$\frac{\partial t_o^{w2}}{\partial\left(t^*-t^+\right)}=\frac{\eta}{\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta}\tag{4.61}$$

$$\begin{aligned}\frac{\partial TC^{CHJE}}{\partial\left(t^--t^*\right)}&=-\alpha s\frac{\partial t_o^{w2}}{\partial\left(t^--t^*\right)}\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\frac{\eta\mu}{\eta+\mu}\left(t_o^{w2}-t^+\right)+\frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s}+t^--t_o^{w2}\right)-\eta\left(t^*-t^+\right)\right]\\&\quad+\alpha\frac{\eta\mu}{\eta+\mu}N-\alpha s\frac{\eta\mu}{\eta+\mu}\left(t_o^{w2}-t^+\right)\end{aligned}\tag{4.62}$$

where

$$\frac{\partial t_o^{w2}}{\partial\left(t^--t^*\right)}=\frac{\theta\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta}{\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta}\tag{4.63}$$

On HC, it holds that $TC^{CDIH}=TC^{CHJE}$. On HJ, it holds that $TC^{CHJE}=TC^{globe}$. The optimal toll window in $DGEC$ is in $CHJE$. As $\partial t_o^{w2}/\partial\left(t^--t^*\right)$ increases with θ and $\partial t_o^{w2}/\partial\left(t^--t^*\right)<1$, for large k, in (4.62), it can be seen that

$$-\alpha s\frac{\partial t_o^{w2}}{\partial\left(t^--t^*\right)}\left(\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\frac{\eta\mu}{\eta+\mu}\left(t_o^{w2}-t^+\right)\right)+\alpha\frac{\eta\mu}{\eta+\mu}N-\alpha s\frac{\eta\mu}{\eta+\mu}\left(t_o^{w2}-t^+\right)\approx 0,$$

which leads to $\partial TC^{CHJE}/\partial\left(t^--t^*\right)<0$, so the optimal toll window is on HC.

In the following, we solve the optimal toll window in $CHJE$ and discuss its properties. To obtain the optimal toll window in $CHJE$, Kuhn tucker condition is used. The mathematical setup is given in the following.

$$Min\ TC^{CHJE}\left(t^*-t^+,t^--t^*\right)$$

subject to

$$t_o^{w1}\le t_o^{w2}$$
$$t_o^{w2}\le t^-$$
$$t_o^{w3}\le t_o^{w2}$$

$$t_o^{globe} \le t_o^{w2}$$

The lagrangian function is constructed in the following

$$L\left(t^* - t^+, t^- - t^*\right) = TC^{CHJE}\left(t^* - t^+, t^- - t^*\right) + U_1\left[-\frac{\eta\mu^2}{\eta+\mu-\eta\mu}\frac{N}{s} + \frac{\eta+\mu+\mu^2}{\eta+\mu-\eta\mu}\mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right)\right]$$

$$+U_2\left[-\theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - (1-\theta)\mu\left(t^- - t^*\right) + \eta\left(t^* - t^+\right)\right] + U_3\left[(\mu+\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}\right.$$

$$\left.-\left(\theta\frac{\eta+\mu-\eta\mu}{\eta+\mu} + \frac{\mu^2+2\eta\mu}{\eta+\mu}\right)\eta\left(t^* - t^+\right) - \left(\frac{\eta+\mu+\mu^2}{\eta+\mu}\right)(\mu+\theta)\left(t^- - t^*\right)\right]$$

$$+U_4\left\{\left(\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta\right)(2+\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s} + 2\left(\frac{\eta\mu}{\eta+\mu}\right)^2\theta\frac{N}{s} - \right.$$

$$\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\left(\mu+2\eta+\frac{\eta+\mu-\eta\mu}{\eta+\mu}\theta\right)\left(t^* - t^+\right) -$$

$$\left.\left[\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\frac{\eta+\mu+\mu^2}{\eta+\mu}\theta - \frac{\eta\mu^2}{\eta+\mu}(1-\theta)\right]\left(t^- - t^*\right)\right\}$$

It can be acquired that

$$\frac{\partial L}{\partial\left(t^* - t^+\right)} = \frac{\partial TC^{CHJE}}{\partial\left(t^* - t^+\right)} - \eta U_1 + \eta U_2 - \left(\theta\frac{\eta+\mu-\eta\mu}{\eta+\mu} + \frac{\mu^2+2\eta\mu}{\eta+\mu}\right)\eta U_3$$

$$-\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\left(\mu+2\eta+\frac{\eta+\mu-\eta\mu}{\eta+\mu}\theta\right)U_4$$

$$\frac{\partial L}{\partial\left(t^- - t^*\right)} = \frac{\partial TC^{CHJE}}{\partial\left(t^- - t^*\right)} + \frac{\eta+\mu+\mu^2}{\eta+\mu-\eta\mu}\mu U_1 - (1-\theta)\mu U_2 - \left(\frac{\eta+\mu+\mu^2}{\eta+\mu}\right)(\mu+\theta)U_3$$

$$-\left[\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\frac{\eta+\mu+\mu^2}{\eta+\mu}\theta - \frac{\eta\mu^2}{\eta+\mu}(1-\theta)\right]U_4$$

According to Kuhn tucker condition, the optimal toll window satisfies the given constraints and the following conditions:

$$\frac{\partial L}{\partial\left(t^* - t^+\right)} = 0 \qquad (4.64)$$

$$\frac{\partial L}{\partial\left(t^- - t^*\right)} = 0 \qquad (4.65)$$

$$U_1 \ge 0,\ U_2 \ge 0,\ U_3 \ge 0,\ U_4 \ge 0$$

$$U_1\left[-\frac{\eta\mu^2}{\eta+\mu-\eta\mu}\frac{N}{s} + \frac{\eta+\mu+\mu^2}{\eta+\mu-\eta\mu}\mu\left(t^- - t^*\right) - \eta\left(t^* - t^+\right)\right] = 0$$

$$U_2\left[-\theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-(1-\theta)\mu\left(t^- - t^*\right)+\eta\left(t^* - t^+\right)\right]=0$$

$$U_3\left[(\mu+\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}-\left(\theta\frac{\eta+\mu-\eta\mu}{\eta+\mu}+\frac{\mu^2+2\eta\mu}{\eta+\mu}\right)\eta\left(t^* - t^+\right)-\left(\frac{\eta+\mu+\mu^2}{\eta+\mu}\right)(\mu+\theta)\left(t^- - t^*\right)\right]=0$$

$$U_4\left\{\left(\mu+\theta-\frac{\eta\mu}{\eta+\mu}\theta\right)(2+\theta)\frac{\eta\mu}{\eta+\mu}\frac{N}{s}+2\left(\frac{\eta\mu}{\eta+\mu}\right)^2\theta\frac{N}{s}-\right.$$
$$\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\left(\mu+2\eta+\frac{\eta+\mu-\eta\mu}{\eta+\mu}\theta\right)\left(t^* - t^+\right)-$$
$$\left.\left[\left(\mu+\theta+\frac{\eta\mu}{\eta+\mu}(1-\theta)\right)\frac{\eta+\mu+\mu^2}{\eta+\mu}\theta-\frac{\eta\mu^2}{\eta+\mu}(1-\theta)\right]\left(t^- - t^*\right)\right\}=0$$

Because the feasible region is a convex set and the constraints are comprised of linear functions, the toll window obtained using Kuhn tucker conditions must be the optimal toll window. Let $\left(t^* - t^+\right)^{wop}$ and $\left(t^- - t^*\right)^{wop}$ denote the optimal toll window in $CHJE$. In the following, we will examine the optimality conditions of the boundary solutions and interior solutions.

We first consider the optimal toll window that satisfies $t_o^{w1}=t_o^{w2}$ (on CH), namely

$$-\frac{\eta+\mu+\mu^2}{\eta+\mu-\eta\mu}\mu\left(t^- - t^*\right)^{\text{opw}}+\eta\left(t^* - t^+\right)^{\text{opw}}=-\frac{\eta\mu^2}{\eta+\mu-\eta\mu}\frac{N}{s} \tag{4.66}$$

then it holds that $U_1\geq 0$, $U_2=0$, $U_3=0$ and $U_4=0$. From (4.64) and (4.65), it can be acquired that

$$\frac{\partial TC^{CHJE}}{\partial\left(t^* - t^+\right)}-\eta U_1=0 \tag{4.67}$$

$$\frac{\partial TC^{CHJE}}{\partial\left(t^- - t^*\right)}+\frac{\eta+\mu+\mu^2}{\eta+\mu-\eta\mu}\mu U_1=0 \tag{4.68}$$

Making use of (4.66), (4.67) and (4.68), it can be obtained that

$$\left(t^* - t^+\right)^{\text{opw}}=\frac{\mu\left(2\eta\mu+\eta+\mu^2+\mu\right)N}{2(\eta+\mu)^2(\mu+1)s} \tag{4.69}$$

$$\left(t^- - t^*\right)^{\text{opw}}=\frac{\eta\left(3\eta\mu+\eta+2\mu^2+\mu\right)N}{2(\eta+\mu)^2(\mu+1)s} \tag{4.70}$$

Plugging (4.69) and (4.70) into (4.60) and (4.62) then combining with (4.67) and (4.68) can solve out U_1. Since $U_1\geq 0$, from (4.68), it requires $\partial TC^{CHJE}/\partial\left(t^- - t^*\right)\leq 0$, which, from (4.62), is equivalent to

$$\frac{\partial t_o^{w2}}{\partial\left(t^- - t^*\right)} \geq \frac{\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}\left(t_o^{w2} - t^+\right)}{\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \frac{\eta\mu}{\eta+\mu}\left(t_o^{w2} - t^+\right) + \frac{\eta\mu}{\eta+\mu}\left(\frac{N}{s} + t^- - t_o^{w2}\right) - \eta\left(t^* - t^+\right)},$$

which, from (4.63), is equivalent to

$$\theta \geq \theta_b$$

where,

$$\theta_b = \frac{\mu}{2\mu+1} \tag{4.71}$$

θ_b represents the critical unit overtaking cost, above which the optimal tolling scheme should make no commuter have incentive to overtake. The total cost under such a tolling scheme is given by

$$TC^{\text{braking}} = \alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N\left(1 - \frac{1}{4}\frac{\eta+\mu^2+\mu}{\left(1+\mu\right)\left(\eta+\mu\right)}\right) \tag{4.72}$$

Now we consider the optimal toll window that satisfies $t_o^{w2} = t^-$ (on EC), namely

$$\left(1-\theta\right)\mu\left(t^- - t^*\right)^{\text{opw}} - \eta\left(t^* - t^+\right)^{\text{opw}} = -\theta\frac{\eta\mu}{\eta+\mu}\frac{N}{s} \tag{4.73}$$

then it holds that $U_2 \geq 0$, $U_1 = 0$, $U_3 = 0$ and $U_4 = 0$. From (4.64) and (4.65), it can be acquired that

$$\frac{\partial TC^{CHJE}}{\partial\left(t^* - t^+\right)} + \eta U_2 = 0 \tag{4.74}$$

$$\frac{\partial TC^{CHJE}}{\partial\left(t^- - t^*\right)} - \left(1-\theta\right)\mu U_2 = 0 \tag{4.75}$$

Making use of (4.73), (4.74) and (4.75), it can be obtained that

$$\left(t^* - t^+\right)^{\text{opw}} = \frac{\frac{\mu}{\eta+\mu}\theta + \frac{1}{2}\frac{\mu}{\eta}\left(1-\theta\right)}{1 + \frac{\mu}{\eta}\left(1-\theta\right)}\frac{N}{s} \tag{4.76}$$

$$\left(t^- - t^*\right)^{\text{opw}} = \frac{\frac{1}{2} - \frac{\mu}{\eta+\mu}\theta}{1 + \frac{\mu}{\eta}\left(1-\theta\right)}\frac{N}{s} \tag{4.77}$$

It is easy to see that (4.76) and (4.77) are same as (4.30) and (4.29). That is because, if $t_o^{w2} = t^-$, the bottleneck is fully utilized (i.e. the equilibrium profile with capacity waste at t^- becomes the profile without capacity waste). Plugging (4.76) and (4.77) into (4.60) and (4.62) then combining

(4.74) and (4.75) can solve out U_2 . Since $U_2 \geq 0$, from (4.75), it requires that $\partial TC^{CHJE} \big/ \partial\left(t^- - t^*\right) \geq 0$, which, from (4.62), is equivalent to

$$\frac{\partial t_o^{w2}}{\partial\left(t^- - t^*\right)} \leq \frac{\frac{1}{2}\frac{\eta\mu}{\eta+\mu}\frac{N}{s}}{\frac{3}{2}\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \eta\left(t^* - t^+\right)^{\text{opw}}} \tag{4.78}$$

where $\left(t^* - t^+\right)^{\text{opw}}$ is given by (4.76). From (4.63), (4.78) is equivalent to

$$\theta \leq \theta_o$$

where

$$\theta_o = \frac{\mu(\eta+\mu)}{2\mu^2 + \eta + \mu} \tag{4.79}$$

θ_o represents the critical unit overtaking cost, below which the optimal tolling scheme should make the bottleneck fully utilized. Plugging (4.76) and (4.77) into $t_o^{w3} \leq t_o^{w2}$ gives

$$\theta \leq \theta_E \tag{4.80}$$

where θ_E is given in Proposition 4.2. The smaller one of θ_o and θ_E determines the optimality condition. The criticalness of θ_o and θ_E depends on η and μ, and it is straight forward that $\theta_b < \theta_o$. The total cost under such a tolling scheme is given by

$$TC^{\text{overtaking}} = \alpha\frac{\eta\mu}{\eta+\mu}\frac{N}{s}N - \alpha s(1-\theta)\left[\frac{\eta\mu}{\eta+\mu}\frac{N}{s} - \mu\left(t^- - t^*\right)_{globe}^{EC}\right]\frac{1}{2}\frac{N}{s}$$

Now, we consider the optimal toll window that satisfies $t_o^{w2} = t^-$ (on EC) and $t_o^{w3} = t_o^{w2}$ (on EJ), i.e., at E, then it holds that $U_1 = 0$, $U_2 \geq 0$, $U_3 \geq 0$ and $U_4 = 0$, from (4.64) and (4.65), it requires

$$\frac{\partial TC^{CHJE}}{\partial\left(t^* - t^+\right)} + \eta U_2 - \left(\theta\frac{\eta+\mu-\eta\mu}{\eta+\mu} + \frac{\mu^2 + 2\eta\mu}{\eta+\mu}\right)\eta U_3 = 0 \tag{4.81}$$

$$\frac{\partial TC^{CHJE}}{\partial\left(t^- - t^*\right)} - (1-\theta)\mu U_2 - \left(\frac{\eta+\mu+\mu^2}{\eta+\mu}\right)(\mu+\theta)U_3 = 0 \tag{4.82}$$

As the optimal toll window satisfies $t_o^{w2} = t^-$ and $t_o^{w3} = t_o^{w2}$, it can be easily solved as $\left(t^* - t^+\right)^E$ and $\left(t^- - t^*\right)^E$. Plugging $\left(t^* - t^+\right)^E$ and $\left(t^- - t^*\right)^E$ into (4.81) and (4.82) can solve U_2 and U_3. The conditions $U_2 \geq 0$ is equivalent to $\theta \geq \theta_E$, and $U_3 \geq 0$ is equivalent to condition $\theta \leq \tilde{\theta}$ ($\tilde{\theta} < 1$), thus it holds $\theta_E \leq \theta \leq \tilde{\theta}$. $\tilde{\theta}$ is obtained by solving

$$\left[\mu^2 + 2\eta\mu + (\eta+\mu-\eta\mu)\theta\right]\frac{\partial TC^{CHJE}}{\partial\left(t^- - t^*\right)} \geq \frac{1}{\eta}\left(\eta+\mu+\mu^2\right)(\mu+\theta)\frac{\partial TC^{CHJE}}{\partial\left(t^* - t^+\right)}$$

Due to the form of $\tilde{\theta}$ is too complicated, thus omitted here. The total cost under such an optimal tolling scheme is given as

$$TC^{\text{overtaking}\prime} = \alpha N \frac{\eta\mu}{\eta+\mu}\left[\frac{N}{s} - \frac{(1+\mu)(1-\theta)}{1+\mu(2-\theta)}\left(t^{-}-t^{+}\right)^{E}\right]$$

Now we consider the interior solutions. If the optimal toll window satisfies $t_o^{w1} < t_o^{w2}$, $t_o^{w2} < t^{-}$, $t_o^{w3} < t_o^{w2}$ and $t_o^{globe} < t_o^{w2}$, then it holds that $U_1 = 0$, $U_2 = 0$, $U_3 = 0$ and $U_4 = 0$. From (4.64) and (4.65), it can be acquired that

$$\frac{\partial TC^{CHJE}}{\partial\left(t^{*}-t^{+}\right)} = 0 \tag{4.83}$$

$$\frac{\partial TC^{CHJE}}{\partial\left(t^{-}-t^{*}\right)} = 0 \tag{4.84}$$

Combining (4.83) and (4.84) can solve out the interior optimal toll window, the optimal toll window is given as

$$\left(t^{*}-t^{+}\right)^{opw} = \frac{\left(2\mu^3+2\eta\mu+2\mu^2+\eta+\mu\right)\theta^2-2\mu^3\theta+\eta\mu^2+\mu^3}{(\eta+\mu)^2(\mu+\theta)^2}\mu\frac{N}{s}$$

$$\left(t^{-}-t^{*}\right)^{opw} = -\frac{\left(-2\eta\mu^2-\eta^2+\eta\mu+2\mu^2+\eta+\mu\right)\theta^2+\left(\eta^2\mu+5\eta\mu^2+2\mu^3\right)\theta-2\eta^2\mu-3\eta\mu^2-\mu^3}{(\eta+\mu)^2(\mu+\theta)^2}\mu\frac{N}{s}$$

then plugging the solution into condition $t_o^{w1} < t_o^{w2}$ gives condition $\theta < \theta_b$, and $t_o^{w2} < t^{-}$ gives condition $\theta > \theta_o$, thus it holds $\theta_o < \theta < \theta_b$. Since $\theta_b < \theta_o$, according to Kuhn tucker condition, the interior solution cannot be the global optimal solution. It can be verified that no more feasible solution exists.

Before giving the global optimal toll window, the criticalness among θ_o, θ_E and θ_b needs to be analyzed. The criticalness among θ_o, θ_E and θ_b determines the optimality conditions obtained from Kuhn tucker condition. Comparing the three critical unit overtaking costs (θ_o, θ_E and θ_b) results in two lines in the (μ,η) two-dimensional space as shown in Figure 6. Specifically, $\eta = \frac{\mu^2-1}{\mu^2+1}\mu$ corresponds to $\theta_o = \theta_E$, and $\eta = \frac{\mu-1}{\mu+1}\mu$ corresponds to $\theta_b = \theta_E$. The criticalness among θ_o, θ_E and θ_b is as given in Figure 4.6.

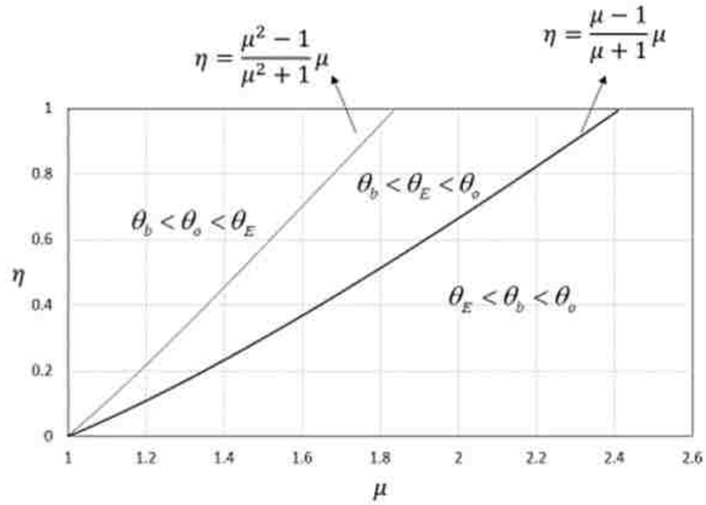

Figure 4.6. criticalness among θ_o , θ_E and θ_b

Setting $TC^{\text{braking}} = TC^{\text{overtaking}}$ gives a critical unit overtaking cost denoted by $\hat{\theta}$, where

$$\hat{\theta} = \frac{\mu + \frac{\mu + \eta}{\mu + 2\eta}}{\mu + 1 + \frac{\eta \mu^2}{(\mu + \eta)(\mu + 2\eta)}}$$

Comparing the optimal tolling schemes of equilibrium profiles shown in Figure 1 (i.e. no capacity waste exists, and all overtaking commuters have schedule late delay) and in Figure 4 (i.e. capacity waste only exists at t^- , and all overtaking commuters have schedule late delay), the unconstrained optimal tolling schemes (i.e., global optimal tolling scheme) is given in the following.

If $\eta \geq \frac{\mu^2 - 1}{\mu^2 + 1}\mu$, then it holds $\theta_b < \theta_o \leq \theta_E$, and θ_E does not need to be considered. According to Kuhn tucker condition, if $\theta < \theta_b$, the optimal toll window is given by (4.76) and (4.77), and the optimal toll level is given by (4.32). This corresponds to equilibrium profile of the overtaking model without capacity waste in Figure 4.1 with $t_o = t^-$, i.e., last overtaking commuter arrives at t^- and $t_y = t^+$, i.e., first toll payer arrives at t^+. If $\theta > \theta_o$, according to Kuhn tucker condition, the optimal toll window is given by (4.69) and (4.70), and the optimal toll level can be obtained by making use of (4.12), (4.42) and (4.46). This corresponds to the equilibrium profile of the overtaking model with capacity waste given in Figure 4.4 with $t_o = \tilde{t}_z = t_b$, i.e., no commuter has incentive to overtake and $t_y = t^+$, i.e., first toll payer arrives at t^+. The overtaking model just reduces to the braking model. If $\theta_b \leq \theta \leq \theta_o$, the optimal total cost is determined by the smaller one of TC^{braking} and $TC^{\text{overtaking}}$. If $\theta < \hat{\theta}$, then $TC^{\text{braking}} > TC^{\text{overtaking}}$. If $\theta > \hat{\theta}$, then $TC^{\text{braking}} < TC^{\text{overtaking}}$. From the Kuhn tucker condition, if $\theta < \theta_b$, it holds $TC^{\text{braking}} > TC^{\text{overtaking}}$, and if $\theta > \theta_o$, it holds $TC^{\text{braking}} < TC^{\text{overtaking}}$. Either $\hat{\theta} < \theta_b$ or $\hat{\theta} > \theta_o$ will contradict with the Kuhn tucker condition, thus it must hold $\theta_b \le \hat{\theta} \le \theta_o$, which shows that if $\theta \in \left[0, \hat{\theta}\right]$, the optimal total cost is $TC^{\text{overtaking}}$, and if $\theta \in \left(\hat{\theta}, 1\right)$, the optimal total cost is TC^{braking}. Condition $\theta_E \le \theta \le \tilde{\theta}$ does not need to be considered, since if $\tilde{\theta} < \theta_E$, this condition is invalid. If $\tilde{\theta} > \theta_E$, as $TC^{\text{overtaking}}$ and $TC^{\text{overtaking}\prime}$ both increases with k, at θ_E, it holds $TC^{\text{overtaking}} = TC^{\text{overtaking}\prime}$, and for $\theta \ge \theta_o$, it holds $TC^{\text{braking}} < TC^{\text{overtaking}}$, then $TC^{\text{overtaking}\prime}$ cannot be the optimal solution.

If $\frac{\mu-1}{\mu+1}\mu < \eta < \frac{\mu^2-1}{\mu^2+1}\mu$, it holds $\theta_b < \theta_E < \theta_o$, then θ_o does not need to be considered. For $\theta < \theta_b$, $TC^{\text{overtaking}}$ is the optimal solution. For $\theta_b \le \theta \le \theta_E$, the smaller one of TC^{braking} and $TC^{\text{overtaking}}$ is the optimal solution. If $\hat{\theta} \ge \theta_E$, then for $\theta \in [0, \theta_E]$, $TC^{\text{overtaking}}$ is the optimal solution. If $\hat{\theta} < \theta_E$, then for $\theta \in \left[0, \hat{\theta}\right]$, the optimal total cost is $TC^{\text{overtaking}}$, and for $\theta \in \left(\hat{\theta}, 1\right)$, the optimal total cost is TC^{braking}. If $\tilde{\theta} > \theta_E$, for $\theta > \tilde{\theta}$, TC^{braking} is the optimal solution. For $\theta_E < \theta \le \tilde{\theta}$, the smaller one of TC^{braking} and $TC^{\text{overtaking}\prime}$ is the optimal solution. If $\tilde{\theta} < \theta_E$, then condition $\theta_E \le \theta \le \tilde{\theta}$ is invalid, thus for $\theta \in (\theta_E, 1)$, TC^{braking} is the optimal solution.

If $\eta \le \frac{\mu-1}{\mu+1}\mu$, it holds $\theta_E \le \theta_b < \theta_o$, then θ_o does not need to be considered. According to the Kuhn tucker condition, it must hold $\tilde{\theta} \ge \theta_b$. For $\theta \le \theta_E$, the optimal total cost is $TC^{\text{overtaking}}$. For $\theta_E < \theta < \theta_b$, the optimal total cost is $TC^{\text{overtaking}\prime}$, the optimal toll window is given by (4.33) and (4.34), namely the toll window corresponding to E in Figure 4.5. The optimal toll level is given by (4.37). The equilibrium profile is the overtaking model without capacity waste given by Figure 4.1 with $t_o = t^-$ and $t_y = t^+$. For $\theta > \tilde{\theta}$, TC^{braking} is the optimal solution. For $\theta_b \le \theta \le \tilde{\theta}$, the smaller one of TC^{braking} and $TC^{\text{overtaking}\prime}$ is the optimal solution.

In summary, for small unit overtaking cost, the optimal tolling scheme is featured by the overtaking model's equilibrium profile where the first toll payer arrives at t^+, the last overtaking commuter arrives at t^-, and no capacity waste exists; for large unit overtaking cost, the optimal tolling scheme is to set the toll high enough to prevent users from overtaking, because from system cost perspective, it is better to make commuters braking instead of overtaking (i.e., the toll is pushed to critical level to make no commuter overtake). This is because reducing overtaking cost is more beneficial for system when unit overtaking cost is high. Although the wasted tolling period can be fully utilized through lowering the toll to make commuters overtake, the system cost will be increased by doing so.

4.7. Conclusion

In this chapter, a new “overtaking model” is developed to study the coarse tolling problem during morning peak hour. The overtaking behavior is featured by that the toll payers can overtake those braking commuters (toll non-payers) to take advantage of the tolling period to pay toll to pass the bottleneck. Such overtaking behavior can easily be observed in the morning commute period. The overtaking behavior is incurred a constant unit cost. The optimal tolling scheme is investigated based on equilibrium profile with capacity waste and without capacity waste. Unlike ADL and Laih model, in overtaking model, tolling scheme causing capacity waste could be better than tolling scheme without capacity waste. It is found that, the optimal tolling scheme is affected by the unit overtaking cost and one critical unit overtaking costs are defined. If the unit overtaking cost is small, the optimal tolling scheme is featured by that the first toll payer arrives at t^+, the last overtaking commuter arrives at t^-, and no capacity waste exists. If the unit overtaking cost is high, the optimal tolling scheme makes no commuter want to overtake, and only braking behavior exists. Thus, the overtaking model reduces to the braking model.

CHAPTER 5

TOLL ROAD PROFIT MAXIMIZATION UNDER MIXED TRAVEL BEHAVIORS OF CARS AND TRUCKS

The previous two chapters investigates the problem of bottleneck coarse tolling. In the scenario of bottleneck model, the purpose of road pricing is to minimize traveller' total system cost to reduce traffic congestion, which serves the government to maintain traffic volume. Meanwhile, road pricing can also serve as a way of infrastructure financing. For example, the government usually finance the construction and maintenance of a public road by charging tolls, so understanding how to achieve road profit maximization is necessary for government to effectively operate and manage roads. On the other hand, if government has budget constraint, it will authorize a private firm to construct and operate a road. Private toll roads are built by the firm to charge toll for revenue for a period until the franchise expires or the government debt is retired. After the debt is retired, the franchise of the road is turned back to government control. Profit maximization is typically the goal of a private firm, so for the government, understanding the profit-oriented behavior of the firm is necessary for choosing suitable regulations. Motivated by these considerations, we conduct this research of toll road profit maximization by considering mixed travel behavior of cars and trucks.

This chapter extended Guo and Xu (2016) by considering mixed travel behaviors of cars and trucks in that trucks choose routes deterministically (i.e., choose the route with the lowest actual cost), while cars follow stochastic user equilibrium in route choice (i.e., choose the route with the lowest perceived cost). This setup is realistic insofar as truckers, because of their commercial nature, are very sensitive to trip costs and tend to have much lower perception errors than passenger car drivers. We derive the equilibrium flow pattern under any combination of car-toll and truck-toll, and identify an integrated equilibrium range within which each road is used by both cars and trucks. We find that, depending on the per-truck pavement damage cost, the firm may take a car-strategy, a truck-strategy, or a car-truck mixed strategy. The perception error of car users, the VOT and traffic demand of cars and trucks are critical in determining different profit-maximizing strategy.

5.1. Model introduction

Consider two roads or links, link 1 and link 2, connecting one origin and one destination, where link 1 is a private toll road and link 2 is a free public road. There are two groups of vehicles, cars and trucks, indexed by $g = L, H$, where L and H represent cars (light vehicles) and trucks (heavy vehicles), respectively. Total travel demands of cars and trucks are fixed, and denoted v_L and v_H respectively. Let v_{gi} be the traffic volume of group g on link i, $g = L, H$, $i = 1,2$, where $v_{g1} + v_{g2} = v_g$, $g = L, H$. If n is the congestion PCE of trucks, then the total PCE units on link i

is $N_i = v_{Li} + nv_{Hi}$, $i=1,2$. Let $t_{gi}(N_i)$ be the travel time of group g on link i, which is assumed to be an increasing and continuously differentiable function of N_i. Let θ_L be the logit-model parameter that measures the perception error of car users . The perception error decreases with θ_L, so a large value of θ_L corresponds to a small perception error. Let P_L and P_H be the tolls charged on cars and trucks, respectively, and β_L and β_H be their values of time (VOT). The equilibrium conditions are

$$t_{H1}(N_1) + \frac{P_H}{\beta_H} \le t_{H2}(N_2), \text{ if } v_{H1} > 0 \tag{5.1}$$

$$t_{H1}(N_1) + \frac{P_H}{\beta_H} \ge t_{H2}(N_2), \text{ if } v_{H2} > 0 \tag{5.2}$$

$$v_{L1} = v_L \frac{e^{-\theta_L\left[t_{L1}(N_1)+\frac{P_L}{\beta_L}\right]}}{e^{-\theta_L\left[t_{L1}(N_1)+\frac{P_L}{\beta_L}\right]} + e^{-\theta_L t_{L2}(N_2)}} \tag{5.3}$$

$$v_{L2} = v_L \frac{e^{-\theta_L t_{L2}(N_2)}}{e^{-\theta_L\left[t_{L1}(N_1)+\frac{P_L}{\beta_L}\right]} + e^{-\theta_L t_{L2}(N_2)}} \tag{5.4}$$

$$v_{g1} + v_{g2} = v_g, \ g = L, H \tag{5.5}$$

$$v_{Li} + nv_{Hi} = N_i, \ i = 1,2 \tag{5.6}$$

$$v_{gi} \ge 0, \ g = L, H, \ i = 1,2 \tag{5.7}$$

Condition (5.1) stipulates that, if link 1 is used by trucks in equilibrium, then the generalized travel time (including toll) of link 1 for trucks must not exceed the travel time of link 2. Condition (5.2) is interpreted similarly. Condition (5.3) and (5.4) stipulate that, the car users follow logit-based stochastic user equilibrium. In this paper, logit model is adopted to represent the route choice probability of car users. As the model shows, no matter how high car toll is charged, there are always cars on link 1. For a given car toll, if θ_L increases to positive infinity, car users will have no perception error and follow deterministic user equilibrium, namely, choosing the route with shortest travel time. If θ_L is zero, namely, the perception error of cars users is infinitely large, there are always half amount of cars on link1 and the other half on link 2, regardless of the toll level.

The differences and connections between our model and that of Guo and Xu (2016) are as follows. Both models feature two parallel links connecting one OD pair, two user types with fixed total demands, and deterministic user equilibrium to describe truck users' route choice. The major difference is that, in this paper, car users follow stochastic user equilibrium, which means that, car users do not have to always choose the shortest path. The perception error of car users can play a significant role in their decision making of route choice, hence further affecting truck users' route choice decisions.

5.2 The integrated equilibrium

Following Arnott et al. (1992) and de Palma et al. (2008), we refer to the user equilibrium as an *integrated* equilibrium if both cars and trucks use each road, i.e., if $v_{gi} > 0$, $g = L, H$, $i = 1, 2$. An integrated equilibrium can occur only if conditions (5.1) and (5.2) both hold as equalities for truck users. Since car users follow stochastic user equilibrium, there are always cars on both link 1 and link 2. Let $N = v_L + nv_H$ denote the total PCE units in the network where $N_1 + N_2 = N$, and v_L and v_H are given. For any P_L and P_H, condition (5.8)-(5.10) uniquely determines the flow pattern of cars and trucks:

$$t_{H1}(N_1) + \frac{P_H}{\beta_H} = t_{H2}(N_2) \tag{5.8}$$

$$v_{L1} = v_L \frac{1}{1 + e^{\theta_L \left[t_{L1}(N_1) + \frac{P_L}{\beta_L} - t_{L2}(N_2) \right]}} \tag{5.9}$$

$$v_{H1} = \frac{N_1 - v_{L1}}{n} \tag{5.10}$$

Condition (5.8) is the integrated equilibrium condition for trucks. From (5.8), we can see that N_1 is a function of P_H. Given P_H, the total traffic flow on link 1 is uniquely determined. We can further obtain that

$$\frac{dN_1}{dP_H} = -\frac{1}{\beta_H \left[t'_{H1}(N_1) + t'_{H2}(N_2) \right]}$$

It is straightforward that dN_1/dP_H is less than zero, so N_1 is a decreasing function of P_H. From (5.9) and (5.10), it is easy to see that, v_{L1} and v_{H1} are both functions of P_L and P_H. If we denote

$$A = 1 + e^{\theta_L \left[t_{L1}(N_1) + \frac{P_L}{\beta_L} - t_{L2}(N_2) \right]}$$

from (5.9), we can obtain that

$$\frac{\partial v_{L1}}{\partial P_H} = v_L \frac{(1-A)\theta_L \left[t'_{L1}(N_1) + t'_{L2}(N_2) \right] \frac{dN_1}{dP_H}}{A^2}$$

$$\frac{\partial v_{L1}}{\partial P_L} = v_L \frac{(1-A)\theta_L \frac{1}{\beta_L}}{A^2}$$

From (5.10) we can obtain that

$$\frac{\partial v_{H1}}{\partial P_H} = \frac{1}{n}\frac{dN_1}{dP_H} + v_L \frac{(A-1)\theta_L \left[t'_{L1}(N_1) + t'_{L2}(N_2) \right] \frac{dN_1}{dP_H}}{nA^2}$$

$$\frac{\partial v_{H1}}{\partial P_L} = v_L \frac{(1-A)\theta_L \frac{1}{\beta_L}}{nA^2}$$

It is obvious that we have $\partial v_{L1}/\partial P_H > 0$, $\partial v_{L1}/\partial P_L < 0$, $\partial v_{H1}/\partial P_H < 0$ and $\partial v_{H1}/\partial P_L > 0$. These tell us that, given the car toll, the flow of cars increases with truck toll but the flow of trucks decreases with it. Given the truck toll, the flow of trucks increases with car toll but the flow of cars decreases with it.

5.3. The critical pricing curves for trucks $P_H^{\text{crit1}} = f(P_L)$ and $P_H^{\text{crit2}} = g(P_L)$

Under the integrated equilibrium, given the car toll, the flow of trucks decreases with truck toll. Since the maximum flow of trucks on link 1 is v_H and minimum flow is zero, there must exist two critical toll levels for trucks. For a given P_L, under condition (5.8)-(5.10), we define that, $v_{H1} = v_H$ under P_H^{crit1}, and $v_{H1} = 0$ under P_H^{crit2}. Obviously $P_H^{\text{crit1}} < P_H^{\text{crit2}}$. Namely, P_H^{crit1} is the truck toll that attracts all trucks to link 1 and P_H^{crit2} is the truck toll that drives away all trucks to link 2. We use v_{L1}^{crit1} and v_{L1}^{crit2} to denote the flow of cars on link1 under P_H^{crit1} and P_H^{crit2}, respectively. Based on (5.8) and (5.9), we can obtain the following equations:

$$t_{H1}\left(v_{L1}^{\text{crit1}} + nv_H\right) + \frac{P_H^{\text{crit1}}}{\beta_H} = t_{H2}\left(v_L - v_{L1}^{\text{crit1}}\right) \tag{5.11}$$

$$v_{L1}^{\text{crit1}} = v_L \frac{1}{1+e^{\theta_L\left[t_{L1}\left(v_{L1}^{\text{crit1}}+nv_H\right)+\frac{P_L}{\beta_L}-t_{L2}\left(v_L-v_{L1}^{\text{crit1}}\right)\right]}}$$

$$t_{H1}\left(v_{L1}^{\text{crit2}}\right) + \frac{P_H^{\text{crit2}}}{\beta_H} = t_{H2}\left(v_L - v_{L1}^{\text{crit2}} + nv_H\right) \tag{5.12}$$

$$v_{L1}^{\text{crit2}} = v_L \frac{1}{1+e^{\theta_L\left[t_{L1}\left(v_{L1}^{\text{crit2}}\right)+\frac{P_L}{\beta_L}-t_{L2}\left(v_L-v_{L1}^{\text{crit2}}+nv_H\right)\right]}}$$

We can see that, P_H^{cirt1} and P_H^{crit2} are both functions of P_L. We use $P_H^{\text{crit1}} = f(P_L)$ and $P_H^{\text{crit2}} = g(P_L)$ to represent the critical pricing curves for trucks. In the (P_L, P_H) two-dimensional space, a toll pair (P_L, P_H) where $P_H^{\text{crit1}} \le P_H \le P_H^{\text{crit2}}$ gives an integrated equilibrium. Any truck toll that falls out of this range gives a non-integrated equilibrium for truck users. If $P_H < P_H^{\text{crit1}}$, all trucks use link 1. If $P_H > P_H^{\text{crit2}}$, all trucks use link 2. If we denote

$$B = 1 + e^{\theta_L\left[t_{L1}\left(v_{L1}^{\text{crit1}}+nv_H\right)+\frac{P_L}{\beta_L}-t_{L2}\left(v_L-v_{L1}^{\text{crit1}}\right)\right]}$$

$$C = 1 + e^{\theta_L\left[t_{L1}\left(v_{L1}^{\text{crit2}}\right)+\frac{P_L}{\beta_L}-t_{L2}\left(v_L - v_{L1}^{\text{crit2}}+nv_H\right)\right]}$$

we can obtain that

$$\frac{dP_H^{\text{crit1}}}{dP_L} = \frac{v_L\left(B-1\right)\theta_L \dfrac{\beta_H}{\beta_L}\left[t'_{H1}\left(v_{L1}^{\text{crit1}}+nv_H\right)+t'_{H2}\left(v_L - v_{L1}^{\text{crit1}}\right)\right]}{B^2 + v_L\left(B-1\right)\theta_L\left[t'_{L1}\left(v_{L1}^{\text{crit1}}+nv_H\right)+t'_{L2}\left(v_L - v_{L1}^{\text{crit1}}\right)\right]}$$

$$\frac{dP_H^{\text{crit2}}}{dP_L} = \frac{v_L\left(C-1\right)\theta_L \dfrac{\beta_H}{\beta_L}\left[t'_{H1}\left(v_{L1}^{\text{crit2}}\right)+t'_{H2}\left(v_L - v_{L1}^{\text{crit2}}+nv_H\right)\right]}{C^2 + v_L\left(C-1\right)\theta_L\left[t'_{L1}\left(v_{L1}^{\text{crit2}}\right)+t'_{L2}\left(v_L - v_{L1}^{\text{crit2}}+nv_H\right)\right]}$$

We can see that, dP_H^{crit1}/dP_L and dP_H^{crit2}/dP_L are both greater than zero, so both $f\left(P_L\right)$ and $g\left(P_L\right)$ are increasing functions of P_L. Figure 5.1 shows these two critical pricing curves in the $\left(P_L, P_H\right)$ two-dimensional space.

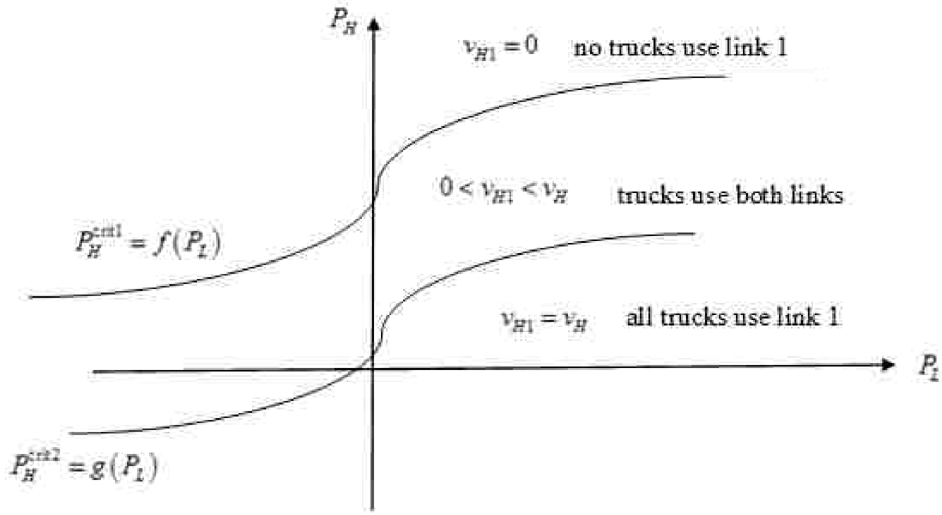

Figure 5.1 two critical pricing curves for trucks

As shown by Figure 5.1, the area between the two curves are the integrated equilibrium area. Any points outside gives a non-integrated equilibrium for truck users. From (5.11) and (5.12), it can be further seen that, $f\left(P_L\right)$ and $g\left(P_L\right)$ both have two asymptotic lines. For $f\left(P_L\right)$, we have

$$\beta_H\left[t_{H2}\left(0\right)-t_{H1}\left(N\right)\right] < f\left(P_L\right) < \beta_H\left[t_{H2}\left(v_L\right)-t_{H1}\left(nv_H\right)\right]$$

This means that, curve $P_H^{\text{crit1}} = f\left(P_L\right)$ is always between line $P_H = \beta_H\left[t_{H2}\left(0\right)-t_{H1}\left(N\right)\right]$ and line $P_H = \beta_H\left[t_{H2}\left(v_L\right)-t_{H1}\left(nv_H\right)\right]$. With P_L increasing, the value of P_H^{crit1} will be infinitely approaching $\beta_H\left[t_{H2}\left(v_L\right)-t_{H1}\left(nv_H\right)\right]$. With P_L decreasing, the value of P_H^{crit1} will be infinitely approaching $\beta_H\left[t_{H2}\left(0\right)-t_{H1}\left(N\right)\right]$. For $g\left(P_L\right)$, we have

$$\beta_H\left[t_{H2}\left(nv_H\right)-t_{H1}\left(v_L\right)\right] < g\left(P_L\right) < \beta_H\left[t_{H2}\left(N\right)-t_{H1}\left(0\right)\right]$$

This means that, curve $P_H^{\text{crit}2} = g(P_L)$ is always between line $P_H = \beta_H\left[t_{H2}(nv_H) - t_{H1}(v_L)\right]$ and line $P_H = \beta_H\left[t_{H2}(N) - t_{H1}(0)\right]$. With P_L increasing, the value of P_H^{crit2} will be infinitely approaching $\beta_H\left[t_{H2}(N) - t_{H1}(0)\right]$. With P_L decreasing, the value of P_H^{crit2} will be infinitely approaching $\beta_H\left[t_{H2}(nv_H) - t_{H1}(v_L)\right]$. The position of these four asymptotic lines depend on the value of v_L and nv_H, if $v_L \geq nv_H$, we have $\beta_H\left[t_{H2}(v_L) - t_{H1}(nv_H)\right] \geq \beta_H\left[t_{H2}(nv_H) - t_{H1}(v_L)\right]$, so the upper asymptotic line of $f(P_L)$ is above the lower asymptotic line of $g(P_L)$; if $v_L < nv_H$, we have $\beta_H\left[t_{H2}(v_L) - t_{H1}(nv_H)\right] < \beta_H\left[t_{H2}(nv_H) - t_{H1}(v_L)\right]$, so the upper asymptotic line of $f(P_L)$ is below the lower asymptotic line of $g(P_L)$. Since we assume that $t_{H2}(0) < t_{H1}(N)$ and $t_{H2}(N) > t_{H1}(0)$, the upper asymptotic line of $g(P_L)$ is always above the lower asymptotic line of lower asymptotic line $f(P_L)$.

5.4. Profit maximization by the private toll road

In this section we examine the profit-maximizing behavior of the firm that operates link 1. Let m_H be the pavement damage cost (road maintenance cost) caused by one truck using link 1 so that $m_H v_{H1}$ is the total pavement damage cost caused by trucks. Because one heavy truck causes thousands of times as much pavement damage as a passenger car (see, e.g., Holguín-Veras and Cetin, 2009, Table 6), we ignore the pavement damage cost caused by cars. For simplicity, we assume that pavement quality is kept constant by immediately repairing any damage. The firm's profit maximization problem is given by

$$\max_{P_L, P_H} \Pi(P_L, P_H) = P_L v_{L1} + (P_H - m_H) v_{H1}$$

subject to (v_{L1}, v_{H1}) satisfying equilibrium conditions (5.1)-(5.7) under (P_L, P_H)

Lemma 5.1: The maximum profit is attained by a toll pair (P_L, P_H) satisfying $P_H^{\text{crit}1} \leq P_H \leq P_H^{\text{crit}2}$, where $P_H^{\text{crit}1} = f(P_L)$ and $P_H^{\text{crit}2} = g(P_L)$.

Proof: For a toll pair (P_L, P_H) such that $P_H > P_H^{\text{crit}2}$, we have $v_{H1} = 0$ and $v_{L1} = v_{L1}^{\text{crit2}}$. Note that we also have $v_{H1} = 0$ and $v_{L1} = v_{L1}^{\text{crit2}}$ under $(P_L, P_H^{\text{crit2}})$, we readily have $\Pi(P_L, P_H) = \Pi(P_L, P_H^{\text{crit2}})$. For a toll pair (P_L, P_H) such that $P_H < P_H^{\text{crit1}}$, we have $v_{H1} = v_H$, $v_{L1} = v_{L1}^{\text{crit1}}$. Note that we also have $v_{H1} = v_H$ and $v_{L1} = v_{L1}^{\text{crit1}}$ under $(P_L, P_H^{\text{crit1}})$, we readily have $\Pi(P_L, P_H) < \Pi(P_L, P_H^{\text{crit1}})$. This completes the proof.

With Lemma 5.1, the profit-maximization problem can be converted to the following form:

$$\max_{P_L,P_H} \Pi\left(P_L,P_H\right) = P_L v_{L1} + \left(P_H - m_H\right) v_{H1}$$

subject to $\left(v_{L1}, v_{H1}\right)$ satisfying integrated equilibrium conditions (5.8)-(5.10) under $\left(P_L, P_H\right)$

$$P_H^{\text{cirt1}} \le P_H \le P_H^{\text{crit2}}$$

$$P_H^{\text{crit1}} = f\left(P_L\right),\ P_H^{\text{crit2}} = g\left(P_L\right)$$

The Kuhn-Tucker condition is used to solve this problem.

Let $L\left(U_1, U_2, P_L, P_H\right) = -P_L v_{L1} - \left(P_H - m_H\right) v_{H1} + U_1\left(P_H^{crit1} - P_H\right) + U_2\left(P_H - P_H^{crit2}\right)$, the optimal toll levels $\left(P_L^*, P_H^*\right)$ satisfies

$$\frac{\partial L}{\partial P_L} = 0,\ \frac{\partial L}{\partial P_H} = 0$$

$$P_H^{crit1} \le P_H^* \le P_H^{crit2}$$

$$U_1\left(P_H^{crit1} - P_H^*\right) = 0$$

$$U_2\left(P_H^* - P_H^{crit2}\right) = 0$$

$$U_1 \ge 0,\ U_2 \ge 0$$

Using the Kuhn-Tucker conditions to solve the problem, we obtain two critical levels of pavement damage cost, m_{H1} and m_{H2}, given by

$$\begin{aligned} m_{H1} = {} & \beta_H\left[t_{H2}\left(v_L - v_{L1}^{\text{crit1*}}\right) - t_{H1}\left(v_{L1}^{\text{crit1*}} + n v_H\right)\right] - v_{L1}^{\text{crit1*}} n \beta_L\left[t_{L1}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] \\ & - v_H n \beta_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{H2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] \end{aligned} \tag{5.13}$$

where $v_{L1}^{\text{crit1*}}$ solves

$$\theta_L = \frac{\ln\left(\dfrac{v_L - v_{L1}^{\text{crit1*}}}{v_{L1}^{\text{crit1*}}}\right) - \dfrac{v_L}{v_L - v_{L1}^{\text{crit1*}}}}{\left\{\begin{aligned} & t_{L1}\left(v_{L1}^{\text{crit1*}} + n v_H\right) - t_{L2}\left(v_L - v_{L1}^{\text{crit1*}}\right) + v_{L1}^{\text{crit1*}}\left[t_{L1}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] \\ & + v_H \frac{\beta_H}{\beta_L}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{H2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] \end{aligned}\right\}} \tag{5.14}$$

$$\begin{aligned} m_{H2} = {} & \beta_H\left[t_{H2}\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right) - t_{H1}\left(v_{L1}^{\text{crit2*}}\right)\right] \\ & - v_{L1}^{\text{crit2*}} n \beta_L\left[t_{L1}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right)\right] \end{aligned} \tag{5.15}$$

where $v_{L1}^{\text{crit2*}}$ solves

$$\theta_L = \frac{\ln\left(\dfrac{v_L - v_{L1}^{\text{crit2*}}}{v_{L1}^{\text{crit2*}}}\right) - \dfrac{v_L}{v_L - v_{L1}^{\text{crit2*}}}}{t_{L1}\left(v_{L1}^{\text{crit2*}}\right) - t_{L2}\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right) + v_{L1}^{\text{crit2*}}\left[t_{L1}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right)\right]}$$

(5.16)

The optimal solution is summarized in the following proposition.

Proposition 5.1. The optimal toll pair $\left(P_L^*, P_H^*\right)$ satisfies

(a) If $m_H \le m_{H1}$, we have $P_H^* = f\left(P_L^*\right)$ and

$$P_L^* = \frac{v_L \beta_L}{\left(v_L - v_{L1}^{\text{crit1*}}\right)\theta_L} + v_{L1}^{\text{crit1*}} \beta_L \left[t_{L1}{}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{L2}{}'\left(v_L - v_{L1}^{\text{crit1*}}\right) \right]$$
$$+ v_H \beta_H \left[t_{H1}{}'\left(v_{L1}^{\text{crit1*}} + n v_H\right) + t_{H2}{}'\left(v_L - v_{L1}^{\text{crit1*}}\right) \right]$$
$$P_H^* = \beta_H \left[t_{H2}\left(v_L - v_{L1}^{\text{crit1*}}\right) - t_{H1}\left(v_{L1}^{\text{crit1*}} + n v_H\right) \right]$$

(b) If $m_H \ge m_{H2}$, we have $P_H^* = g\left(P_L^*\right)$ and

$$P_L^* = \frac{v_L \beta_L}{\left(v_L - v_{L1}^{\text{crit2*}}\right)\theta_L} + v_{L1}^{\text{crit2*}} \beta_L \left[t_{L1}{}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}{}'\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right) \right]$$
$$P_H^* = \beta_H \left[t_{H2}\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right) - t_{H1}\left(v_{L1}^{\text{crit2*}}\right) \right]$$

(c) If $m_{H1} < m_H < m_{H2}$, we have $f\left(P_L^*\right) < P_H^* < g\left(P_L^*\right)$ and

$$P_L^* = \frac{v_L \beta_L}{\left(v_L - v_{L1}^*\right)\theta_L} + v_{L1}^* \beta_L \left[t_{L1}{}'\left(N_1^*\right) + t_{L2}{}'\left(N_2^*\right) \right] + v_{H1}^* \beta_H \left[t_{H1}{}'\left(N_1^*\right) + t_{H2}{}'\left(N_2^*\right) \right]$$
$$P_H^* = m_H + v_{L1}^* n \beta_L \left[t_{L1}{}'\left(N_1^*\right) + t_{L2}{}'\left(N_2^*\right) \right] + v_{H1}^* n \beta_H \left[t_{H1}{}'\left(N_1^*\right) + t_{H2}{}'\left(N_2^*\right) \right]$$
$$N_1^* = v_{L1}^* + n v_{H1}^*,\ N_2^* = N - N_1^*$$

$v_{L1}^{\text{crit1*}}$ is the flow of cars on link 1 at the optimal solution if the optimal $\left(P_L^*, P_H^*\right)$ is on curve $P_H^{\text{crit1}} = f\left(P_L\right)$. It is straightforward that we have $v_{H1} = v_H$. m_{H1} can be viewed as a truck-strategy taking criterion for the firm. If $m_H \le m_{H1}$, the firm should take truck-strategy. It is easy to see that, if link 1 is much longer than link 2, we have $t_{H2}\left(v_L - v_{L1}^{\text{crit1*}}\right) < t_{H1}\left(v_{L1}^{\text{crit1*}} + n v_H\right)$, which leads to $m_{H1} < 0$. Since in real world the pavement damage cost can never be negative, under this circumstance, the truck-strategy will never be used.

$v_{L1}^{\text{crit2*}}$ is the flow of cars on link 1 at the optimal solution if the optimal $\left(P_L^*, P_H^*\right)$ is on curve $P_H^{\text{crit2}} = g\left(P_L\right)$. It is straightforward that we have $v_{H1} = 0$. m_{H2} can be viewed as a car-strategy taking criterion for the firm. If $m_H \ge m_{H2}$, the firm should take car-strategy. It can be seen that, if link 1 is much longer than link 2, we have $t_{H2}\left(v_L - v_{L1}^{\text{crit2*}} + n v_H\right) < t_{H1}\left(v_{L1}^{\text{crit2*}}\right)$, which leads to $m_{H2} < 0$. Since in real world, the pavement damage cost is positive, under this circumstance, car strategy will always be used.

v_{L1}^{*} and v_{H1}^{*} are the flows of cars and trucks on link1 at the optimal solution if the optimal $\left(P_L^*, P_H^*\right)$ is between curve $P_H^{\text{crit1}} = f\left(P_L\right)$ and curve $P_H^{\text{crit2}} = g\left(P_L\right)$. It is straightforward that we have $v_{H1} > 0$ and $v_{L1} > 0$. If $m_{H1} < m_H < m_{H2}$, the firm should take a car-truck mixed strategy.

5.5. Impact of perception error on profit-maximizing strategy

In this section, we will analyze how the perception error of car users can impact the profit-maximizing strategy for the firm. As mentioned in the first section, the perception error decreases with θ_L. The profit-maximizing strategy is determined by m_{H1} and m_{H2}. From (5.14), with other parameters v_L, v_H, β_L and β_H given, $v_{L1}^{\text{crit1*}}$ can be seen as a function of θ_L, namely, given θ_L, we can solve $v_{L1}^{\text{crit1*}}$. From (5.13), m_{H1} can be seen as a function of $v_{L1}^{\text{crit1*}}$. With the value of $v_{L1}^{\text{crit1*}}$, the value of m_{H1} can be obtained, so m_{H1} is actually a function of θ_L. m_{H2} can be interpreted similarly.

In (5.14), denote

$$D = v_{L1}^{\text{crit1*}}\left[t_{L1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] + v_H\frac{\beta_H}{\beta_L}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{H2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right]$$

We assume D increases with $v_{L1}^{\text{crit1*}}$, i.e., $dD/d\,v_{L1}^{\text{crit1*}} > 0$. $dD/d\,v_{L1}^{\text{crit1*}}$ is obtained as:

$$\begin{aligned}\frac{dD}{dv_{L1}^{\text{crit1*}}} &= t_{L1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right) + v_{L1}^{\text{crit1*}}\left[t_{L1}''\left(v_{L1}^{\text{crit1*}} + nv_H\right) - t_{L2}''\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] \\ &\quad + v_H\frac{\beta_H}{\beta_L}\left[t_{H1}''\left(v_{L1}^{\text{crit1*}} + nv_H\right) - t_{H2}''\left(v_L - v_{L1}^{\text{crit1*}}\right)\right]\end{aligned}$$

In (5.16), denote

$$E = v_{L1}^{\text{crit2*}}\left[t_{L1}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right)\right]$$

Similarly, we assume E increases with $v_{L1}^{\text{crit2*}}$, i.e., $dE/d\,v_{L1}^{\text{crit2*}} > 0$. $dE/d\,v_{L1}^{\text{crit2*}}$ is obtained as:

$$\frac{dE}{dv_{L1}^{\text{crit2*}}} = t_{L1}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right) + v_{L1}^{\text{crit2*}}\left[t_{L1}''\left(v_{L1}^{\text{crit2*}}\right) - t_{L2}''\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right)\right]$$

The assumptions made for D and E hold if BPR link travel time function is adopted. With these assumptions, it is easy to obtain that $dm_{H1}/dv_{L1}^{\text{crit1*}} < 0$ and $dm_{H2}/dv_{L1}^{\text{crit2*}} < 0$. From (5.13)-(5.16), we can obtain the following proposition.

Proposition 5.2. As θ_L increases from zero to infinity,

(a) the firm is more likely to take either car-strategy or truck-strategy but less likely to take mixed-strategy if link 2 is long and truck-to-car VOT ratio is high.

(b) the firm is more likely to take car-strategy if link 2 is long and truck-to-car VOT ratio is low.

(c) the firm is more likely to take truck-strategy if link 2 is short.

Proof: In (5.14) and (5.16), denote

$$A' = t_{L1}\left(v_{L1}^{\text{crit1*}} + nv_H\right) - t_{L2}\left(v_L - v_{L1}^{\text{crit1*}}\right) + v_{L1}^{\text{crit1*}}\left[t_{L1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right]$$
$$+ v_H \frac{\beta_H}{\beta_L}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{H2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right]$$

$$B' = t_{L1}\left(v_{L1}^{\text{crit2*}}\right) - t_{L2}\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right) + v_{L1}^{\text{crit2*}}\left[t_{L1}'\left(v_{L1}^{\text{crit2*}}\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right)\right]$$

$$C' = \ln\left(\frac{v_L - v_{L1}^{\text{crit1*}}}{v_{L1}^{\text{crit1*}}}\right) - \frac{v_L}{v_L - v_{L1}^{\text{crit1*}}}$$

$$D' = \ln\left(\frac{v_L - v_{L1}^{\text{crit2*}}}{v_{L1}^{\text{crit2*}}}\right) - \frac{v_L}{v_L - v_{L1}^{\text{crit2*}}}$$

It can be seen that, A' is an increasing function of $v_{L1}^{\text{crit1*}}$, B' is an increasing function of $v_{L1}^{\text{crit2*}}$, C' is a decreasing function of $v_{L1}^{\text{crit1*}}$ and D' is a decreasing function of $v_{L1}^{\text{crit2*}}$. Since θ_L is greater than zero, A' and C' must have the same positive or negative sign, so do B' and D'. If $v_{L1}^{\text{crit1*}}$ or $v_{L1}^{\text{crit2*}}$ is approaching zero, the value of C' or D' is approaching plus infinity. If $v_{L1}^{\text{crit1*}}$ or $v_{L1}^{\text{crit2*}}$ is approaching v_L, the value of C' or D' is approaching minus infinity. We further assume $t_{L1}(0) < t_{L2}(N)$ and $t_{L1}(N) > t_{L2}(0)$, so the minimum value of B' is always less than zero and the maximum value of A' is always greater than zero. Based on the minimum value of A' and maximum value of B', we can obtain the sign of the numerator and denominator of θ_L. There are totally eight scenarios. The minimum value of A' and maximum value of B' can be acquired as

$$\min A' = t_{L1}(nv_H) - t_{L2}(v_L) + v_H \frac{\beta_H}{\beta_L}\left[t_{H1}'(nv_H) + t_{H2}'(v_L)\right]$$

$$\max B' = t_{L1}(v_L) - t_{L2}(nv_H) + v_L\left[t_{L1}'(v_L) + t_{L2}'(nv_H)\right]$$

Scenario one:

$\min A' > 0$, $\max B' < 0$

Scenario two:

$\min A' < 0$, $\max B' < 0$. In C'/A', with $v_{L1}^{\text{crit1*}}$ increasing from zero to v_L, A' equals zero first.

Scenario three:

$\min A' > 0$, $\max B' > 0$. In D'/B', with $v_{L1}^{\text{crit2*}}$ increasing from zero to v_L, D' equals zero first.

Scenario four:

$\min A' < 0$, $\max B' > 0$. In C'/A', with $v_{L1}^{\text{crit1*}}$ increasing from zero to v_L, A' equals zero first. In D'/B', with $v_{L1}^{\text{crit2*}}$ increasing from zero to v_L, D' equals zero first.

Scenario five:

$\min A' < 0$, $\max B' < 0$. In C'/A', with $v_{L1}^{\text{crit1*}}$ increasing from zero to v_L, C' equals zero first.

Scenario six:

$\min A' < 0$, $\max B' > 0$. In C'/A', with $v_{L1}^{\text{crit1}*}$ increasing from zero to v_L, C' equals zero first. In D'/B', with $v_{L1}^{\text{crit2}*}$ increasing from zero to v_L, D' equals zero first.

Scenario seven:

$\min A' > 0$, $\max B' > 0$. In D'/B', with $v_{L1}^{\text{crit2}*}$ increasing from zero to v_L, B' equals zero first.

Scenario eight:

$\min A' < 0$, $\max B' > 0$. In C'/A', with $v_{L1}^{\text{crit1}*}$ increasing from zero to v_L, A' equals zero first. In D'/B', with $v_{L1}^{\text{crit2}*}$ increasing from zero to v_L, B' equals zero first.

In scenario one to four, we have $A' > 0$ and $B' < 0$. With θ_L increasing, m_{H2} decreases and m_{H1} increases, meaning that the space of using car-strategy or truck-strategy is becoming bigger but the space of using car-truck mixed strategy is becoming smaller. In terms of link distance, $A' > 0$ and $B' < 0$ implies that link 2 is long and truck-to-car VOT ratio is high. Under this circumstance, as θ_L increases, the firm is more likely to take either car-strategy or truck-strategy but less likely to take mixed-strategy.

In scenario five and six, we have $A' < 0$ and $B' < 0$. With θ_L increasing, m_{H1} and m_{H2} both decrease, implying that the space of using car-strategy is becoming bigger but the space of using truck-strategy is becoming smaller. In terms of link distance, $A' < 0$ and $B' < 0$ implies that link 2 is long and truck-to-car VOT ratio is low. Under this circumstance, as θ_L increases, the firm is more likely to take car-strategy.

In scenario seven and eight, we have $A' > 0$ and $B' > 0$. With θ_L increasing, m_{H1} and m_{H2} both increase, implying that the space of using truck-strategy is becoming bigger but the space of using car-strategy is becoming smaller. In terms of link distance, $A' > 0$ and $B' > 0$ implies that link 2 is short. Under this circumstance, as θ_L increases, the firm is more likely to take truck-strategy. This completes the proof.

We can see that, the perception error plays an important role in firm's decision making of which strategy to choose. Under different circumstances, with different perception error, the firm's likelihood of taking each strategy is different.

5.6. Impact of β_L and β_H on profit-maximizing strategy

In this section, we will discuss how the VOTs of car and truck users affect the profit-maximizing strategy. It is straightforward that, if a traveler's VOT is large, she will be more time sensitive and thus more likely to choose the shorter path. In (5.14) and (5.16), given other parameters such as v_L, v_H and θ_L, $v_{L1}^{\text{crit1}*}$ can be seen as a function of β_L and β_H but $v_{L1}^{\text{crit2}*}$ is not related with either

β_L or β_H. Based on (5.13) and (5.15), m_{H1} and m_{H2} can both be seen as functions of β_L and β_H.

From (5.14) we can obtain that

$$\frac{\partial v_{L1}^{\text{crit1*}}}{\partial \beta_L}=\frac{\beta_L^{-2}v_H\beta_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]}{\frac{1}{\theta_L}\left[\frac{v_L}{\left(v_L-v_{L1}^{\text{crit1*}}\right)v_{L1}^{\text{crit1*}}}+\frac{v_L}{\left(v_L-v_{L1}^{\text{crit1*}}\right)^2}\right]+t_{L1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)+\frac{dD}{dv_{L1}^{\text{crit1*}}}}$$

It is obviously that we have $\partial v_{L1}^{\text{crit1*}}/\partial\beta_L>0$. From (5.13), we can acquire that

$$\frac{\partial m_{H1}}{\partial\beta_L}=-\frac{\partial v_{L1}^{\text{crit1*}}}{\partial\beta_L}\beta_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]$$
$$-nv_{L1}^{\text{crit1*}}\left[t_{L1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]-n\beta_L\frac{\partial v_{L1}^{\text{crit1*}}}{\partial\beta_L}\frac{dD}{dv_{L1}^{\text{crit1*}}}$$

which readily gives us $\partial m_{H1}/\partial\beta_L<0$. This tells us that, given β_H, m_{H1} decreases with β_L, which means the space of taking truck-strategy is decreasing. For the VOT of truck users, from (5.14), we can obtain that

$$\frac{\partial v_{L1}^{\text{crit1*}}}{\partial \beta_H}=\frac{-\frac{v_H}{\beta_L}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]}{\frac{1}{\theta_L}\left[\frac{v_L}{\left(v_L-v_{L1}^{\text{crit1*}}\right)v_{L1}^{\text{crit1*}}}+\frac{v_L}{\left(v_L-v_{L1}^{\text{crit1*}}\right)^2}\right]+t_{L1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)+\frac{dD}{dv_{L1}^{\text{crit1*}}}}$$

It is obviously that we have $\partial v_{L1}^{\text{crit1*}}/\partial\beta_H<0$. From (5.13), we can acquire that

$$\frac{\partial m_{H1}}{\partial\beta_H}=t_{H2}\left(v_L-v_{L1}^{\text{crit1*}}\right)-t_{H1}\left(v_{L1}^{\text{crit1*}}+nv_H\right)-nv_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]$$
$$-\beta_H\frac{\partial v_{L1}^{\text{crit1*}}}{\partial\beta_H}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]-n\beta_L\frac{\partial v_{L1}^{\text{crit1*}}}{\partial\beta_H}\frac{dD}{dv_{L1}^{\text{crit1*}}}$$

If it holds

$$t_{H2}\left(v_L-v_{L1}^{\text{crit1*}}\right)-t_{H1}\left(v_{L1}^{\text{crit1*}}+nv_H\right)-nv_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]<0$$

based on (5.13), we have $m_{H1}<0$, so truck-strategy will never be adopted. If $m_{H1}>0$, or namely, truck-strategy can be used, it always holds that

$$t_{H2}\left(v_L-v_{L1}^{\text{crit1*}}\right)-t_{H1}\left(v_{L1}^{\text{crit1*}}+nv_H\right)-nv_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]>0$$

which readily gives us $\partial m_{H1}/\partial\beta_H>0$. This tells us that given β_L, m_{H1} increases with β_H, which means the space of taking truck-strategy is increasing.

Since $v_{L1}^{\text{crit2*}}$ is not related with either β_L or β_H, we only need to consider m_{H2}. From (5.15), it is easy to obtain that

$$\frac{\partial m_{H2}}{\partial \beta_L} = -nv_{L1}^{\text{crit2*}}\left[t_{L1}'\left(v_{L1}^{\text{crit2*}}\right)+t_{L2}'\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right)\right]$$

$$\frac{\partial m_{H2}}{\partial \beta_H} = t_{H2}\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right) - t_{H1}\left(v_{L1}^{\text{crit2*}}\right)$$

It is straightforward that $\partial m_{H2}/\partial \beta_L < 0$, which means that, given β_H, m_{H2} decreases with β_L, so the space of taking car-strategy is increasing. If $t_{H2}\left(v_L - v_{L1}^{\text{crit2*}} + nv_H\right) < t_{H1}\left(v_{L1}^{\text{crit2*}}\right)$, $m_{H2} < 0$, so car-strategy will always be used. If $m_{H2} > 0$, we will have $\partial m_{H2}/\partial \beta_H > 0$, which means that given β_L, m_{H2} increases with β_H, so the space of taking car-strategy is decreasing. We readily have the following proposition:

Proposition 5.3. If $m_{H1} > 0$ and $m_{H2} > 0$,
(a) Given β_H, m_{H1} and m_{H2} both decrease with β_L.
(b) Given β_L, m_{H1} and m_{H2} both increase with β_H.

Under Proposition 5.3(a), the firm is more likely to take car-strategy. Under Proposition 5.3(b), the firm is more likely to take truck-strategy.

5.7. Impact of v_L on profit-maximizing strategy

In this section we will discuss how the proportion of car users impact the firm's profit-maximizing strategy. The total PCE units in the network N is fixed. We have $v_H = (N - v_L)/n$. From (5.14) and (5.16), with other parameters θ_L, β_L and β_H given, $v_{L1}^{\text{crit1*}}$ and $v_{L1}^{\text{crit2*}}$ can both be seen as functions of v_L. Similarly, m_{H1} and m_{H2} can also be seen as functions of v_L.

From (5.14), we can obtain that

$$\frac{dv_{L1}^{\text{crit1*}}}{dv_L} = \frac{\dfrac{dD}{dv_{L1}^{\text{crit1*}}} + \dfrac{1}{n}\dfrac{\beta_H}{\beta_L}\left[t_{H1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{H2}'\left(v_L - v_{L1}^{\text{crit1*}}\right)\right] + \dfrac{v_{L1}^{\text{crit1*}}}{\theta_L}\left[\dfrac{1}{\left(v_L - v_{L1}^{\text{crit1*}}\right)v_{L1}^{\text{crit1*}}} + \dfrac{1}{\left(v_L - v_{L1}^{\text{crit1*}}\right)^2}\right]}{\dfrac{dD}{dv_{L1}^{\text{crit1*}}} + t_{L1}'\left(v_{L1}^{\text{crit1*}} + nv_H\right) + t_{L2}'\left(v_L - v_{L1}^{\text{crit1*}}\right) + \dfrac{v_L}{\theta_L}\left[\dfrac{1}{\left(v_L - v_{L1}^{\text{crit1*}}\right)v_{L1}^{\text{crit1*}}} + \dfrac{1}{\left(v_L - v_{L1}^{\text{crit1*}}\right)^2}\right]}$$

It can be seen that $dv_{L1}^{\text{crit1*}}/dv_L > 0$.
From (5.13), we can obtain that

$$\frac{dm_{H1}}{dv_L}=\left\{n\beta_L\left[t_{L1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]-\beta_H\left[t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)\right]\right\}\left(\frac{dv_{L1}^{\text{crit1*}}}{dv_L}-1\right)$$

$$-\frac{n\beta_L}{\theta_L}\left[\frac{1}{\left(v_L-v_{L1}^{\text{crit1*}}\right)v_{L1}^{\text{crit1*}}}+\frac{1}{\left(v_L-v_{L1}^{\text{crit1*}}\right)^2}\right]\left(v_{L1}^{\text{crit1*}}-v_L\frac{dv_{L1}^{\text{crit1*}}}{dv_L}\right)$$

Let

$$k=\frac{t_{H1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{H2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)}{t_{L1}'\left(v_{L1}^{\text{crit1*}}+nv_H\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit1*}}\right)}$$

If $k\beta_H/\beta_L<n$, we have $dv_{L1}^{\text{crit1*}}/dv_L<1$; If $dv_{L1}^{\text{crit1*}}/dv_L>1$, we have $k\beta_H/\beta_L>n$. But we cannot obtain the sign of dm_{H1}/dv_L ,actually, through numerical analysis, in most practical cases, it holds $k\beta_H/\beta_L>n$ and

$$\frac{v_{L1}^{\text{crit1*}}}{v_L}<\frac{dv_{L1}^{\text{crit1*}}}{dv_L}<1$$

so we have $dm_{H1}/dv_L>0$

From (5.16), we can obtain that

$$\frac{dv_{L1}^{\text{crit2*}}}{dv_L}=\frac{\frac{v_{L1}^{\text{crit2*}}}{\theta_L}\left[\frac{1}{\left(v_L-v_{L1}^{\text{crit2*}}\right)v_{L1}^{\text{crit2*}}}+\frac{1}{\left(v_L-v_{L1}^{\text{crit2*}}\right)^2}\right]}{\frac{v_L}{\theta_L}\left[\frac{1}{\left(v_L-v_{L1}^{\text{crit2*}}\right)v_{L1}^{\text{crit2*}}}+\frac{1}{\left(v_L-v_{L1}^{\text{crit2*}}\right)^2}\right]+t_{L1}'\left(v_{L1}^{\text{crit2*}}\right)+t_{L2}'\left(v_L-v_{L1}^{\text{crit2*}}+nv_H\right)+\frac{dE}{dv_{L1}^{\text{crit2*}}}}$$

It can be seen that $dv_{L1}^{\text{crit2*}}/dv_L>0$. Based on (5.15), we can obtain

$$\frac{dm_{H2}}{dv_L}=-\beta_H\left[t_{H2}'\left(v_L-v_{L1}^{\text{crit2*}}+nv_H\right)+t_{H1}'\left(v_{L1}^{\text{crit2*}}\right)\right]\frac{dv_{L1}^{\text{crit2*}}}{dv_L}-n\beta_L\frac{dE}{dv_{L1}^{\text{crit2*}}}\frac{dv_{L1}^{\text{crit2*}}}{dv_L}$$

It can be seen that $dm_{H2}/dv_L<0$. m_{H2} decreases with v_L. We readily have the following proposition.

Proposition 5.4. For a fixed N, if the proportion of car users increases, regardless of perception error, the space of using car-strategy is increasing.

This means that, with car accounting for a relatively large portion of total traffic flow, the firm is more likely to take car-strategy.

In this chapter we used three subsections to analyze the impact of perception error, each group user's VOT and proportion of car users on the firm's profit-maximizing strategy. We find out that, under different circumstances, with different perception error, the firm's likelihood of taking each strategy is different. If car users' VOT is relatively large to the truck user, the likelihood of adopting car-strategy is high. If truck users' VOT is relatively large to the car user, the likelihood of adopting truck-strategy is high. If the perception error of car users is small, with truck users accounting for a relatively large portion of total traffic flow, the firm is more likely to take truck-strategy. If the proportion of car users is relatively large, the firm is more likely to take car-strategy, regardless of the perception error.

5.8. Conclusions

In this chapter, we studied the toll road profit maximization problem under mixed travel behaviors of cars and trucks. We considered that truck users follow deterministic user equilibrium and car users follow stochastic user equilibrium. This setup is realistic, as truck users are much more sensitive to travel cost than car users. We first introduced the mixed model by giving conditions of the mixed user equilibrium. Then by giving conditions of the integrated equilibrium, we defined two critical pricing curves for the truck users, which helps us outline the integrated equilibrium area. It is shown that the profit-maximization problem need only consider the integrated equilibrium range (including the two boundary critical curves). By using the Kuhn-Tucker condition to solve the profit maximization problem, we defined two critical pavement damage cost for trucks. Depending on the critical pavement damage cost, the firms can accordingly take car-strategy, truck-strategy or car-truck mixed strategy. Then we analyzed the impacts of the perception error, VOT of each group user and proportion of car users on the firm's profit maximizing strategies. We found that, as the perception error of car users decreases, the firm's likelihood of taking each strategy changes differently under different conditions of link length and truck-to-car VOT ratio. Regarding the impact of each group's VOT, we found that, the likelihood of the firm taking the car-strategy increases with the VOT of cars and decreases with the VOT of trucks, and the likelihood of the firm taking the truck-strategy decreases with the VOT of cars and increases with the VOT of trucks. We also found out that, if the proportion of car users is relatively large, the firm is more likely to take car-strategy, regardless of the perception error.

CHAPTER 6

MAJOR FINDINGS AND EXTENSIONS

This chapter summarizes the major contributions of this dissertation. Unsolved problems and possible extensions are also pointed out.

6.1. Major findings

This dissertation contributes to road pricing studies in the following aspects.

Due to public acceptance reason, for peak hour congestion pricing, it may be politically unacceptable to charge a toll price that is too high or charge a toll for a period that is too long. Motivated by this, we study bottleneck coarse tolling in a constrained optimization setup, where there is a maximum acceptable toll level and a maximum acceptable toll window length. Three widely used coarse tolling models are studied, the ADL, Laih and braking models. The basic user behavioral difference between these three models are: in the ADL model, toll non-payers form a mass arrival at the bottleneck following the last toll payer's arrival at the bottleneck; in the Laih model, a separated waiting facility is built aside of the bottleneck for toll non-payers to wait until the toll ends; in the braking model, toll non-payers can choose to defer their arrival at the bottleneck to avoid paying the toll. In all three models, we consider proportional user heterogeneity, and focus on the case that the unconstrained optimal toll level and toll window length exceed the maximum acceptable upper bounds.

We find that, in the ADL and the Laih models, the constrained optimal coarse tolling chooses the maximum acceptable toll level and toll window length, which is consistent with the traditional insight that, because the toll replaces the queuing delay and thereby reduces the total system cost, increasing the toll in both toll price and tolling period will improve the system efficiency. While this is not surprising, an important new insight regarding the ADL model is established: under constrained optimization, because it is impossible to eliminate the queues at both the starting and the ending moments of the tolling period, in the ADL model the priority is to start the tolling period as late as possible to eliminate the queue at the toll ending moment only. This insight is never reported in traditional unconstrained coarse tolling studies. We find that, if the toll window length constraint is too stringent (the upper bound is too small), then any toll price will make the total system cost greater than the no-toll equilibrium, and thus no toll should be charged. When the toll window length constraint is not too stringent, the optimal solution chooses the maximum acceptable toll window length, while the optimal toll price may be an interior solution (i.e., less than the maximum acceptable level).

Comparing the three models, one common result is that the constrained optimal tolling scheme in all three models chooses the maximum acceptable toll window length. It should be noted that this result is not trivial for the braking model with heterogeneous users, where the trade-off between toll payers and non-payers makes the total system cost nonlinear and non-monotonic in toll window length. This result holds in the braking model with heterogeneous users because we proved that, when toll price is charged at the optimal level, the net effect of increasing the toll window length is positive in improving the total system cost. The Laih model and the braking model have a common feature that the total system cost depends on the toll level and toll window length rather than the specific toll window position. Therefore, in these two models the constrained optimal tolling scheme has a range of toll window positions. By contrast, the ADL model requires the toll window to be positioned as late as possible to minimize the queuing delay of the mass arrival users. A unique feature of the braking model is that the constrained optimal toll price may be less than the maximum acceptable level. This is because, unlike the ADL and Laih models, in the braking model there is a trade-off between toll payers and non-payers, which, depending on the toll window length, may give an interior optimal toll price.

In summary of all three models, in designing bottleneck coarse tolling, when the unconstrained optimal solution exceeds the toll level and toll window length upper bounds, it is generally safe to push the toll window length to its upper bound. When the mass arrival behavior has to be considered, the specific position of the toll window matters, i.e., the later the better. When the braking behavior has to be considered, it may not be optimal to charge the maximum acceptable toll level.

Inspired by the behavioral difference of the ADL, the Laih and the braking model and to better capture the commuters' travel behavior, a new coarse tolling model "overtaking model" is developed to study the coarse tolling problem during morning peak hour. The overtaking behavior is featured by that the toll payers can overtake those braking commuters (toll non-payers) to take advantage of the tolling period to pay toll to pass the bottleneck. This would allow commuters to brake and in the meanwhile can make the bottleneck fully utilized during the tolling period, i.e., eliminate the somewhat unrealistic unused period in the braking model. Such overtaking behavior can easily be observed in the morning commute period. The overtaking model systematically combines the Laih model and the braking model together, capturing both of their properties. The overtaking behavior is incurred a constant unit cost. Specifically, the overtaking model reduces to the Laih model when the unit overtaking cost approaches zero, and reduces to the braking model when the unit overtaking cost is too high. The optimal tolling scheme is investigated based on equilibrium profile with capacity waste and without capacity waste. Unlike the ADL and the Laih model, in overtaking model, tolling scheme causing capacity waste could be better than tolling scheme without capacity waste. It is found that, the optimal tolling scheme is affected by the unit overtaking cost and one critical unit overtaking cost is defined. For small unit overtaking cost, the optimal tolling scheme is featured by the overtaking model's equilibrium profile where the first toll payer arrives at t^+, the last overtaking commuter arrives at t^-, and no capacity waste exists; for large unit overtaking cost, the optimal tolling scheme is to set the toll high enough to prevent users from overtaking, because from system cost perspective, it is better to make commuters braking instead of overtaking (i.e., the toll is pushed to critical level to make no commuter overtake). This is because reducing overtaking cost is more beneficial for system when unit overtaking cost is high. Although the wasted tolling period can be fully utilized through lowering the toll to make commuters overtake, the system cost will be increased by doing so.

In the aspect of infrastructure financing, we studied the toll road profit maximization problem under mixed travel behaviors of cars and trucks. Private provision of public roads (tunnels, bridges, etc.) is increasing around the world. Profit maximization is typically the goal of a private firm, so for the government, understanding the profit-oriented behavior of the firm is necessary for choosing suitable regulations. We considered that truck users follow deterministic user equilibrium and car users follow stochastic user equilibrium. This setup is realistic, as truck users are much more sensitive to travel cost than car users. We first introduced the mixed model by giving conditions of the mixed user equilibrium. Then by giving conditions of the integrated equilibrium, we defined two critical pricing curves for the truck users, which helps us outline the integrated equilibrium area. It is shown that the profit-maximization problem need only consider the integrated equilibrium range (including the two boundary critical curves). By using the Kuhn-Tucker condition to solve the profit maximization problem, we defined two critical pavement damage cost for trucks. Depending on the critical pavement damage cost, the firms can accordingly take car-strategy, truck-strategy or car-truck mixed strategy. Then we analyzed the impacts of the perception error, VOT of each group user and proportion of car users on the firm's profit maximizing strategies. We found that, as the perception error of car users decreases, the firm's likelihood of taking each strategy changes differently under different conditions of link length and truck-to-car VOT ratio. Regarding the impact of each group's VOT, we found that, the likelihood of the firm taking the car-strategy increases with the VOT of cars and decreases with the VOT of trucks, and the likelihood of the firm taking the truck-strategy decreases with the VOT of cars and increases with the VOT of trucks. We also found out that, if the proportion of car users is relatively large, the firm is more likely to take car-strategy, regardless of the perception error.

6.2. Future extensions

Almost every contribution of this dissertation is associated with new or old unsolved problems. Thus it is important to point out possible future extensions of the studies done in this dissertation.

An important extension worth studying is the constrained optimization for the overtaking model. How the constraints on toll level and toll window length affects the overtaking behavior is a promising topic to investigate. Another extension regarding the overtaking model is to consider heterogeneous VOT and investigate how the VOT distribution affects the equilibrium profile and optimal tolling scheme.

For the profit maximization problem, numerical analysis will be conducted to support the current findings and study the impact of link distance on the firm's strategy. To study link distance impact, the BPR function will be used in numerical analysis. Another extension is to model competition between two toll roads in the presence of cars and trucks. Each road is operated by an independent firm trying to maximize its own profit. A competitive Nash-Equilibrium will be derived to solve this problem. The competition between two toll roads can be applied to the two bridges connecting Windsor and Detroit. The profit maximizing strategy taken by each firm under competition could have important insights on the government's regulations imposed on toll roads.

REFERENCES

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of Urban Economics 27(1), 111-130.

Arnott, R., de Palma, A., Lindsey, R., 1992. Route choice with heterogeneous drivers and group-specific congestion costs. Regional Science and Urban Economics 22 (1), 71-102.

Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. The American Economic Review 83 (1), 161-179.

Arnott, R., de Palma, A., Lindsey, R., 1994. The welfare effects of congestion tolls with heterogeneous commuters. Journal of Transport Economics and Policy 28 (2), 139-161.

Arnott, R., Kraus, M., 1995. Financing capacity in the bottleneck model. Journal of Urban Economics 38 (3), 272–290.

Beaty, C., Burris, M., Geiselbrecht, T., 2012. Executive Report: Toll Roads, Toll Rates, and Driver Behavior. FHWA/TX-14/0-6737-1, Texas A&M Transportation Institute.

Canadian Shipper, 2002. Hwy 407 is no model for addressing future transportation needs: OTA. http://www.canadianshipper.com/transportation-and-logistics/hwy-407-is-no-model-for-addressing-future-transportation-needs-ota/1000003799/

Cervero, R.B., 1998. *The Transit Metropolis.* Island Press, Washington, D.C.

Cheng, D., Ishak, S., 2016. Maximizing toll revenue and level of service on managed lanes with a dynamic feedback-control toll pricing strategy. Canadian Journal of Civil Engineering 43(1), 18-27.

Chicago Tribune, 2015. New year brings 40% toll hike for truckers in Illinois. http://www.chicagotribune.com/news/ct-trucks-toll-increase-met-20141226-story.html

Chu, H.-C., Meyer, M.D., 2008. Screening process for identifying potential truck-only toll lanes in a metropolitan area: the Atlanta Georgia case. Transportation Research Record 2066, 79-89.

Chu, H.-C., Meyer, M.D., 2010. Methodology for assessing safety benefits of truck diversion from truck-only toll lanes to arterials. Transport Reviews 30(6), 717-731.

De Palma, A., Lindsey, R., 2000. Private toll roads competition under various ownership regimes. The Annals of Regional Science. 34(1), 13-35.

De Palma, A., Kilani, M., Lindsey, R., 2008.The merits of separating cars and trucks. Journal of Urban Economics 64(2), 340-361.

Fan Wei, 2015. Optimal congestion pricing toll design for revenue maximization: comprehensive numerical results and implications. Canadian Journal of Civil Engineering, 2015, Vol.42 (8), 544-551.

Geiselbrecht, T., Baker, T., Beaty, C., Wood, N., Chigoy, B., Overmyer, S., Prozzi, J., 2015. Incentives for Truck Use of SH 130 Final Report. PRC 14-23F, Texas A&M Transportation Institute.

Guo, X., Xu, D. 2016. Profit maximization by a private toll road with cars and trucks, Transportation Research Part B 91, 113-129.

Holguín-Veras, J., Cetin, M., 2009. Optimal tolls for multi-class traffic: Analytical formulations and policy implications. Transportation Research Part A 43(4), 445-467.

Holguín-Veras, J., Jara-Díaz, S., 1999. Optimal space allocation and pricing for priority service at container ports. Transport Research Part B 33 (2), 81-106.

Holguín-Veras, J., Sackey, D., Hussain, S., Ochieng, V., 2003. Economic and financial feasibility of truck toll lanes. Transportation Research Record 1833, 66-72.

Jia, Z., Wang, D. Z. W., Cai, X., 2016. Traffic managements for household travels in congested morning commute. Transportation Research Part E 91, 173-189.

Knight, F.H., 1924. Some fallacies in the interpretation of social costs. *Quarterly Journal of Economics* 38, 582-606.

Knockaert, J., Verhoef, E.T., Rouwendal, J., 2016. Bottleneck congestion: differentiating the coarse charge. Transportation Research Part B 83, 59-73.

Laih C.H. 1994 Queuing at a bottleneck with single-Step and multistep tolls. Transportation Research Part A 28,197-208.

Laih, C.H. 2004. "Effects of the Optimal Step Toll Scheme on Equilibrium Commuter Behavior." Applied Economics 36 (1), 59-81.

Lindsey, R., Verhoef, E.T., 2001. Traffic congestion and congestion pricing. In: Button, K.J., Hensher, D.A. (Eds.), Handbook of Transport Systems and Traffic Control. Elsevier Science, Oxford, 77-105.

Lindsey, R., van den Berg, V.A.C., Verhoef, E.T. 2012. "Step Tolling with Bottleneck Queuing Congestion." Journal of Urban Economics 72 (1), 46–59.

Liu, L.N., McDonald, J.F., 1998. Efficient congestion tolls in the presence of unpriced congestion: a peak and off-peak simulation model. *Journal of Urban Economics*, 44, 352-366.

Liu, W., Zhang, F., Yang, H. 2016. Modeling and managing morning commute with both household and individual travels. Transportation Research Part B (in press), http://dx.doi.org/10.1016/j.trb.2016.12.002.

Mills, G., 1995. Welfare and profit divergence for a tolled link in a road network. *Journal of Transport Economics and Policy,* 29, 137-146.

Nie, Y.M. 2015. A new tradable credit scheme for the morning commute problem. Networks and Spatial Economics 15 (3), 719–741.

Pigou, A.C., 1920. *The Economics of Welfare*. MacMillan, London.

Ren, H., Xue, Y., Long, J., Gao, Z. 2016. A single-step-toll equilibrium for the bottleneck model with dropped capacity. Transportmetrica B 4 (2), 92-110.

Samuel, P., Poole, R.W., Jr., Holguín-Veras, J., 2002. Toll truckways: a new path toward safer and more efficient freight transportation. Policy Study No. 294, Reason Public Policy Institute.

Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge, London and New York.

Swan, P.F., Belzer, M.H., 2010. Empirical evidence of toll road traffic diversion and implications for highway infrastructure privatization. Public Works Management & Policy 14(4), 351-373.

Tian, L., Yang, H., Huang, H., 2013. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. Transportation Research Part E 54, 1-13.

Tsekeris, T., Voß, S., 2009. Design and evaluation of road pricing: state-of-the-art and methodological advances. Netnomics 10(1), 5-52.

Van den Berg, V., Verhoef, E.T., 2011. Congestion tolling in the bottleneck model with heterogeneous values of time. Transportation Research Part B 45 (1), 60-78.

Van den Berg, V.A.C., 2014. Coarse tolling with heterogeneous preferences. Transportation Research Part B 64, 1-23.

Vickrey W.S. 1969. Congestion theory and transport investment. American Economic Review 59, 251-261.

Xiao, F., Yang, H., Han, D., 2007. Competition and efficiency of private toll roads. Transportation Research Part B. 41(3), 292-308.

Xiao, F., Qian, Z., Zhang, H.M., 2011. The morning commute problem with coarse toll and nonidentical commuters. Networks and Spatial Economics 11(2), 343-369.

Xiao, F., Shen, W., Zhang, H.M., 2012. The morning commute under flat toll and tactical waiting. Transportation Research Part B 46 (10), 1346-1359.

Xiao, F., Qian, Z., Zhang, H.M., 2013. Managing bottleneck with tradable credits. Transportation Research Part B 56, 1-14.

Yang, H., Huang, H., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transportation Research Part B 31(6), 425-440.

Yang, H., Huang, H.-J., 2005. Mathematical and Economic Theory of Road Pricing. Elsevier, Oxford.

Yang, L., Saigal, R., Zhou, H., 2012. Distance-based dynamic pricing strategy for managed toll lanes. Transportation Research Record: Journal of Transportation Research Board, 2283(10), 90–99.

Zhang, F., Liu, W., Wang, X., Yang, H., 2017. A new look at the morning commute with household shared-ride: How does school location play a role? Transportation Research Part E 103, 198-217.

Yang, H., Xiao, F., Huang, H., 2009. Private Road Competition and Equilibrium with Traffic Equilibrium Constraints. Journal of Advanced Transportation. 43(1), 21-45.

VITA AUCTORIS

NAME	Da Xu
PlACE OF BIRTH	Jinan, China
YEAR OF BIRTH	1988
EDUCATION	University of Windsor, Windsor, Ontario 2012-2013 M.M. University of Windsor, Windsor, Ontario 2014-2015 M.A.Sc. University of Windsor, Windsor, Ontario 2016-2019 Ph.D.