Efficient and Equitable Travel Demand Management Using Price and Quantity Controls

by

Siyu Chen

B.S., University of California, Berkeley (2017)
M.S., Massachusetts Institute of Technology (2019)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2022

Author ....
Department of Civil and Environmental Engineering
Jan 24, 2022

Certified by..
Moshe E. Ben-Akiva
Edmund K. Turner Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by..
Ravi Seshadri
Assistant Professor, Department of Technology, Management and Economics, DTU
Thesis Supervisor

Accepted by ....
Colette L. Heald
Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

Efficient and Equitable Travel Demand Management Using Price and Quantity Controls

by

Siyu Chen

Submitted to the Department of Civil and Environmental Engineering
on Jan 24, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Traffic congestion is a serious problem that imposes significant costs on the economy, environment, and society. Congestion pricing as a demand management instrument has been known to be a cost-effective approach to deal with congestion. However, the issue of equity remains one of the major challenges to the successful design, acceptance, and deployment of congestion pricing. Although refunding revenues in a personalized manner has the potential to improve its acceptance by being Pareto-improving, there is limited research on methodologies to do so.

An alternative approach to travel demand management termed tradable mobility credits (TMC) has been gaining attention recently. It is a type of quantity control which can avoid the flow of money from users to the regulator and has been shown to have better performance than pricing under demand and supply uncertainty. Despite these promises, several important questions remain with regard to the design and functioning of the market within the TMC schemes, an aspect critical to the effective operationalization of these schemes.

The objective of this thesis is to design the efficient, equitable and Pareto improving congestion tolling for both price and quantity controls. First, we develop a market design for TMC schemes that ensures TMC is used for mobility management and avoids undesirable behavior such as hoarding, frequent selling and speculation, excessive activity at boundary (of token expiration), and negotiation cost. The developed design considers all aspects of market including token allocation, expiration, transaction fee, price adjustment and market rules governing trading. In addition, a heuristic approach to model disaggregate selling behavior is developed and the resulting simple selling strategy is derived. The developed market design addresses a growing and imminent need to develop methodologies to realistically model TMC schemes that are suited for real-world deployments.

Second, we develop a bi-level optimization framework for personalized distribution to make congestion tolling (both price and quantity controls) efficient, equitable, and Pareto improving. The system optimization determines the toll policy with the objective to maximize social welfare while the user optimization can be formulated with different objectives (e.g. to achieve Pareto improvement or maximize social welfare) to determine an individual-specific distribution of revenue for pricing or mobility credits for TMC. The developed personalized congestion tolling is promising as it addresses the important issue of equity and has the potential to improve public acceptance.

The performance of the designed instruments are demonstrated via microsimulation in a daily commute context between a single origin-destination pair. The simulation experiments employ a day-to-day assignment framework wherein transportation demand is modeled using a logit-mixture model with the nonlinear income effects and supply is modeled using a standard bottleneck model. The evaluation framework includes four main categories: social welfare, distributional impacts, behavior change, and level of congestion.

Thesis Supervisor: Moshe E. Ben-Akiva
Title: Edmund K. Turner Professor of Civil and Environmental Engineering

Thesis Supervisor: Ravi Seshadri
Title: Assistant Professor, Department of Technology, Management and Economics, DTU

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my thesis co-advisors, Professor Moshe Ben-Akiva and Professor Ravi Seshadri. Moshe's unmatched expertise in the areas of transportation were invaluable. Not only he provided me invaluable suggestions on several research projects, but also he set very high standards for me and pushed me to do better.

I am sincerely grateful to Professor Ravi Seshadri for his patience and immense knowledge. He was always available whenever I had a question about my research or writing.

Besides my co-advisors, I would like to thank my committee members: Professor Jinhua Zhao for his insightful comments and encouragement, and Professor Carlos Lima De Azevedo for his continuous support since the beginning of my graduate school. I must also thank Professor Jimi Oke, and Dr. Arun Akkinepally for their great amount of assistance along my research journey.

I am very grateful to my friends and colleagues: Yifei Xie, Youssef Aboutaleb, Mazen Danaf, Renming Liu, Yu Jiang, Yundi Zhang, Peiyu Jing, Kexin Chen, Emma Desoto, Haizheng Zhang, Xiang Song. I feel fortunate to study and work with them. I got inspired by them from time to time.

I also had great pleasure of taking multiple useful courses with amazing professors at MIT. Their knowledge, expertise and wisdom were valuable to me.

This dissertation would not have been possible without funding from the U.S. National Science Foundation (ID:CMMI-1917891).

Last but not least, I am deeply indebted to my girlfriend for all her love and encouragement, and my parents for their continuous support and endless love.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Congestion pricing (price control) is a well-researched mechanism to deal with traffic congestion by charging a toll to internalize congestion externalities. However, it is often perceived as a flat tax and receives low public acceptance. It is natural to think about whether and how revenues from congestion pricing can be used to benefit "losers" and improve its acceptance. In contrast, another variant of congestion tolling, termed tradable mobility credits has been gaining attention recently. It is a type of quantity control which can avoid the flow of money from users to the regulator and has been shown to have better performance than pricing under demand and supply uncertainty. However, several key elements of such schemes (revenue refunding and tradable mobility credits) still need more investigation for operationalizing them.

We start by introducing the background of these two types of controls and the motivation for this research. Then, we describe the research objectives and approach. Next, key contributions of this research are explained, and the outline of the thesis is described.

1.1 Background and Motivation

Traffic congestion is a serious problem that imposes significant costs on the economy, environment and society. For example, [Schrank et al., 2015] reported that peak hour trips took 35% more time on average than non-peak hour trips in 2016 compared to

20% in 2010 in United States. Also, it was found that congestion caused an extra 3.1 billion gallons of fuel consumption in U.S. [Global, 2018] finds that Americans lose 97 hours per year due to congestion, equivalent to 1348 USD. It is expected that the world's population residing in urban areas will increase from 50% currently to 70% by 2050 [United Nations Department of Economical and Social Affairs (DESA), 2018], which will lead to the increased travel demand and more severe congestion.

Treating road space as a scarce resource, there are three perspectives to deal with congestion. The first is to simply increase road capacity. However, it is less attractive and often difficult to implement because of financial, spatial, and environmental constraints. It has also been shown to be self-defeating because the increased capacity will be absorbed quickly by the induced demand [Goodwin, 1996, Goodwin and Noland, 2003, Duranton and Turner, 2011]. The second is to reduce demand (e.g. car ownership or usage) but it suppresses activities and hinders economic growth. The third is transportation system management [Bull et al., 2003, Kuhn et al., 2017], including traffic management (e.g. reversible lanes, signal control, ramp metering) and demand management (e.g. congestion pricing).

Congestion pricing as a demand management instrument has been known to be a cost-effective approach to deal with congestion (e.g. [Lindsney and Verhoef, 2001]). It was first introduced by [Pigou, 1920] as a tax to internalize the costs of a negative externality. It can influence travelers to alter their travel decisions on whether to travel, departure time, mode and destination choice [Saleh and Sammer, 2016]. Successful applications, such as Singapore's Electronic Road Pricing Scheme (ERP), London's Congestion Charge (CC), and Stockholm's Congestion Tax have highlighted the effectiveness of congestion-pricing schemes.

However, the issue of equity remains one of several challenges to the successful design, acceptance and deployment of congestion pricing as evident in the failures of previous plans in the Greater Manchester, Edinburgh and New York City. The out-of-pocket charges hamper low-income users from using road facilities and make road usage a privilege of high-income users [Lindsney and Verhoef, 2001, Gu et al., 2018].

[Jaensirisak et al., 2003] find that congestion pricing could be more acceptable if it increases everyone's benefit (so called Pareto-improving) besides the net social benefit. It is natural to consider how pricing revenue can be used to benefit users. In the literature, some studies have studied refunding revenue uniformly to users but this is not guaranteed to be Pareto-improving with realistic assumptions like heterogeneity (e.g. [Small, 1992, Arnott et al., 1994]). To the best of our knowledge, there has been no study on refunding revenue in a personalized way, although it has the potential to allocate revenue more efficiently and equitably. Moreover, the analysis of impacts of congestion tolling is limited with unrealistic assumptions (e.g. homogeneity, inelastic demand, constant income effect).

An alternative approach of travel demand management that has received increasing attention in the transportation domain in recent years is quantity control – in particular, tradable mobility credit (TMC) schemes (e.g. [Fan and Jiang, 2013, Grant-Muller and Xu, 2014, Dogterom et al., 2017].)

Within a TMC system, a regulator provides an initial endowment of mobility credits or tokens to all potential travelers. In order to use the road network or transportation system, users need to spend a certain number of tokens (i.e., tariff) that could vary with the attributes or performance of the specific mobility alternative used. The tokens can be bought and sold in a market that is monitored by the regulator at a price determined by the token demand and supply.

In principle, TMC schemes are appealing since they offer a means of directly controlling quantity (important when the elasticity of demand to prices in the short term may be low), they are revenue neutral in that there is no transfer of money to the regulator, they improve equity even with uniform token allocation, and they are shown to be more efficient than price control under uncertainty.

Despite these promises, several important questions remain with regard to the design and functioning of the market within TMC schemes, an aspect critical to the effective operationalization of these schemes. For instance, how should the allocation and expiration of tokens be designed? What rules should govern trading behavior in the market so as to avoid undesirable speculation and trading (see [Brands et al., 2020]

Table 1.1: Research approach

	Uniform (U)	Personalized (I)
Pricing (P)	PU	PI
TMC (M)	MU	MI

for more on this), and yet ensure efficiency and revenue neutrality? How should the regulator intervene in the market in the presence of special or non-recurrent events? What is the role and impact of transaction fees? Despite the large body of literature on TMCs, issues of market design, market dynamics and behavior of individuals in the market has received relatively little attention despite being critical to the successful real-world deployment of a TMC scheme.

1.2 Research Approach and Contributions

The objective of this thesis is to design efficient, equitable and Pareto improving congestion tolling for both price and quantity controls. The research approach consists of two dimensions: mechanism and distribution. For mechanism, we consider both pricing and tradable mobility credits (TMC); while for distribution, it can be either uniform or personalized. This leads to four instruments as shown in Table 1.1.

Each instrument is labeled by two characters, for which the first character represents mechanism and the second one represents distribution. For pricing, it is revenue to be distributed (more precisely re-distributed) while it is tokens (to be allocated) for TMC. In this study, besides these four instruments, the baseline considered is a *no toll* scenario denoted as NT and the traditional congestion pricing without re-distribution of toll revenues, which is denoted as $P-$.

This thesis has two major contributions as follows:

1. Develop a market design for TMC schemes that ensures TMC is used for mobility management and avoids undesirable behavior such as hoarding, frequent selling and speculation, excessive activity at boundary (of token expiration), and negotiation cost.

2. Develop a bi-level optimization framework for personalized distribution to make congestion tolling (both price and quantity controls) efficient, equitable, and Pareto improving.

The performance of the designed instruments are demonstrated via microsimulation in a daily commute context. Under congestion tolling, travelers are subject to a toll profile in units of dollars (for pricing) or tokens (for TMC). Taking advantage of microsimulation, we can examine agent-level travel and market behavior and the distributional impacts of the instruments. Heterogeneity, nonlinear income effects and user learning are considered to make our analysis realistic. A detailed discussion of the simulation framework can be found in Chapter 5.

The framework of pricing without distribution is illustrated in Figure 1-1. Users' day-to-day mobility choices considering traffic conditions and a toll profile are simulated in the network. User benefits and regulator revenue at equilibrium (convergence of the day-to-day dynamic process) are used to calculate social welfare and other metrics. The toll profile is optimized to maximize social welfare.

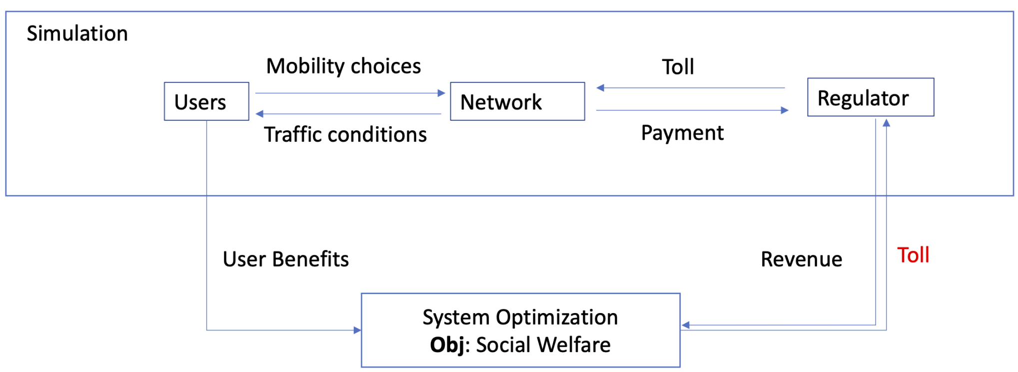


Figure 1-1: Illustration of pricing without distribution

For pricing with uniform distribution, the only change is to have the available regulator revenue distributed equally to users, while for personalized distribution, the framework is illustrated in Figure 1-2. A bi-level optimization formulation is adopted to determine the toll and personalized refunds. The system optimization determines the toll with the objective of maximizing social welfare; the user optimization determines personalized refunds with the objective of benefiting all users.

They are interdependent such that user optimization depends on system optimization policies and user behavior also affects system optimization. This bi-level optimization is an application of a general framework of online analytics for transportation system management termed Tri-POP, which combines prediction, optimization and personalization (POP) [Atasoy et al., 2020]. A detailed discussion of this can be found in Chapter 6.

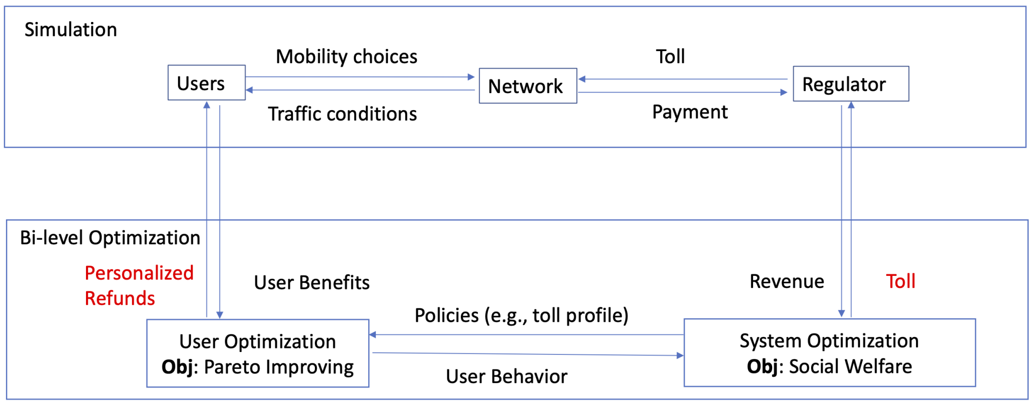


Figure 1-2: Illustration of pricing with personalized distribution

Regarding TMC, the framework is illustrated in Figure 1-3. Travelers now are subject to a toll charged in tokens instead of dollars. They receive a token allocation from the regulator which can be either uniform or personalized. In order to buy or sell tokens, they can send requests to a market in which all requests are satisfied by the regulator. Such a setting can avoid negotiation costs and information acquisition costs compared to peer-to-peer trading or auction markets [Brands et al., 2020]. The token price is adjusted based on demand and supply of tokens. Details of the market design and market behavior are discussed in Chapter 4.

1.3 Thesis Outline

The rest of the thesis has six chapters and is organized as follows: Chapter 2 reviews the relevant literature about congestion pricing and TMC in detail; Chapter 3 discusses a detailed market design for TMC schemes to manage travel demand; Chapter 4 introduces the simulation framework for numerical demonstration and toll profile

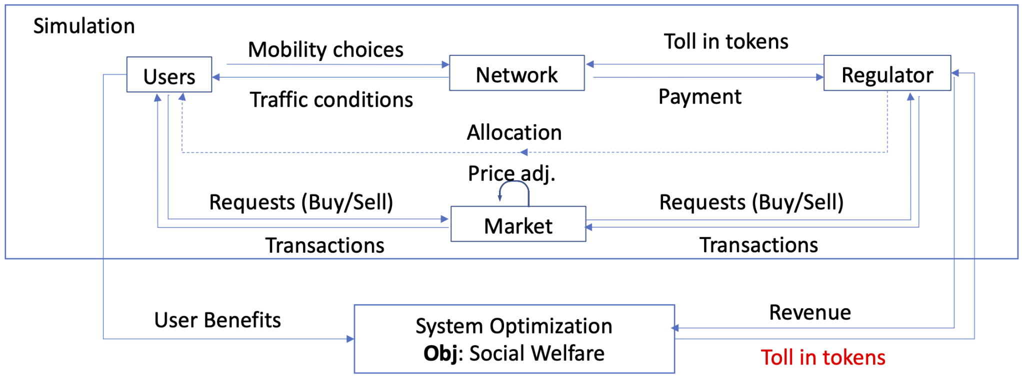


Figure 1-3: Illustration of tradable mobility credits

optimization formulations; Chapter 5 discusses the methodology for personalized distribution; Chapter 6 examines the designed instruments via numerical experiments; Chapter 7 summarizes contributions, findings, as well as future research directions.

Chapter 2

Literature Review

Congestion pricing has a long history since Pigou developed the initial concept in 1920. There is a vast body of literature on it. The review in this chapter focuses on congestion pricing design and its public acceptance. Recently, tradable mobility credits have received increasing attention leveraging on its successful application in other fields. It has the potential to address some drawbacks of congestion pricing, such as equity issues. This chapter reviews its design, mathematical modeling and economic properties.

2.1 Congestion pricing

2.1.1 First-best Pricing

The introduction of congestion pricing can be traced back to the seminal work by [Pigou, 1920] and [Knight, 1924]. Assume a single road network with homogeneous travelers and elastic demand. Let v denote the traffic flow, $c(v)$ denote the travel cost. The marginal social cost $MC(v)$ is the additional cost of having one more traveler to the road, which includes the additional cost on all other travelers already on the road and her own cost. It can be written as

$$MC(v) = \frac{\partial c(v) \cdot v}{\partial v} = c(v) + v \cdot \frac{\partial c(v)}{\partial v} \tag{2.1}$$

At equilibrium, marginal benefit (marginal willingness-to-pay) is equal to average cost $c(v)$ without tolling, which is less than marginal social cost $MC(v)$. Therefore, the optimal toll to have social welfare maximized should be equal to $v \cdot \frac{\partial c(v)}{\partial v}$, which is the difference between marginal social cost and average cost. It is also known as "Pigouvian toll".

Marginal cost pricing can be extended to a network. For an unpriced network, a user equilibrium is reached when all used routes for an OD pair have the same travel time equal to marginal benefits for that OD pair, and there are no unused routes with lower travel time. These user equilibrium conditions are also referred to as Wardrop's first principle [Wardrop, 1952]. However, the optimal flow pattern, referred to as system optimum, is different from the user equilibrium flow pattern. The conditions of system optimum are referred to as Wardrop's second principle such that all used routes for an OD pair have the same marginal cost equal to the marginal benefits for that OD pair, and there are no unused routes with lower marginal costs. It can be shown that for every link l, marginal cost pricing with toll as $V_l \cdot \frac{\partial c(v_l)}{\partial v_l}$ achieves the system optimum. But marginal cost pricing is not necessarily the only pricing scheme to achieve system optimum. [Hearn and Ramana, 1998] pointed out that the system optimum can be achieved as long as the sum of link tolls over each route equal to the sum of marginal external costs of all routes' links.

Another stream of research on congestion pricing is from a dynamic perspective. It starts with the basic bottleneck model proposed by [Vickrey, 1969]. In Vickrey's basic bottleneck model, congestion is modeled by a single bottleneck with a finite capacity s. The first in first out (FIFO) queue starts forming once flow rate is equal to the capacity and queuing time is proportional to queue length. Free flow travel time is assumed to be 0 for simplicity. It only considers homogeneous travelers: namely, the value of time α, the value of schedule delay early β and value of schedule delay late γ, and the desired arrival time t^* are identical for all travelers. Also, the total demand Q is inelastic. The generalized price of a user i's trip for arrival at t can be written as:

$$p_i(t) = \alpha T(t) + \tau(t) + \max(0, \beta(t^* - t)) + \max(0, \gamma(t - t^*)) \tag{2.2}$$

where $T(t)$ denotes travel time for arrival at t; $\tau(t)$ denotes a toll defined in terms of arrival time t; $\max(0, \beta(t^* - t))$ denotes schedule delay early cost and $\max(0, \gamma(t - t^*))$ denotes schedule delay late cost.

For the no toll equilibrium, since every user should bear the same generalized cost and the duration of the peak is $\frac{Q}{s}$, the equilibrium generalized cost can be easily calculated as:

$$p = \delta \frac{Q}{s} \quad \text{with} \quad \delta = \frac{\beta\gamma}{\beta + \gamma} \tag{2.3}$$

The departure rate of no toll equilibrium is equal to $\frac{s\alpha}{\alpha - \beta}$ for early arrivals and $\frac{s\alpha}{\alpha + \gamma}$ for late arrivals. The former expression requires $\alpha > \beta > 0$, which is consistent with empirical data.

It can be shown that a triangular toll schedule that at each instant is equal to the value of queuing delay in the no toll equilibrium can eliminate queuing delay completely (i.e. departure rate is equal to s). The generalized cost for every traveler is still the same as the no toll equilibrium since queuing delay is replaced fully by a toll. More details about the bottleneck model can be found in [Small, 2015, Li et al., 2020].

2.1.2 Second-best Pricing

Research reviewed in the previous section on first-best pricing provide important insights, but because of constraints and distorted markets, second-best pricing analysis is more relevant for realistic applications. Five main cases are summarized by [Lindsey and Verhoef, 2001, Small et al., 2007]: pricing in networks, pricing by time-of-day, heterogeneity of users, demand and supply uncertainty and interactions with the rest of the economy.

There are many studies about the case where not every congested link is tolled because of political constraints, acceptability constraints and economic constraints. The simplest case is the classic two-route problem with one route allowed to have a toll while the other must remain untolled. When users are homogeneous in all aspects except willingness to pay and congestion is static, the optimal toll can be solved by maximizing social surplus under equilibrium conditions. [Verhoef and Small, 2004] applied simulation to study the change of relative efficiency of second-best pricing compared to first-best pricing by varying the relative size of priced capacity from 0 to 1. They found that the welfare losses of second-best pricing from spillovers on untolled capacity are substantial unless a significant portion of capacity is priced. Considering departure-time adjustments, dynamic models (bottlenecks) of the two-route problem estimate the relative welfare gains from second-best pricing greater than those of static models [Braid, 1996, De Palma and Lindsey, 2000].

For more general networks, the toll design problems are usually formulated as bi-level optimization problems [Yan and Lam, 1996, Yang and Bell, 1997, Ferrari, 1999, Verhoef, 2002a, Verhoef, 2002b]. A bi-level optimization problem contains two inter-dependent problems, where upper-level problem is to optimize some system performance measure based on lower-level user responses; lower-level problem is to optimize users' objectives subject to toll solution from upper-level. Common upper-level objectives include social benefits [Yan and Lam, 1996, Ferrari, 1999, Verhoef, 2002b], toll revenue [Dial, 1999], or multiple objectives such as congestion, equity, and emissions [Wang et al., 2014b]. For the lower-level, it is often formulated as an user equilibrium problem to minimize generalized travel cost (travel time and toll) or stochastic user equilibrium problem to minimize perceived travel cost [Liu et al., 2014].

Various algorithms have been developed to solve the formulated problems, including sensitivity analysis based approaches [Yan and Lam, 1996], genetic algorithms based approaches [Shepherd and Sumalee, 2004], heuristic methods using Lagrangian multipliers [Verhoef, 2002b, Verhoef, 2002a] and manifold suboptimization [Lawphongpanich and Yin, 2010].

A second case is when the toll cannot be fully time varying. The simplest case is to have a single fixed ("flat" or "uniform") toll throughout the peak period. A more sophisticated case is a "step" or coarse toll with one or more non-zero toll values. [Arnott et al., 1990a] considered a case of two parallel bottlenecks with an optimal time varying toll, an uniform toll and a step toll. They found that the step toll in general yields more efficiency gains than the uniform toll. [Chu, 1999] confirmed this with an equilibrium simulation model of peak period commuting along a highway. [Laih, 1994] showed that the efficiency gains of step tolls increase with the number of steps.

Regarding behavioral aspects of coarse tolling, there are three different ways to equalize price before and after a toll change. The first is "Laih model" [Laih, 1994], which considers two separate queues for tolled users and untolled users after the toll is lifted; the second "ADL model" by [Arnott et al., 1990a] has a mass departure right after the toll is lifted; the third "Braking model" by [Lindsey et al., 2012, Xiao et al., 2012] incorporates more realistic behavior that users who are about to pass the tolling point just before toll is lifted would brake and wait until toll is lifted.

[De Palma et al., 2005] studied both network and time-of-day aspects of second-best pricing on a circular network. Using the dynamic queue-based METROPOLIS model, they found that welfare gains are higher with area pricing (pricing all trips within a cordon) than with cordon pricing and higher with step tolls than with flat tolls. Also, there is a trade-off between welfare gain and acceptability. Specifically, the highest welfare gain (61% of the first-best pricing) is achieved by area pricing with step tolls. While cordon pricing with step tolls achieves less relative welfare gain as 44%, it has the highest proportion of positive consumer-surplus changes as 41%.

A third case of the second-best problem concerns user heterogeneity. [Verhoef et al., 1995] found that within a static model the second-best toll is a weighted average of the marginal external costs of different groups. It is found that heterogeneity plays a significant role in benefit analysis of the second-best pricing [Small and Yan, 2001, Verhoef and Small, 2003]. [Verhoef and Small, 2003] found that the welfare benefits of the second-best pricing of one of two parallel links can be dramatically underestimated if heterogeneity is ignored.

Another case of second-best pricing is about uncertainty, including idiosyncratic and objective uncertainties [Small et al., 2007]. Idiosyncratic uncertainty is about individual idiosyncratic behavior, such that they perceive traffic conditions differently. A standard way to model this is the Stochastic User Equilibrium (SUE). [Yang, 1999] demonstrated that marginal cost pricing is still applicable in logit based SUE. [Maher et al., 2005] extended analysis on stochastic system optimum to a general utility maximizing framework. [Jiang et al., 2011] developed a multi-criterion dynamic user equilibrium traffic assignment model considering heterogeneous users for analyzing a variety of road pricing scenarios in large-scale networks.

Objective uncertainty refers to unpredictable traffic conditions because of accidents, bad weather, unusual events, or others. Instruments like Advanced Traveler Information System (ATIS) are developed to guide travelers to make better decisions. However, information provision could be welfare-reducing if travelers' responses are not considered [Rapoport et al., 2014]. [Verhoef et al., 1996] studied the joint application of information and congestion pricing in a stochastic two-route network. They found information provision and flat tolling are highly complementary. [De Palma and Lindsey, 1998] had similar findings and also found that when price is not responsive, information provision could reduce welfare gains from congestion pricing. [de Cea Ch et al., 2009] extended the work of [Verhoef et al., 1996] to incorporate various information penetration. They found when the impact of non-recurrent events is small, congestion pricing is more efficient. [Liu and Yang, 2021] studied heterogeneous travelers in a two-route network and demonstrated how willingness of travelers change with information price and toll.

The last case is about the influence of pricing on the rest of the economy. For instance, tolling could discourage labor supply as it raises the living costs for low income workers. [Mayeres and Proost, 2001] used a general-equilibrium model to evaluate the efficiency effects of tolling in Belgium and found inefficient spending of congestion pricing revenues can weaken the net benefits of the policy. [Parry and Bento, 2001, Van Dender, 2003] obtained similar findings. [Anas, 2020] applied a general-equilibrium model to the greater LA region and found that recycling the toll revenue by cutting the income taxes of low income workers can bring additional benefits.

2.1.3 Public Acceptance

Despite numerous theoretical studies supporting congestion pricing, there are still many concerns in practical applications. Successful applications, such as Singapore's Electronic Road Pricing Scheme (ERP), London's Congestion Charge (CC), and Stockholm's Congestion Tax have highlighted the effectiveness of congestion-pricing schemes. However, low public acceptance remains as the main challenge to the successful design and deployment of congestion pricing as indicated by the failures of previous plans in the Greater Manchester, Edinburgh and New York City [Altshuler, 2010, Schuitema et al., 2010].

As pointed out in [de Palma and Lindsey, 2011], congestion pricing practices can be divided into four categories: facility-based, zonal-based, cordon-based and distance-based. Zonal and cordon are often combined together as area-based pricing. Since area-based pricing has more popularity than others, there are more studies on public acceptance about them [Sørensen et al., 2014, Hensher and Li, 2013, Noordegraaf et al., 2014]. A widely used approach to study public acceptance is called qualitative case study approach [Seale et al., 2003] based on the principle that certain closely related cases can serve as a foundation for generalizations.

[Gu et al., 2018] have a detailed review of evidence in the literature for various cases and show that there are four main factors of public acceptance, including privacy, complexity, uncertainty and equity. Privacy concern is mainly about travelers' information among various stakeholders [Ison and Rye, 2005]. It was the main reason that led to the rejection of Hong Kong congestion pricing [Hau, 1990]. Singapore ERP and London congestion pricing addressed privacy concerns by not recording personal information and itineraries of travelers [Noordegraaf et al., 2014].

The failures of Edinburgh and the Greater Manchester congestion pricing were partially because the proposed schemes were more complicated compared to those of Stockholm and Milan [Hensher and Li, 2013]. They considered two pricing cordons while Stockholm only considered one. It was found that transition from complexity to simplicity is effective to gain public acceptance [Hensher and Li, 2013].

Regarding uncertainty, [De Borger et al., 2008] identified two sources of uncertainty as uncertainty about the effectiveness of the proposed scheme and uncertainty in terms of revenue allocation. Through surveying 368 Edinburgh residents, [Gaunt et al., 2007,Allen et al., 2006] found that the lack of understanding of the effectiveness of congestion pricing scheme was one of key reasons for the rejection. In terms of revenue allocation, [Farrell and Saleh, 2005] found using stated preference surveys that congestion pricing revenues should be used to improve public transit, while [Ubbels and Verhoef, 2006] found that congestion pricing revenues should be used to replace car taxes or lower fuel taxes also using surveys. Therefore, [De Borger and Proost, 2012] compared two alternatives and found that applying the generated revenues to improve public transit receives more support.

With respect to equity, the out-of-pocket charges hamper the low-income from using road facilities and make road usage a privilege of the high-income [Lindsney and Verhoef, 2001,Gu et al., 2018]. Through studying the New York congestion pricing experience, [Schaller, 2010] found that it is not enough for congestion pricing to be just perceived as benefiting society without benefiting drivers individually. [Jaensirisak et al., 2003] has similar findings that congestion pricing could be more acceptable if it increases everyone's benefit (so called Pareto-improving) besides the total social benefit. Successful applications in London, Stockholm and Milan all have some measures to address equity concerns. For example, London provides full exemption to disabled people and discounts to residents of the central zone [Santos and Fraser, 2006] and Milan also provides full exemption to disabled people and discounts to frequent users [Rotaris et al., 2010]. [Jaensirisak et al., 2005] found that a fixed charge was more preferred than a variable one while [Francke and Kaniok, 2013] found that distance-based charge with fixed kilometer charge was the most preferred than others on average.

To resolve the equity problem of congestion pricing, several studies in the literature try to find congestion pricing schemes that are Pareto improving directly. [Lawphongpanich and Yin, 2010] investigated searching anonymous nonnegative and Pareto improving (in terms of travel time) toll schemes. They only considered a single mode and homogeneous users. Also, they did not consider other system objectives (e.g. maximize social welfare or minimize travel time). They found that a Pareto improving toll might not always exist. [Nie and Liu, 2010] found that self-financing and Pareto improving pricing schemes might not always exist when the continuously distributed VOT is highly skewed to the right (lower end). [Xiao and Zhang, 2014] developed a link-based toll and subsidy (via negative toll) scheme on a one origin network to achieve Pareto-improving, system optimal and revenue neutral. However, it is not guaranteed to exist for a network with multiple origins and destinations.

Another stream of research tried to achieve Pareto improving schemes through revenue refunding/distribution. [Small, 1992] showed that it is possible for congestion pricing to be progressive if all users receive an equal travel allowance. [Adler and Cetin, 2001] proposed an analytical model for a two-node two-route bottleneck network, in which toll revenues collected from users on a more desirable route are transferred users on a less desirable route. It is shown that their approach can eliminate queuing time and reduce travel cost of all users. Under the assumptions that all users have the same choice set and all alternatives have the same monetary cost, [Eliasson, 2001] showed that a tolling scheme that reduces aggregate travel time and refunds the revenues equally to all users will make everyone better off. Using Vickrey's model [Vickrey, 1969], [Arnott et al., 1994] found that it is not always possible to get Pareto-improvement with an equal lump sum refund. [Guo and Yang, 2010] studied OD-based Pareto-improving congestion pricing revenue refunding schemes for the fixed demand case with multiclass users. They proved that the existence of OD-based uniform Pareto-improving refunding scheme is ensured if the degree of user heterogeneity of

VOT is not too large. [Guo et al., 2012] extended it to elastic demand. [Liu et al., 2009] investigated the performance of distributing equal allowance in multi-modal systems.

2.1.4 Summary

Congestion pricing has low public acceptance, which prevents its widespread application. Despite the recognition of the potential of using congestion pricing revenues to improve its acceptance, current methods focus on simple equal refunding, which is not guaranteed to achieve Pareto improvement in realistic settings with user heterogeneity. Recently, there has been increasing interest in practice on providing discounts to low income toll road users to counter their losses. Some active programs can be found in Illinois, Los Angeles and Norfolk. Besides toll discounts, there are other applications of using incentives or discounts to benefit low income users. For example, Portland Bureau of Transportation started a transportation incentive program to increase low income users' mobility in 2018 [McNeil et al., 2021]; Bay Area Metropolitan Transportation Commission and Ford GoBike/Motivate provided discounted bike share memberships in 2018. However, to the best of our awareness, the literature has not attempted to study methodologies for providing discounts or incentives in a personalized way given it is achievable with current technology development. Personalization has the potential to allocate resources more efficiently and further improve efficiency and equity. It is also a promising means of designing Pareto improving revenue refunding schemes. In addition, existing literature on the analysis of the welfare impacts of congestion tolling suffers from unrealistic assumptions (e.g. homogeneity, inelastic demand, linear income effect) and pays less attention to distributional impacts.

2.2 Tradable Mobility Credits

Recently, tradable mobility credits (TMC; also termed TCS or Tradable Credit Schemes in the literature), a type of quantity control, has been receiving increasing attention in literature leveraging its successful applications in other fields. The main idea of TMC is that users receive an initial allocation of credits/tokens from the regulator. In order to use the road network, they have to spend a certain number of tokens which can vary with the alternative used. They can buy and sell tokens in a market, in which token price is determined by token demand and supply.

In the literature, some studies use permits instead of credits to describe a more restrictive scheme such that road usage is restricted to travelers who have road and time specific permits (e.g. [Wang et al., 2018, Wada and Akamatsu, 2013]), which can be treated as a special type of credit-based schemes.

The main advantages of TMC include: 1) no transfer of money from users to the regulator (i.e. revenue neutral); 2) still effective when the price elasticity in short term is low because of controlling quantity; 3) progressive distributions of benefits even with uniform token allocations because of trading; and 4) more efficient than pricing under demand or supply uncertainty.

2.2.1 Methodological Development

Early work on the use of tradable mobility credits in transportation date back several years [Verhoef et al., 1997, Raux, 2007, Goddard, 1997]. Existing literature on TMC can be broadly classified into three domains and includes comprehensive reviews corresponding to each. The first domain includes studies that propose and conceptualize the scheme design, implementation and credit distribution of individual TMC [Fan and Jiang, 2013]. The second domain focuses on formulating a mathematical programming approach to study user equilibrium and market equilibrium under different assumptions, such as with/without transaction costs, fixed/elastic demand, homogeneous/heterogeneous users [Grant-Muller and Xu, 2014]. The third domain includes works that empirically investigate individual behavior under TMC schemes, considering risk aversion, mental accounting, loss aversion and other concepts [Dogterom et al., 2017].

The initial endowment of credits plays a key role within a TMC system and it involves both determining the number of available credits (which is closely related to the credit tariff rates), selecting eligible credit recipients, and defining the allocation policy to recipients [Fan and Jiang, 2013]. [Fan and Jiang, 2013] lists two main methods of initial allocation, including free allocation and pay-for-permits allocation. [Wada and Akamatsu, 2013] adopted a auction mechanism where the central authority sells the permits in an auction market to travelers.

The free allocation approach is adopted by most of studies, which can be further classified into a uniform allocation and a non-uniform allocation, such as an origin-destination specific credit distribution investigated in [Yang and Wang, 2011a] which might make every user better off. However, there is evidence from other fields suggesting that such a non-uniform allocation can cause entry barriers [Hepburn, 2006, Newell et al., 2005]. Hence, it is important to systematically investigate the initial credit allocation in a transportation context. [Xiao et al., 2019] propose a cyclic tradable credit scheme (CTCS) which does not need periodic collection and redistribution of credits. The link-based credit charge can be either positive or negative (as a subsidy) to have credits circulated within the system such that the number of total credits collected from the travelers is constrained to be zero.

Regarding the credit charging scheme, it could be time-place specific (e.g. [Wong, 1997, Buitelaar et al., 2007]) wherein the mobility credits are distinct for every network link and time-interval. However, operationalizing such a system is likely to be extremely complicated and would necessitate a sophisticated booking system and trading market. Alternatively, time-place dependent TMCs (e.g. [Verhoef et al., 1997, Raux, 2007, Yang and Wang, 2011a]) -which are the most widely discussed in the literature- are universal and usable on the entire network, but the token tariff rates themselves can differ across the network links and time of entry. The time-place dependent TMCs could further have tariffs that are quantified by trips (e.g [Fiorello et al., 2010]), by days (e.g. [Goddard, 1997]) or by vehicle-miles-traveled (e.g. [Verhoef et al., 1997, Raux, 2007, Yang and Wang, 2011a]). The nature of the token charging scheme itself is another aspect of the TMC scheme design that warrants more systematic investigation.

Adopting mathematical programming or Variational Inequality (VI) approaches, a large body of current literature examines user and market equilibrium under different assumptions. In the model of [Yang and Wang, 2011b], the regulator distributes a pre-specified number of credits to travelers, charges a link-specific credit tariff and allows trading of credits within a market. They demonstrate that for a given set of credit rates in a general network, the user equilibrium (UE) link flow pattern is unique under standard assumptions and identify additional conditions (relatively mild) to ensure uniqueness of the credit price at the market equilibrium. Extensions to their model have been proposed to incorporate heterogeneity in the value of time [Wang et al., 2012] and multiple user classes [Zhu et al., 2015] using variational inequality formulations to establish existence and uniqueness properties of the network and market equilibrium. [He et al., 2013] employ a similar equilibrium approach considering allocations of credits to not just individual travelers, but to transportation firms such as logistics companies and transit agencies.

The effect of transaction costs in a TMC scheme on efficiency with two types of markets (auction-based and negotiated) is considered by [Nie, 2012]. [Bao et al., 2014] consider travelers' loss aversion by formulating a reference-dependent user equilibrium model. The transaction costs are also incorporated in the travel disutility function. They demonstrate the system optimum link flow pattern may not be achievable.

[Xiao et al., 2019] extend [Xiao and Zhang, 2014]'s investigate Pareto-improving toll and subsidy schemes to tradable mobility credits using [Yang and Wang, 2011b]'s tradable credit system. With their cyclic tradable credit scheme and under certain conditions on the OD matrix of the network, the Pareto-improving solution is shown to exist.

In contrast with the aforementioned TMC schemes, [Kockelman and Kalmanje, 2005, Gulipalli and Kockelman, 2008] proposed a system of credit-based congestion pricing (termed CBCP) where credits are allowances used to pay tolls. [Bao et al., 2019] showed that the equilibrium under TMC scheme with Vickrey's bottleneck model of congestion is not unique while with Chu's dynamic flow congestion model it is unique.

While most attention focuses on static equilibrium, literature on the dynamics of the credit price considering travelers day-to-day learning is scarce. [Ye and Yang, 2013] incorporated a day-to-day learning model under route choice setting to model the dynamic evolution of traffic flow and credit price. The work investigated the conditions for stability and convergence. [Miralinaghi and Peeta, 2016] is the first study to address the long-term planning problem by modeling the multi-period TMC. The framework developed allows the central authority solve the planning problem at the end of each perid with current conditions and future forecast but subject to critical assumptions like perfect information about future credit prices and homogeneous travelers. [Miralinaghi et al., 2019] extend this framework to consider heterogeneity of value of time and schedule delay and travelers' loss aversion. The proposed model can be combined with day-to-day models of [Ye and Yang, 2013] to represent credit price and flow evolutions in each period.

On the other hand, the comparison of efficiency and equity properties of tradable credits and congestion pricing has received relatively less attention. [Akamatsu and Wada, 2017] showed that TMC is equivalently efficient as congestion under perfect information in a general network, while TMC is advantageous when demand information is not perfect. [de Palma et al., 2018] performed a comparative analysis of the two instruments in a simple transportation network, which has parallel roads for travelers to choose and road usage is free if there is no pricing or TMC in place. They showed that without uncertainty price and quantity are equivalent as in the regular market case studied by [Weitzman, 1974]; under uncertainty, TMC can outperform the pricing instrument in terms of efficiency if the congestion cost is strongly convex. [Xiao et al., 2019] demonstrate that TMC can improve equity compared to several traditional tolling schemes.

Experiments about TMC are also limited. An online experiment about tradable carbon allowance scheme for personal travel is conducted by [Aziz et al., 2015]. They found that high income groups are less sensitive to carbon cost increase for work trips while middle and low income groups are highly sensitive to carbon cost increase for non-work trips. [Tian et al., 2019] conducted an online experiment focusing on behavioral effects under TMC schemes. They observed loss aversion, an immediacy effect and a learning effect and showed that the proposed TMC is efficient and financially sustainable.

Recently, [Brands et al., 2020] empirically tested a complete market design for TMC scheme though a lab-in-the-field experiment where participants make virtual travel choices and real transactions in a tradable parking permits setting. The results showed that credit prices stay within a desired range, and the number of bought and sold quantities kept close to each other, in accordance with a theoretical market equilibrium. [Brands et al., 2021] conducted a 8-week field experiment with TMC applied to manage parking. They adopted the market design proposed in [Brands et al., 2020]. They observed that active users adjusted behavior as intended and participation required less effort than people anticipated. From survey responses, TMC schemes are perceived as a fairer and better alternative to paid parking.

2.2.2 Summary

In summary, despite the large body of research on TMCs, several gaps remain. First, the modeling of the market has received little attention and almost all the studies employ an equilibrium approach to model the credit market (with the notable exception of [Ye and Yang, 2013] who model the price and flow dynamics of a tradable credit scheme). The literature has –to the best of our knowledge– thus far not attempted to model realistically the disaggregate behavior of individuals within the market that could enable the consideration of empirically observed phenomena such as loss aversion, endowment effects, mental accounting, day-to-day learning [Dogterom et al., 2017]. Second, design aspects of the credit market have received little attention (despite being a critical step towards real-world deployment) including features such as token allocation/expiration, trading, intervention, and transaction fees, and the impact of these on behavior of individuals in the market and efficiency. Finally, income effects (that impact both efficiency and equity) have received relatively little attention (with the exception of [Wu et al., 2012] who consider it in a route choice setting). A part of this thesis aims to address these gaps by proposing and examining alternative market designs of the TMC system and investigating their performance relative to congestion pricing using realistic models of traveler behavior (with heterogeneity and income effects), congestion and their interactions.

Chapter 3

Market Design for Tradable Mobility Credits

This chapter aims to design a market (including allocation/expiration of credits, transaction fees, price adjustment, and rules governing trading) for a tradable mobility credits (TMC) system, and develop a methodology that explicitly models the disaggregate behavior of individuals within the market.

3.1 System Design

Before we describe the elements of market design, we first describe a real-time architecture of the TMC system to operationalize the scheme depicted in Figure 1-3. The proposed TMC system comprises three main components, an online bi-level optimization model, a market model and a smartphone app (see Figure 3-1).

From the user's perspective, the proposed TMC system is a smartphone app which includes (1) a personalized trip planner, (2) interfaces to allow the user to manage her/his token account and (3) a trading module. Prior to a trip, travelers open the trip planner, which presents a menu with different travel alternatives along with their predicted attributes. Each alternative is also associated with a token tariff, which is calculated based on the alternative-specific contribution to the system's congestion. The app tracks and verifies realized trips, and charges tokens accordingly. Validated

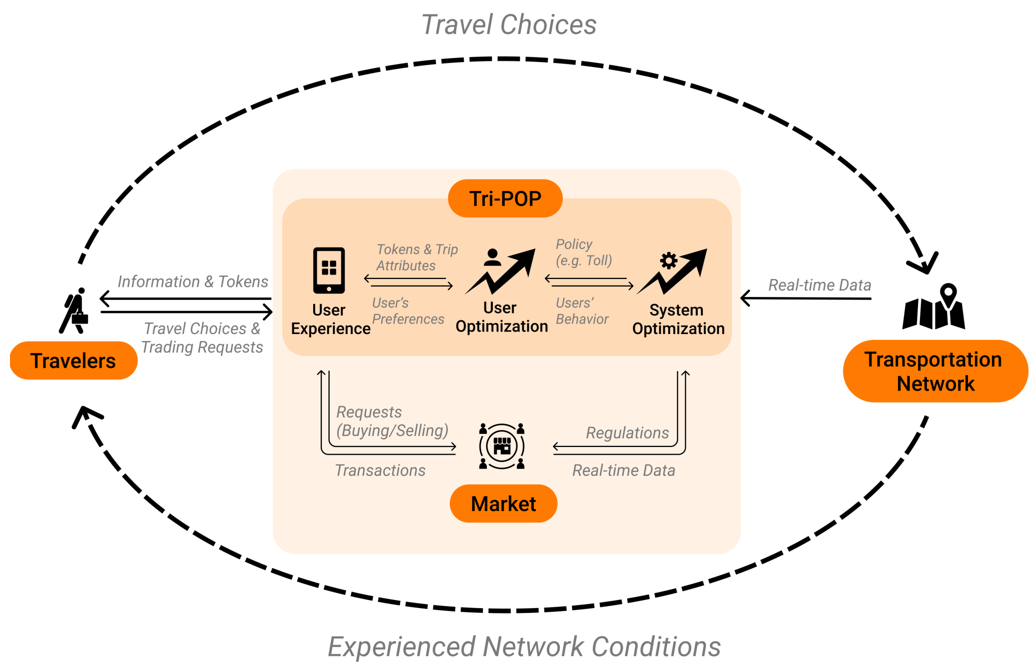


Figure 3-1: The proposed TMC system architecture

frequent travelers will be endowed with a budget of tokens (e.g., mobility credits) obtained through free allocation by the regulator. Moreover, it learns individual user's preferences from previous choices and presents personalized menus, which increases the user's benefit [Song et al., 2018].

The second component of the proposed TMC system is the bi-level optimization module, which is responsible for setting the token charges or tariff for each travel alternative in real time ('system-level' optimization) and provides personalized 'user-optimal' menu to travelers ('user-level' optimization). The system-level optimization utilizes a simulation-based predictive system that uses real-time data from the market and from sensors in the transportation system [Araldo et al., 2019]. The overall policy objectives for the proposed TMC system in terms of congestion, emissions, network performance, quality of service and sustainability are defined via the system-level optimization. It is an application of a general framework of online analytics for smart mobility and transportation system management named as Tri-POP, which combines prediction, optimization and personalization (POP). More details about Tri-POP are provided in Chapter 5.

The third component of the proposed TMC system is the Market in which users can sell or buy tokens. If a user chooses a travel alternative associated with a certain token amount and her token budget is insufficient, she can buy the remaining needed tokens. On the other hand, a user can sell excess tokens in her budget at any time. The token market price at which these exchanges occur adjusts dynamically based on demand and supply of tokens. If demand exceeds supply, the price increases and vice-versa. The market enables the TMC system to achieve revenue neutrality, which means the system avoids taxes, user charges or incentive funding programs. The operator can also intervene in the market, reducing or increasing the number of tokens available and thus allowing for a better management of non-recurrent situations.

Note that the key focus in this chapter is on the analysis and design of the market within the proposed TMC system. Chapter 5 focuses on developing the bi-level optimization framework for pricing and TMC with personalized distribution.

3.1.1 Market Design

In this section, the features of the market within the proposed TMC scheme and the behavior of users in this market are described in detail. Within the TMC scheme, the regulator provides a token endowment to all travelers (more on the nature of this endowment in the following paragraph). The application we explore in the numerical experiments involves a daily commute where in order to use the network at a particular time-of-day (e.g., for a given departure time interval), travelers have to pay a pre-specified toll in tokens that does not vary from day to day. In other words, the toll in tokens is dynamic and varies by time-of-day, but is fixed across days. The rationale for this assumption is that modifying the toll in tokens from day to day would involve communicating the tariff or toll structure on a daily basis, which is complicated, particularly in large general networks (for instance, the electronic road pricing or ERP scheme in Singapore includes dynamic tolls, which are revised only every three months or longer). The developed market design is generic and can be used for other applications as well including parking management.

The regulator operates a market where tokens can be bought and sold at a prevailing market price, and may also levy pre-specified transaction fees for buying and selling. The market price of the token varies across days and is adjusted by the regulator to achieve revenue neutrality, considering the demand and supply of tokens in the market. Note that all transactions take place between an individual and the regulator directly, who guarantees all buying and selling requests. This central market with a regulator who acts as a price setting intermediary is similar to the *virtual bank* in [Brands et al., 2020], who note that such a market can significantly reduce transaction costs (associated with information acquisition, negotiation, finding a potential buyer or seller etc.) compared to designs that involve consumer to consumer trading (and over existing designs such as Dutch and English auctions, sealed-bid auctions and Vickerey auction markets). The regulator may also intervene in the token market within a day by controlling token market price, token allocation, and transaction fees to manage non-recurrent events.

With regard to the token allocation or endowment, we adopt a 'continuous time' approach wherein tokens are acquired (provided by the regulator) at a certain rate over the entire day and each token has a lifetime (i.e., it expires after a certain period specified by the regulator). The expiration of tokens will avoid undesirable consequences of the TMC system that can compromise public acceptability such as speculative behavior and hedging in the market. The 'continuous' allocation avoids concentrated trading activities and excessive trading near boundary (a time period when a large amount of tokens expire at the same time for lump-sum allocation.). It also provides more degrees of freedom for the regulator to intervene than that of a 'lump sum' allocation which distributes tokens at the beginning of each day. A comparison between the two allocation approaches will be performed through numerical experiments presented in Chapter 6.

As a result, each individual acquires tokens at a constant rate r over the entire day (credited into a *wallet*) and each token has a lifetime L to avoid speculation and hoarding. Let $x_n^d(t)$ denote traveler n's token account (or wallet) balance at time t on day d. A full wallet state indicates that the number of tokens in the wallet has reached a maximum (Lr), and in the absence of travelling or selling, does not change since the acquisition of new tokens is balanced by an expiry of old tokens. Thus, a full wallet implies that the oldest token in an individual's account has an age of L. In contrast, when the account is not in a full wallet state, it increases by an amount $r\Delta_t$ in a unit time interval Δ_t.

Several additional assumptions regarding market design are noteworthy – these serve to avoid quantity buildup and market manipulation. First, travelers can only buy tokens from the regulator at the time of traveling for immediate use, i.e., only if they wish to travel and are short of tokens. Second, when they sell tokens to the regulator, they have to sell all tokens in their wallet. Third, buying and selling cannot happen at the same time, i.e. travelers can sell all tokens anytime except at the time of buying. Note that the second assumption differs from the design of [Brands et al., 2020], who assume that tokens can be traded per piece, and implications of this assumption warrant more investigation, particularly when the market prices vary within-day. Since a large part of our experiments do not involve within-day dynamic prices and given that it considerably simplifies the modeling of selling behavior, we defer the relaxation of this assumption to future research.

3.1.2 Account Evolution

Let $T(t)$ denote the toll in tokens to travel at time t, $\tilde{t}_n^d$ represent the departure time of traveler n on day d and D represent the duration of one day. Note that in the simulation (Chapter 4) and experiments (Chapter 6), time will be discretized into time intervals of a specified size; for now, we treat it as continuous. Let r denote the allocation rate, L denote token lifetime, and $x_n^d(t)$ denote traveler n's token account balance at time t on day d. At time t on day d, traveler n can perform one and only one of the following actions:

1. Perform a trip if $t = \tilde{t}_n^d$.

    - If $x_n^d(t) \geq T(t)$, she consumes $T(t)$. Her account balance at $t + \Delta_t$, $x_n^d(t +$

Δ_t), can be written as:

$$x_n^d(t + \Delta_t) = min\left(x_n^d(t) - T(t) + r\Delta_t, Lr\right) \tag{3.1}$$

where the cap Lr ensures that tokens with life greater than L expire.

- If $x_n^d(t) < T(t)$, she needs to buy $T(t) - x_n^d(t)$ tokens. Her account balance $x_n^d(t + \Delta_t)$ becomes:

$$x_n^d(t + \Delta_t) = r\Delta_t \tag{3.2}$$

since all of $x_n^d(t)$ and the newly bought tokens are used to travel.

2. Does nothing. Her account balance $x_n^d(t + \Delta_t)$ becomes:

$$x_n^d(t + \Delta_t) = min\left(x_n^d(t) + r\Delta_t, Lr\right) \tag{3.3}$$

3. Sells all tokens $x_n^d(t)$. Her account balance becomes:

$$x_n^d(t + \Delta_t) = r\Delta_t \tag{3.4}$$

3.1.3 Buying and Selling

The token market price p^d is fixed within day d (in the absence of non-recurrent events) and is only adjusted day to day. Details of the price adjustment process are discussed in Section 3.1.4. We assume that the regulator levies a two-part (fixed and proportional) transaction fee for both buying and selling transactions. Let F_S^P, F_B^P ($F_S^P, F_B^P \geq 0$) denote the proportional part of selling and buying transaction fees (this component of the transaction fee is proportional to the amount of the trade), and F_S^F, F_B^F ($F_S^F, F_B^F \geq 0$) denote the fixed part of selling and buying transaction fees. The effect of transaction fees on market behavior and efficiency will be examined in Chapter 6.

The revenue obtained from selling y tokens ($y \leq Lr$) tokens with transaction fees on day d at time t can be written as,

$$S(y) = yp_s^d - F_S^F \tag{3.5}$$

where $p_s^d = p^d(1 - F_S^P)$, which is the token market price adjusted for the proportional selling transaction fee. Transaction fees and price are not expressed in function inputs for conciseness.

The cost of buying y tokens ($y \leq Lr$) tokens with transaction fees at time t on day d can be written as,

$$B(y) = yp_b^d + F_B^F \tag{3.6}$$

where $p_b^d = p^d(1 + F_B^P)$, which is the token market price adjusted for the proportional buying transaction fee.

3.1.4 Price Adjustment

The marketplace dictates the token price p^d on day d, which is adjusted according to an apriori rule established by the regulator to achieve revenue neutrality. The price p^d is modified daily with a deterministic rule considering the regulator revenue K^{d-1} (net revenue from all buying and selling transactions of users) from the previous day as follows

$$p^d = \begin{cases} p^{d-1} & K^{d-1} \in [-K_t, K_t] \\ p^{d-1} + \Delta p & K^{d-1} < -K_t \\ p^{d-1} - \Delta p & K^{d-1} > K_t \end{cases} \tag{3.7}$$

where Δp currently is a constant parameter representing the price change amount.

K_t is a constant parameter representing a regulator revenue threshold to adjust the price and ensures that price will not fluctuate for small regulator revenues close to zero. Price is ensured to be positive and below a certain cap p_m as follows:

$$p^d = \max\left(0, \min\left(p^d, p_m\right)\right) \tag{3.8}$$

Although token price is typically constant within a day, the regulator may intervene in the market to adjust the market price during a day in the presence of unusual events. For example, if road capacity drops because of an accident, or if demand increases due to a concert, the regulator can intervene, increasing token price in certain period to discourage travel and reduce congestion. Numerical experiments are conducted to study this in Chapter 6.

Market elements discussed in this section are summarized in Table 3.1.

Table 3.1: Market elements for the tradable mobility credits system

<table>
<tr><th>Elements</th><th>Design</th><th>Motivation</th></tr>
<tr><td rowspan="2">Allocation</td><td>Lump-sum</td><td>Simple; automated trading</td></tr>
<tr><td>Continuous</td><td>Avoid concentrated trading;
additional control</td></tr>
<tr><td>Expiration</td><td>Lifetime</td><td>Avoid quantity buildup</td></tr>
<tr><td rowspan="2">Transaction fee</td><td>Proportional</td><td rowspan="2">Avoid undesirable market
behavior (e.g. frequent selling)</td></tr>
<tr><td>Fixed</td></tr>
<tr><td>Price adjustment</td><td>Day to day
constant adjustment</td><td>Balance demand and supply</td></tr>
<tr><td colspan="3">Market rules governing trading</td></tr>
</table>

3.2 Market Behavior

As buying behavior is governed by the previously specified buying rule, this section mainly discusses individual selling behavior. It is assumed that individual selling decision is based on their mobility decision (departure time $\tilde{t}_n$). In general, the decision to sell can be formulated as a dynamic programming or optimal control problem, where the optimal selling strategy is characterized by Bellman's equation [Kirk, 2004]. However, this may be unrealistic as a model of individual decision making. Instead, a simpler heuristic approach is developed to model an individual's selling strategy.

At time t on day d, assume traveler n has an upcoming planned trip at a time denoted by $\tilde{t}_n$, where $\tilde{t}_n = \tilde{t}_n^d$ if $t \leq \tilde{t}_n^d$, and $\tilde{t}_n = \tilde{t}_n^{(d+1)}$, if $t > \tilde{t}_n^d$. Given the next trip, a conditional profit function $\Pi_n^d(t)$, which represents the profit obtained by selling all tokens at time t (with no future selling until the next departure $\tilde{t}_n$) can be written as follows,

$$
\begin{aligned}
\Pi_n^d(t) &= S\left(x_n^d(t)\right) - \mathbb{I}\left(T(\tilde{t}_n) \geq \hat{x}_n(\tilde{t}_n)\right) \cdot B\left(T(\tilde{t}_n) - \hat{x}_n(\tilde{t}_n)\right) \\
&= x_n^d(t)p_s^d - F_S^F - \mathbb{I}\left(T(\tilde{t}_n) \geq \hat{x}_n(\tilde{t}_n)\right) \cdot \left(\left(T(\tilde{t}_n) - \hat{x}_n(\tilde{t}_n)\right) p_b^d + F_B^F\right)
\end{aligned} \tag{3.9}
$$

where $\hat{x}_n(\tilde{t}_n)$ represents the expected account balance at next travel $\tilde{t}_n$. Since it is assumed there will be no future selling until the next departure $\tilde{t}_n$, it can be written as,

$$
\hat{x}_n(\tilde{t}_n) = \min\left[(\tilde{t}_n - t)r, Lr\right] \tag{3.10}
$$

For other notation in the conditional profit function $\Pi_n^d(t)$, $T(\tilde{t}_n)$ represents toll in tokens of traveling at departure time $\tilde{t}_n$. A buying cost is incurred only if the toll at $\tilde{t}_n$ is greater than or equal to traveler n's expected account balance (i.e. $T(\tilde{t}_n) \geq \hat{x}_n(\tilde{t}_n)$), which is represented by the indicator function. Note that in defining the profit function above, we have made the critical assumption that if a decision to sell at the current time is made, no further selling will occur until the next trip. This simplification allows us to derive an optimal selling strategy analytically and is partly justifiable given that we also assume that during selling, an individual needs to sell all tokens in her wallet, and that prices do not vary within-day.

Under our assumptions, at time t on day d, traveler n will consider selling tokens only if the profit value is positive, i.e., $\Pi_n^d(t) > 0$. If the profit value is positive, she may still decide to wait if the derivative of the profit function is positive (meaning that the profit is expected to increase if she defers the decision to sell). Therefore, the selling strategy depends on both the profit function and its derivative, which can be analyzed from the following three cases:

1. $T(\tilde{t}_n) < \hat{x}_n(\tilde{t}_n)$ (no tokens need to be bought for the next trip)

   The profit function $\Pi_n^d(t)$ can be written as

$$\Pi_n^d(t) = x_n^d(t) p_s^d - F_S^F \tag{3.11}$$

   and the derivative can be written as

$$\frac{d\Pi_n^d(t)}{dt} = \begin{cases} 0 & x_n^d(t) = Lr \\ rp_s^d & \text{otherwise} \end{cases} \tag{3.12}$$

   which implies that profit will continue to increase until a full wallet is reached. It does not make sense to wait longer at a full wallet because new acquired tokens are canceled out with the expired tokens. Hence, the selling should be at full wallet.

   However, it is worth noting that, without fixed transaction fees, the selling revenue at full wallet is the same as selling every time when one receives new tokens. In fact, as long as one avoids token expiration, any selling strategy is equivalent in the absence of fixed transaction fees. It is fixed transaction fees that avoid frequent selling.

2. $T(\tilde{t}_n) > \hat{x}_n(\tilde{t}_n)$ (tokens need to be bought for the next trip)

   The profit function $\Pi_n^d(t)$ can be written as

$$\Pi_n^d(t) = x_n^d(t)p_s^d - F_S^F - \left(\left(T(\tilde{t}_n) - \hat{x}_n(\tilde{t}_n)\right) p_b^d + F_B^F\right) \tag{3.13}$$

and its derivative can be written as

$$\frac{d\Pi_n^d(t)}{dt} = \begin{cases} -rp_b^d & x_n^d(t) = Lr \\ rp_s^d - rp_b^d & \text{otherwise} \end{cases} \tag{3.14}$$

which is always negative since $p_s^d < p_b^d$ given F_B^P or F_S^P is greater than 0. This implies that profit obtained from waiting and selling at any time in the future (until the next trip) is guaranteed to be less than the profit from selling now. Hence, she should sell now if profit is positive.

Without transaction fees, the profit function $\Pi_n^d(t)$ can be written as

$$\Pi_n^d(t) = x_n^d(t)p^d - \left(T(\tilde{t}_n) - \hat{x}_n(\tilde{t}_n)\right) p^d \tag{3.15}$$

and its derivative can be written as

$$\frac{d\Pi_n^d(t)}{dt} = \begin{cases} -rp^d & x_n^d(t) = Lr \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

which means as long as account balance is not full, it does not matter whether one sells now or later. However, once we introduce fixed transaction fees, it is better to sell at a full wallet to minimize the number of transactions. With additional proportional transaction fees, it is better to sell immediately and not worth waiting anymore as the derivative is always negative.

3. $T(\tilde{t}_n) = \hat{x}_n(\tilde{t}_n)$ (the expected account balance is just enough to cover the toll of the next trip)

The profit function $\Pi_n^d(t)$ can be written as

$$\Pi_n^d(t) = x_n^d(t)p_s^d - F_S^F \tag{3.17}$$

but its derivative does not exit because the conditional profit function is discontinuous at t due to the transaction fees of buying. To avoid any buying transaction fees (either fixed or proportional), it is optimal to sell immediately if profit $\Pi_n^d(t)$ is positive. Without transaction fees, similarly, it does not matter to sell now or later as long as token expiration is avoided.

Based on the analysis, the effect of fixed transaction fees is to prevent multiple transactions while the effect of proportional transaction fees is to make one sell as soon as possible when the conditional profit is positive (if buying is required at next travel). The proportional transaction fee is not preferable because it does not prevent frequent selling but instead prevents selling at a full wallet. Numerical experiments in Chapter 6 will provide further justification for the use of only a fixed transaction fee from an efficiency perspective.

The selling strategy for an individual n at any time t on day d considering positive transaction fees is summarized in Algorithm 1.

3.3 Summary

This chapter introduces the detailed market design for tradable mobility credits schemes to manage travel demand and avoid undesirable market behavior, including hoarding, frequent selling, excessive activity at boundary and negotiation cost. Market elements like token allocation, expiration, transaction fee, price adjustment and market rules governing trading are explained in detail. A heuristic approach to model disaggregate selling behavior is developed and the resulting simple selling strategy is derived. The effect of proportional and fixed transaction fees on selling behavior are discussed analytically.

Algorithm 1: Selling Rule

```
input: d, t, n, p^d, t̃_n, x_n^d(t), L, r
At time t on day d, calculate Π_n^d(t);
and expected account balance x̂_n(t̃_n) = min[(t̃_n − t)r, Lr];
if Π_n^d(t) > 0 then
    if T(t̃_n) ≥ x̂_n(t̃_n) then
        Sell now;
    else
        if x(t) = Lr then
            Sell now;
        else
            Do nothing;
        end
    end
else
    Do nothing;
end
```

Chapter 4

Simulation Framework

This chapter describes the modeling and simulation framework for evaluating the performance of the designed instruments including both tradable mobility credits and pricing. Following this, the design of the personalized pricing and TMC schemes are discussed in the next chapter. The overall simulation framework is shown in Figure 4-1.

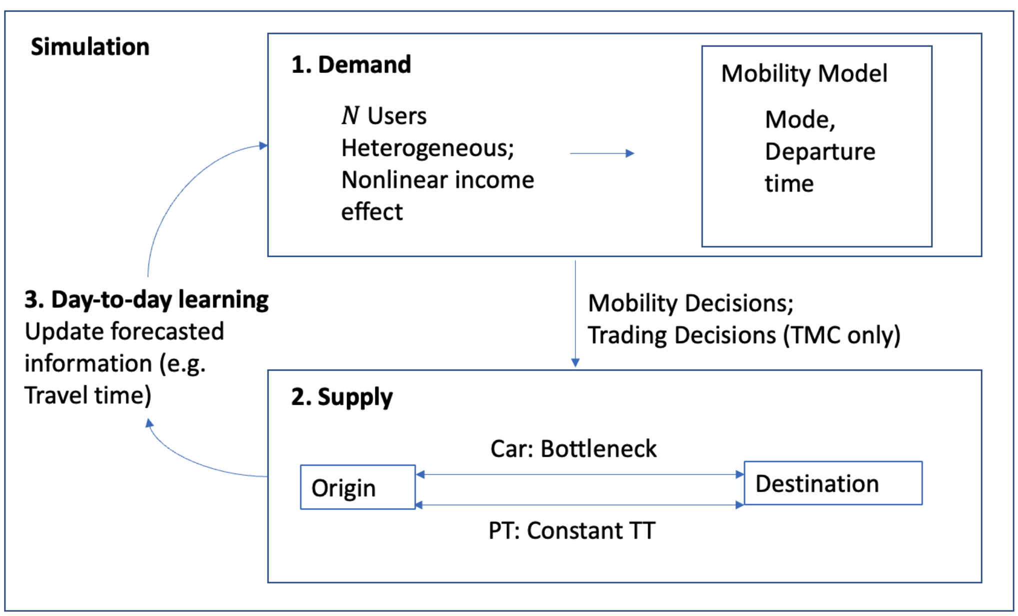


Figure 4-1: Simulation Framework

N travelers perform a daily commute between a single origin-destination pair. For the sake of simplicity, each traveler performs a single morning trip and a single evening trip. Only their morning commute trip will be explicitly simulated and their evening trip is assumed to be a mirror of the morning trip.

At the beginning of each day, every traveler uses forecasted information of travel times, schedule delays and their account balance over the entire day to make a *pre-day mobility decision*, which is the combination choice of mode (between car and PT) plus departure time (over a individual set of departure time choices) for their morning commute trip. Travelers who choose to drive may be subject to a time-of-day toll. For TMC, the time-of-day toll profile is in units of tokens. Note that mobility credits can only be used for toll road payment. The individual mobility decision is modeled using a logit mixture model allowing for heterogeneity and non-linear income effects.

Next, the determined mobility decisions along with trading decisions (TMC only) are simulated on a simple network connected by a single driving path and an alternative public transit (PT) line. Congestion (for driving) is modeled by a point queue model (bottleneck of finite capacity), in which a queue develops once flow exceeds capacity. Travel time of PT is assumed to be constant.

Travelers' day-to-day learning is modeled through an exponential smoothing filter to update forecasts of travel times and account balance. The day-to-day framework in Figure 4-1 is used to simulate the evolution of the system state (departure flows, travel times) until a measure of convergence has been reached. The performance measures (overall welfare, distribution of user benefits, congestion, and mode shares) at convergence are used to evaluate the given instruments.

In the following sections, we first describe the models of demand, supply and day-to-day learning, termed the *system model*, in more detail. Next, we discuss social welfare computation and the simulation-based toll optimization problem (to determine optimal tolls) for the different instruments. The evaluation framework is introduced at the end. Relevant notation is shown in Table 4.1. Note that models and optimization formulation introduced in this chapter pertain to pricing without revenue distribution (P-) and TMC with uniform allocation (MU). Other instruments (e.g. pricing with revenue distribution) can be incorporated with modifications discussed in more detail in Chapter 5.

Table 4.1: Notation

Variables	**Description**
h	Departure time interval
$\bar{t}$	Simulation time step
d	Day d
t_h	Start time of interval h
Δ_h	Duration of departure time interval
Δ_t	Duration of simulation time step
Δ_a	Size of desired arrival window
n	Individual n
α_n	Value of time of individual n
β_{En}, β_{Ln}	Value of schedule delay early/late of individual n
λ	Coefficient of nonlinear income effect
γ	Nonlinear income effect adjustment parameter
μ_n	Random component scale parameter of individual n
ϵ_{in}	Random utility component for mobility decision i of individual n
I_n	Disposable income of individual n
H_n	Departure time choice set of individual n
M_n	Mode choice set of individual n
$\hat{t}_n$	Desired arrival time of individual n
η	Departure time window size parameter
p	Market price
$\mathrm{T}^j(h)$	Toll of instrument j in h
$\tilde{\tau}_i$	Forecasted travel time of choice i
$\tilde{c}_{in}$	Expected cost for mobility decision i
$x_n^d(t)$	Account balance of individual n at time t
L	Token lifetime
r	Token allocation rate
t_f	Free flow travel time
$t_v(t)$	Delay in queue at t
$Q(t)$	Number of drivers in queue at t
θ_τ/θ_t	Weights on previous day's forecasts
$(\cdot)^j$	Variable associated with instrument j

4.1 System Model

As noted previously, the setting we consider involves N users traveling between a single origin-destination pair connected by a path containing a bottleneck of finite capacity and a PT line. Users wish to arrive at the destination within a certain "preferred arrival time window" in the morning, and can choose between PT and car. If they decide to drive, they can adjust their departure times to avoid congestion (similar to the model in [Ben-Akiva et al., 1984], which is a dynamic extension of [De Palma and Lefevre, 1983]). The system is modeled using a stochastic process approach that can be viewed as a simplification of the model in [Cascetta and Cantarella, 1991], which considers the stochastic assignment problem in general networks. Day to day adjustment is modeled using suitable learning and forecasting filters, within-day departure time decisions and mode choices are modeled using a logit-mixture model, and supply is modeled using a point queue model. We refer to [Cantarella and Cascetta, 1995] for a nuanced discussion of terminology and a detailed description of deterministic and stochastic process models (with probabilistic assignment or a probabilistic model for users' choice behavior). It is noted that the model of [Ben-Akiva et al., 1984] may be viewed as a deterministic process model with probabilistic assignment.

The mobility demand model, network model, and demand-supply interactions are discussed in detail next.

4.1.1 Demand Model

The demand model (preday mobility decision) is a model of departure time and mode choice. Unless otherwise specified, everything discussed below pertain to day d. The day is discretized into $h = 1 \dots H$ time intervals of size Δ_h (let the set of all time intervals in the day be denoted by $\mathcal{H} = \{1, \dots, h, \dots, H\}$), and it is assumed that each individual n has a preferred/desired arrival time $\hat{t}_n$ (more specifically, users are assumed to wish to arrive within a time window of size $2\Delta_a$ centered around $\hat{t}_n$; this is discussed in more detail later in the section). The day is also discretized into smaller time intervals of size $\bar{t} = 1 \dots \bar{T}$ of size Δ_t, which is the resolution of the supply model and trading (selling) decisions.

The choice set of mode for individual n is defined as $M_n = \{C, PT\}$, where C represents car choice and PT represents PT choice. The choice set of feasible departure time intervals $H_n \subset \mathcal{H}$ is individual-specific and defined as $H_n^C = \{\tilde{t}_{0n} - \eta\Delta_h, \tilde{t}_{0n} - (\eta - 1)\Delta_h, \ldots, \tilde{t}_{0n} + \eta\Delta_h\}$, where η is a parameter, and $\tilde{t}_{0n}$ represents the initial departure time interval on day 0, which is computed based on the preferred arrival time $\hat{t}_n$ and the free flow travel time. Thus, the departure time choice set H_n consists of 2η time intervals of size Δ_h centered around the preferred departure time interval on day 0, $\tilde{t}_{0n}$. Let $i = (m, h)$ represent an individual mobility decision as a combination of mode and departure time choice ($i \in \{m, h | m \in M_n, h \in H_n\}$).

Each individual is assumed to be rational and she wants to maximize her money metric utility from the choice. The utility of the mobility decision i is denoted as U_{in}. The individual utility consists of two parts: a systematic utility V_{in} which is a function of observable variables and a random utility component ϵ_{in} which represents the analyst's imperfect knowledge. ϵ_{in} is assumed to be in i.i.d. extreme value distribution with zero mean and individual specific scale parameter μ_n. It is also assumed that individual random utility term is perfectly correlated across instruments (i.e. remains the same before the change and after the change) assuming before and after periods are not too far apart (e.g. [McFadden, 2001,de Palma and Kilani, 2005]). This assumption can be relaxed in future work according to [Delle Site and Salucci, 2013, Zhao et al., 2008].

The systematic money-metric utility for individual n departing in time interval h by car under instrument j, $V_{in}(\tilde{\boldsymbol{\phi}}_i^j)$, where $i \in \{m = C, h | h \in H_n\}$. $\tilde{\boldsymbol{\phi}}_i^j$ is a vector of forecasted information in the systematic utility that affects the choice of departure time interval of driving. The first is forecasted/expected travel time $\tilde{\tau}_i^j$, which determines the expected schedule delay early (second component) and schedule delay late (third component). The fourth component is expected cost $\tilde{c}_{in}^j$ which is explained in more detail next. The last component is remaining income, which is equal to the disposable income for transportation I_n minus expected cost $\tilde{c}_{in}^j$.

The marginal utility of an additional unit of travel time for individual n is denoted by α_n. For simplicity, we assume travelers have common knowledge of forecasted travel times (more on this in section 4.1.3). The desired arrival time window for individual n is defined as $[\hat{t}_n - \Delta_a, \hat{t}_n + \Delta_a]$, where $\hat{t}_n$ represents the center of the period and Δ_a represents arrival flexibility. If she arrives outside of desired time period, she suffers a schedule delay. The marginal utility of an additional unit of schedule delay early is β_{En} and an additional unit of schedule delay late is β_{Ln}, where $\beta_{En} \leq \alpha_n \leq \beta_{Ln}$ according to empirical results (e.g., [Small, 1982]).

The expected cost $\tilde{c}^j_{in}$ warrants additional discussion. Under the No Toll (NT) scenario, it is equal to the operation cost c_f (fuel cost). Under pricing ($j = P$), it is equal to the toll in dollars charged for departing in time interval h, $T^P(h)$, plus an operation cost c_f, which can be written as

$$\tilde{c}^P_{in} = T^P(h) + c_f \tag{4.1}$$

Under the TMC ($j = M$) scheme, it depends on an individual's expected opportunity cost of tokens $\tilde{R}_{in}$ plus the operation cost c_f as follows:

$$\tilde{c}^M_{in} = \tilde{R}_{in} + c_f \tag{4.2}$$

Recall that the selling revenue of y tokens with transaction fees (F^F_S, F^P_S) and token price (p) can be written as (selling revenue function),

$$S(y) = yp\left(1 - F^P_S\right) - F^F_S \tag{4.3}$$

and similarly, the buying cost of y tokens can be written as (buying cost function),

$$B(y) = yp\left(1 + F^P_B\right) + F^F_B \tag{4.4}$$

Let t_h be the start time of interval h, $\tilde{x}_n(t_h)$ be the expected account balance at time t_h, the beginning time of the time interval h. If a traveler does not need to pay any toll, she can sell the entire day's token allocation completely. Hence, the opportunity cost (or negative opportunity benefit) is equal to the negative of selling revenue of the entire day's allocation, $S(Lr)$.

If a traveler needs to pay toll $T^M(h)$ in h but the expected account balance $\tilde{x}_n(t_h)$ is greater or equal to $T^M(h)$ (no buying), her opportunity cost is equal to the negative of selling revenue of one-day allocation Lr minus toll in tokens $T^M(h)$, which can be written as

$$\tilde{R}_{in} = -S\left(Lr - T^M(h)\right) \tag{4.5}$$

However, if she does not have enough account balance to cover the toll $T^M(h)$, she has to buy additional tokens equal to $T^M(h) - \tilde{x}_n(t_h)$ in order to travel in h. The amount of tokens she can sell for profit is equal to one-day allocation Lr minus her expected account balance $\tilde{x}_n(t_h)$ since all of her tokens will be used for toll payment if she departs in h. The opportunity cost can be written as

$$\tilde{R}_{in} = -S\left(Lr - \tilde{x}_n(t_h)\right) + B\left(T^M(h) - \tilde{x}_n(t_h)\right) \tag{4.6}$$

In summary, the expected opportunity cost $\tilde{R}_{in}$ depends on an individual's forecasted account balance $\tilde{x}_n(t_h)$, market price p, the toll in tokens $T^M(h)$ and transaction fees as follows:

$$\tilde{R}_{in} = \begin{cases} -S\left(Lr - T^M(h)\right) & \tilde{x}_n(t_h) \geq \mathrm{T}^M(h) \\ -S\left(Lr - \tilde{x}_n(t_h)\right) + B\left(T^M(h) - \tilde{x}_n(t_h)\right) & \text{otherwise} \end{cases} \tag{4.7}$$

Note that if transaction fees are zero, the opportunity cost in Equation 4.7 reduces to the one-day allocation minus the toll in tokens times token price, i.e., $\tilde{R}_{in} = -\left(Lr - \mathrm{T}^M(h)\right)p$. In the absence of non-linear income effects, Lr can be overlooked because it is a constant which does not affect the choice.

Regarding the income effect, the diminishing marginal utility of income suggests that as an individual's income increases, the extra benefit to that individual decreases. It is thus natural to model this nonlinear effect of remaining income by a quasiconcave function (as per [McFadden, 2017]). Hence, we add the remaining income plus a natural log of the remaining income to the systematic money-metric utility.

The utility of an individual n driving and departing in time interval h (choosing a mobility decision $i \in \{m = C, h | h \in H_n\}$) under instrument j can be written as,

$$\begin{aligned} U_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) =& V_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) + \epsilon_{in} \\ =& -2\alpha_n \tilde{\tau}_i^j - \beta_{En} SDE\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) - \beta_{Ln} SDL\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) \\ & + I_n - 2\tilde{c}_{in}^j + \lambda ln\left(\gamma + I_n - 2\tilde{c}_{in}^j\right) + \epsilon_{in} \end{aligned} \tag{4.8}$$

where

$$SDE\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) = \max\left(0, \hat{t}_n - \Delta_a - (t_h + \tilde{\tau}_i^j)\right) \tag{4.9}$$

$$SDL\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) = \max\left(0, (t_h + \tilde{\tau}_i^j) - \hat{t}_n - \Delta_a\right) \tag{4.10}$$

Schedule delay of the evening trip is ignored because it is assumed to be more flexible. The individual departure time choice set H_n is also subject to a budget constraint (i.e. an individual cannot choose a departure time that is not affordable).

The systematic money-metric utility function of user n who departs in time interval h by PT is denoted as $V_{in}(\tilde{\boldsymbol{\phi}}_i^j)$, where $i \in \{m = PT, h | h = \lfloor \hat{t}_n - \tau_{PT} \rfloor\}$. Since the travel time and headway of PT are constant, we only need to consider one departure time interval h, which has a corresponding arrival time closest to the desired arrival time $\hat{t}_n$. For PT, the input vector $\tilde{\boldsymbol{\phi}}_i^j$ for the systematic utility consists of four components: PT travel time τ_{PT}, expected waiting time W_{PT}, expected PT cost $\tilde{c}_{in}^j$ and remaining income $I_n - \tilde{c}_{in}^j$.

The marginal utility of an additional unit of PT travel time of individual n is assumed to be the same as that of car travel time as α_n. The marginal utility of an additional unit of waiting time is β_{Wn}.

Regarding the expected PT cost $\tilde{c}_{in}^{j}$, it is equal to operation cost (i.e. PT fare) c_{PT} under the No Toll (NT) scenario and pricing. Under the TMC scheme, it depends on an individual's expected opportunity cost of tokens $\tilde{R}_{in}$ plus operation cost c_{PT}, where $\tilde{R}_{in}$ is equal to the negative of selling revenue of a full wallet Lr since travelers who choose PT can sell all of her tokens acquired in one day for maximum return. It can be written as $\tilde{R}_{in} = S(Lr)$

Hence, the expected PT cost $\tilde{c}_{in}^{j}$ under the TMC scheme can be written as

$$\tilde{c}_{in}^{M} = -\tilde{R}_{in} + c_{PT} \tag{4.11}$$

The utility of an individual n using PT who departs in interval h (choosing a mobility decision $i \in \{m = PT, h|h = \lfloor \hat{t}_n - \tau_{PT} \rfloor\}$) can be written as,

$$\begin{aligned} U_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) =& V_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) + \epsilon_{in} \\ =& -2\alpha_n \tau_{PT} - 2\beta_{Wn} W_{PT} + I_n - 2\tilde{c}_{in}^{j} + \lambda ln\left(\gamma + I_n - 2\tilde{c}_{in}^{j}\right) + \epsilon_{in} \end{aligned} \tag{4.12}$$

4.1.2 Supply Model

The network is assumed to be a single origin-destination pair connected by a single path containing a bottleneck of fixed capacity s [Arnott et al., 1990b]. A first-in-first-out (FIFO) queue develops once the flow of travelers exceeds s. The free flow travel time is t_f and the extra delay time for a traveler departing from home at time t is $t_v(t)$. Thus, the total travel time for a traveler departing from home at time t is:

$$\tau(t) = t_v(t) + t_f \tag{4.13}$$

Let $Q(t)$ be the number of travelers in the queue at time t. The delay time at time t is derived from the deterministic queuing model as follows:

$$t_v(t) = \frac{Q(t)}{s} \tag{4.14}$$

where $Q(t) = 0$ and $t_v(t) = 0$ when there is no congestion.

Note that within our simulation, the capacity s is defined for time intervals $\bar{t}$ of size Δ_t. The travel time for a given departure time interval h is obtained by averaging the travel times of all travelers departing in h. Further, the exact time of departure of a traveler within the supply model is randomly (uniformly) drawn within the chosen departure time interval h.

For the alternative PT line, it has constant travel time τ_{PT}. Its headway is also constant, which is equal to twice of the expected waiting time W_{PT}.

4.1.3 Day-to-Day Learning

Let τ_i^{d-1} denote the actual or experienced car travel time on day $d-1$ of choice i, where $i \in \{m = C, h | h \in H_n\}$. As we specified in the demand model, travelers are assumed to make their choices of departure time according to forecasted car travel times $\tilde{\tau}_i^d$, $\forall h \in \mathcal{H}$ from their memory and learning. We use an exponential smoothing filter, a type of homogeneous filter [Cantarella and Cascetta, 1995], to model the learning and forecasting process by weighting actual and forecasted costs of the previous day as follows:

$$\tilde{\tau}_i^d = (1-\theta_\tau)\tilde{\tau}_i^{d-1} + \theta_\tau \tau_i^{d-1} \tag{4.15}$$

where $\theta_\tau \in [0,1]$ is a learning weight for the previous day's forecasted travel time.

In order to get individual forecasted account balance on day d $\tilde{x}_n^d(\bar{t})$, the individual forecasted departure time $\tilde{t}_n^d$ is first acquired by applying a similar filter as follows:

$$\tilde{t}_n^d = (1 - \theta_t)\tilde{t}_n^{d-1} + \theta_t t_n^{d-1} \tag{4.16}$$

where $\theta_{\tilde{t}} \in [0, 1]$.

Next, the trading model presented in Section 3.1.3 is applied using the individual forecasted departure time to determine their forecasted account balance over the entire day, which is used to compute the expected toll costs under the TMC scheme through Equation 4.7.

4.2 Toll Optimization

This section discusses time-of-day toll profile optimization problem with the objective to maximize the improvement of social welfare. In the following sections, the social welfare measure used in this study is defined first; then the toll profile optimization problem is formulated; finally, the solution procedure of the optimization problem is presented.

4.2.1 Social Welfare

In order to evaluate the policy impacts of the different instruments, a real-value function termed the social welfare function (SWF) is adopted, which is an aggregate measure of the whole society's well-being. From individualistic ethics, social welfare depends on the welfare of all agents in society. If the utility is cardinal, it can be used to measure individual welfare and it is natural to have SWF as a form of a non-weighted sum of individual utilities [Harsanyi, 1955]. A weighted sum is also possible with weights determined based on analysts' value judgements. This viewpoint dates back to the work of Bentham and such a form of SWF is also know as the Benthamite social welfare function. Note that there are other forms of SWF, like Nash's social welfare function [Kaneko and Nakamura, 1979], Rawlsian minmax social welfare function [Rawls, 2020] and Sen's welfare function [Sen, 2009]. In the case considered in this study, users and the regulator are the only two types of agents in society.

For the regulator, it is assumed that the operation cost is constant across instruments. Therefore, its surplus is only represented by regulator revenue (K). Under pricing without revenue distribution (P-), it can be written as

$$K^{P-} = \sum_{n=1}^{N} \sum_{i \in M_n \times H_n^{P-}} \hat{c}_{in}^{P-} \mathbb{I}_n \left(i | \boldsymbol{T}^{P-} \right) \tag{4.17}$$

where i is the mobility decision, which is a combination of mode and departure time choice; $\mathbb{I}_n(i|\boldsymbol{T}^{P-})$ is a indicator if traveler n chooses mobility choice i or not given toll vector $\boldsymbol{T}^{P-} = \{T^{P-}(h)|h \in \mathcal{H}\}$; $\hat{c}_{in}^{P-}$ is equal to the toll payment for driving ($T^{P-}(h)$) or the PT fare payment for PT (c_{PT}); and H_n^{P-} might be instrument specific because of budget constraints.

Under the TMC scheme, regulator revenue K^M consists of two parts. The first part is the sum of PT fare payments and the second part is the sum of tokens bought minus tokens sold over one day. K^M can be written as

$$K^M = \sum_{n=1}^{N} \left(c_{PT} \mathbb{I}_n \left(PT | T^M(h) \right) + \sum_{\bar{t} \in \{1 \dots \bar{T}\}} \left(B \left(T^M(\bar{t}) - x_n(\bar{t}) \right) \mathbb{I}_n^B \left(\bar{t} | \boldsymbol{T}^M \right) - S \left(x_n(\bar{t}) \right) \mathbb{I}_n^S (\bar{t} | \boldsymbol{T}^M) \right) \right) \tag{4.18}$$

where $B(\cdot)$ is cost of buying function and $S(\cdot)$ is revenue of selling function; $\mathbb{I}_n^B \left(\bar{t} | \boldsymbol{T}^M \right)$ and $\mathbb{I}_n^S \left(\bar{t} | \boldsymbol{T}^M \right)$ are indicators of buying or selling at $\bar{t}$ given toll in tokens $\boldsymbol{T}^M$. Note that the price adjustment mechanism described in Section 3.1.4, is designed to ensure that the regulator revenue of the TMC scheme at equilibrium is close to zero, thereby achieving revenue neutrality.

In the literature, in the presence of nonlinear income effects, three ways are usually used to measure user benefits: Compensating Equivalent (MCE), Hicksian Compensating Variation (HCV), and Hicksian Equivalent Variation (HEV). MCE is equal to the difference in indirect utilities between a “but for” scenario and an “as is” scenario, scaled to money metric units by dividing by the marginal utility of income (MUI) at the “as is” scenario level. It differs from the commonly used Marshallian consumer surplus (MCS) only in the MUI scaling factor. It can be easily computed when the indirect utility function and its derivatives are known. HCV is equal to the net decrease in “but for” scenario income that equates utility in two scenarios while HEV is equal to the net increase in “as is” scenario income that equates utility in two scenarios. Refer to [McFadden, 2017] for a detailed discussion.

A major drawback of these three measures is that their ethical implications are not defensible as pointed out by [Blackorby and Donaldson, 1990]. Well-being measured in units of income treat increases in income as equally socially valuable no matter who receives them. This is not the case for net utility improvement as the nonlinear effect of income improvement is captured by a nonlinear income effect term in the utility specification.

In this study, user benefits (Z^j) under instrument j is the sum of all users' net experienced utilities relative to NT denoted as z_n^j. This is similar to what [De Palma and Lindsey, 2004] used except they summed up the log transformations of individual utilities. Since utilities adopted in this study are money-metric, the net utility amount serves as a meaningful measurement of improvement directly. An individual n's net experienced utility is the difference between maximum utility under instrument j and under NT, which can be written as,

$$z_n^j = \max_{i \in M_n \times H_n^j} \left(U_{in} \left(\boldsymbol{\phi}_i^j \right) \right) - \max_{i \in M_n \times H_n^{NT}} \left(U_{in} \left(\boldsymbol{\phi}_i^{NT} \right) \right) \tag{4.19}$$

where $\boldsymbol{\phi}_i^j$ is a vector of experienced variables under instrument j and $\boldsymbol{\phi}_i^{NT}$ is a vector of experienced variables under NT.

Hence, the user benefits Z^j can be written as

$$Z^j = \sum_{n=1}^{N} z_n^j \tag{4.20}$$

4.2.2 Optimization Formulation

It is possible to consider different time-of-day toll profile specifications. For a smooth and continuous toll profile, one can express the tolling function (or profile) by a mixture of M Gaussian functions, which are specified by $3M$ parameters, mean, variance and amplitude, $m = 1, 2, \ldots, M$. However, in practice, implementing a smooth and continuous toll profile of this nature is complex. Hence, the toll function that we consider is a step toll profile (of the kind implemented in Singapore and Stockholm), which consists of five step toll values and six break points.

The toll profile optimization can be formulated as a simulation-based optimization problem with the objective of maximizing total social welfare (SW) as follows

$$\begin{aligned} \max_{\boldsymbol{T}^j} \quad & Z^j + K^j \\ \text{s.t.} \quad & Z^j, K^j = SM\left(\boldsymbol{T}^j, \boldsymbol{\xi}, \boldsymbol{\psi}\right) \\ & \boldsymbol{T}^j = \{T^j(h) | h \in \mathcal{H}\} \\ & \boldsymbol{T}^j \geq 0 \end{aligned} \tag{4.21}$$

where j can be either $P-$ or MU; toll profile $\boldsymbol{T}^j$ is a set of toll values over the entire day. In the case of pricing, the toll profile is in units of dollars while it is in units of tokens in the case of TMC. For the TMC system, in addition to the toll profile in tokens, other parameters like the allocation rate r and transaction fees have to be determined. For the allocation rate r, making it exogenous will avoid the issue of non-uniqueness (during optimization) of combinations of allocation rate and the toll profile. For transaction fees, they are determined based on not only social welfare optimization, but also undesirable market behavior reduction, which is discussed in detail in Chapter 6.

$\boldsymbol{\xi}$ represents all input data for simulation, such as individual income, preferred arrival time, and choice attributes. $\boldsymbol{\psi}$ represents all model parameters, such as demand model coefficients, bottleneck capacity, user learning weights, and market parameters for the TMC scheme. The $SM(\cdot)$ function is the system model discussed in Section 4.1.

The user benefit Z^j is defined in Equation 4.20 and the regulator revenue K^j is defined in Equation 4.17 and 4.18.

Clearly, the optimization problem in 4.21 has no closed-form since the objective function for a given toll profile is the outcome of a simulation of the stochastic process (a simulation-based optimization problem), or more specifically, the system model presented in 4.1, which includes traveler behavior, regulator states and actions, and the resulting network and market conditions. The social welfare given an instrument j is evaluated at simulation convergence.

4.2.3 Solution Procedure

In order to solve this simulation-based optimization problem, a differential evolution (DE) algorithm is adopted as it is derivative-free and performs well for global optimization problems of this kind [Price, 2013]. In the literature, metaheuristic algorithms have been shown to work well for nonconvex and nonlinear toll design problems (e.g. [Shepherd and Sumalee, 2004,Zhang and Yang, 2004]). In this section, the solution procedure is introduced and some key steps are illustrated in detail.

Let $\mathbf{X}$ represent the decision variables of the simulation-based optimization problem (i.e., the parameters of the toll profile), $\mathbf{X} = (\xi_0, \xi_m, A_m)\,,\ m = 1, 2, \ldots, M$, where ξ_0 represents the starting time of the toll and ξ_m represents the breaking points for the mth step of the toll; A_m represents the toll value for the interval between ξ_{m-1} and ξ_m; M represents the number of steps. The procedure of the differential evolution (DE) algorithms is as follows:

Step 1: *Initial population.* Randomly generate an initial population of feasible step toll profiles of size $\mathcal{N}$, where $\mathcal{N}$ represents the solution population size. Feasibility constraints include $\xi_m \geq \xi_{m-1} + 15$ (minimum duration of one step toll is 15 minutes) and $A_m < 15$ (maximum toll has to be less than 15 dollars).

Step 2: *Simulation evaluation.* Run simulations in parallel given $\mathcal{N}$ solutions and obtain social welfare values at convergence.

Step 3: *Mutation.* Randomly uses individuals from the current solution population to generate variant vectors subject to feasibility constraints. For instance, the gth variable of vector i at generation k, $\mathbf{Y}_{g,i}^{k}$, is given by

$$\mathbf{Y}_{g,i}^{k} = \mathbf{X}_{g,r1}^{k} + F \cdot \left(\mathbf{X}_{g,r2}^{k} - \mathbf{X}_{g,r3}^{k}\right) \tag{4.22}$$

where $r1,\ r2,\ r3$ are randomly selected from $[1, \mathcal{N}]$, $i \neq r1 \neq r2 \neq r3$, and F is a scale factor. If the resulting $\mathbf{Y}_{g,i}^{k}$ is not feasible, discard and re-generate a feasible solution.

Step 4: *Crossover.* Combine variants and original solutions subject to feasibility constraints. For example, a trial vector $\mathbf{U}_{i}^{k}$ is created by combining the variant vector and original vector as follows,

$$\mathbf{U}_{g,i}^{k} = \begin{cases} \mathbf{Y}_{g,i}^{k} & \text{if } rand(0,1) < C_r \text{ or } g = rg \\ \mathbf{X}_{g,i}^{k} & \text{otherwise} \end{cases} \tag{4.23}$$

where $C_r \in [0,1]$ is the crossover rate, $rand(0,1)$ represents a random uniformly distributed variable within $(0,1)$, and rg is a random integer in $[1, 2M]$ ensuring at least one variable of the trial vector $\mathbf{U}_{i}^{k}$ is from the variant vector $\mathbf{Y}_{i}^{k}$. Similarly, $\mathbf{Y}_{g,i}^{k}$ is re-generated if it violates feasibility constraints.

Step 5: *Next generation.* Produce the next generation solutions by comparing the previous solutions and trial solutions. For example, the next generation vector i is acquired by comparing the original vector $\mathbf{X}_{i}^{k}$ and the trial vector $\mathbf{U}_{i}^{k}$ in terms of social welfare as follows,

$$\mathbf{X}_{i}^{k+1} = \begin{cases} \mathbf{U}_{i}^{k} & \text{if } SW(\mathbf{U}_{i}^{k}) > SW(\mathbf{X}_{i}^{k}) \\ \mathbf{X}_{i}^{k} & \text{otherwise} \end{cases} \tag{4.24}$$

Step 6: *Verification of stopping criterion.* If the stopping criterion has not been reached, go to Step 2; otherwise stop.

This algorithm is fairly robust to initial values because of mutation, crossover and selection steps in the algorithm.

4.3 Evaluation Framework

The policy evaluation is based on microsimulation outputs under the baseline NT and various designed instruments. Each instrument is simulated until convergence so that the performance indicators represent the day at equilibrium. The evaluation framework include four main categories: social welfare, distributional impacts, behavior change and level of congestion, incorporating aggregate-level and disaggregate-level indicators.

Social welfare is a monetary aggregated measurement of overall societal well-being. As discussed in previous section, it consists of user benefits and regulator revenue.

In order to analyze how user benefits distribute across users, the Lorenz curve [Gastwirth, 1972] of individual benefit z_n^j is plotted. If the curve is strictly nonnegative, it means the corresponding instrument is Pareto improving. As Pareto improvement does not relate to equity, a commonly used equity index, the Gini coefficient, is also calculated. Note that Gini coefficient is calculated using disposable income I_n plus monetary benefit z_n^j to measure the inequality of new income distribution for a given instrument j.

As there are only two modes considered in this study, the indicator for behavior change is PT usage. The purpose of this evaluation is to help policy-makers adapt relevant public services and prevent undesirable activities accordingly. For example, charging a higher toll may increase PT usage, which may require improved PT supply. Also, giving revenue refund or tokens may have undesirable consequences like increasing number of cancelled trips, which may reduce economic productivity. It is important to consider other travel behavior indicators in future work, such as number of trips, number of activities, and trip distance.

The last category is level of congestion, which is measured by the car trip travel time index (TTI). It is an average ratio of simulated travel time to free flow travel time. It can also be weighted by distance to incorporate the effect of trip length. However, in this study, we only consider a single path, and hence, weighting is not required.

It is worth noting that the simulation is stochastic. Therefore, it is important to examine whether the performance differences are due to policy impacts or stochasticity. On the one hand, the level stochasticity in the simulator is controlled as much as possible using the same random seed for simulations of different scenarios to ensure the same agent has the same random error and coefficients (e.g. VOT) drawn in demand model. On the other hand, each policy is evaluated multiple times with different random seeds in numerical experiments. The level of stochasticity is quantified and considered in comparisons.

4.4 Summary

This chapter discusses the simulation framework for evaluating the performance of various instruments. The demand model, supply model and day-to-day learning are explained in detail. Next, toll design is formulated as a simulation-based optimization problem to maximize social welfare at equilibrium. A metaheuristic algorithm, namely the Differential Evolution algorithm is applied to solve it. Finally, the evaluation framework including four categories of performance indicators are explained. The performance of various instruments are compared against the NT baseline.

Chapter 5

Personalization for Pricing and Tradable Mobility Credits

As a uniform revenue refund has been shown to not guarantee Pareto improvement, it is natural to consider how to refund revenue more efficiently through personalization. Since low income users are known to more likely lose from pricing, personalization has the potential to improve equity at the same time. Given current developments in technology, it becomes easier to infer users' preferences but there has been no research on methodologies for personalized revenue refunding and tradable mobility credits.

Due to the hierarchical nature of the problem, formulating it as a bi-level optimization problem allows us to consider system and user objectives together. This chapter develops a bi-level optimization framework for both price and quantity controls to achieve personalized distribution, which is an application of an online bi-level optimization framework for smart mobility and transportation system management termed Tri-POP [Atasoy et al., 2020].

The rest of chapter first reviews the literature on bi-level optimization and personalized pricing in transportation. Next, optimization problems for uniform and personalized distribution are formulated utilizing the system model discussed in Chapter 4.

5.1 Literature Review

5.1.1 Bi-level Optimization

Bi-level optimization is a mathematical framework suitable for optimizing large-scale hierarchical decision-making processes, wherein the realized outcomes of any policy or decision by the upper level (system) to optimize their goals is affected by the response of lower level entities (users), who want to optimize their own outcomes. The upper level or system optimization contains another optimization problem (user) as a constraint. It is appealing because the objectives of users (e.g. maximize their own benefits) can be in conflict with the system objectives and bi-level optimization provides a way to reconcile conflicting objectives. Refer to [Sinha et al., 2017] for more details about bi-level optimization.

However, because of the hierarchical optimization structure, it also becomes difficult to solve. [Hansen et al., 1992] showed that Bi-level programming is strongly NP-hard. [Vicente et al., 1994] have proven that even evaluating a solution for optimality is also a NP-hard problem. [Deng, 1998] discuss the complexity issues of bi-level optimization and prove that there is no polynomial time algorithm for linear bi-level problems.

Due to its difficult nature, much of the literature considers bi-level optimization with well-behaved functions (e.g. linear, quadratic or convex) with strong assumptions (e.g. continuity, semi-continuity, differentiability). In these cases, a single-level reduction approach can be applied to replace the lower level optimization problem with its KKT conditions and solve the bi-level problem as a single level optimization problem. However, classical methods fail for complex formulations. Recently, more studies adopt evolutionary algorithms to solve bi-level optimization. A detailed survey of classical and evolutionary algorithms can be found in [Sinha et al., 2017].

As pointed out by [Migdalas, 1995], the transportation planning process has hierarchical characteristics for applications of bi-level optimization, such that the public sector at system level wants to improve network performance while users at another level make choices (e.g. route, mode, departure time, destination) to improve their own benefits. For example, [Constantin and Florian, 1995] formulate the problem of optimizing the frequencies of transit lines as bi-level optimization and solve it by sub-gradient algorithm.

More applications of bi-level optimization in the transportation field can be found in toll design. For example, [Labbé et al., 1998] study a bi-level optimization of pricing, in which the system optimization objective is to maximize revenues and the user optimization works like an assignment problem with the objective to find the shortest path. [Yin, 2002] consider a multi-objective system optimization to minimize the total travel cost and maximize the total revenue with a user-equilibrium traffic assignment at the user level. The formulated problem is solved by a genetic algorithm-based (GAB) approach. [Wang et al., 2014b] consider multiple objectives both in system optimization and user optimization. System objectives include minimizing system travel, total vehicle emissions and negative health impacts; user objectives include minimizing travel time and toll. A combination of a metaheuristic and a classical optimization algorithm is used to solve it.

The design of tradable mobility credits has also been explored using bi-level optimization formulations. For example, [Wu et al., 2012] formulated the design of congestion pricing and tradable credit schemes as mathematical programs with equilibrium constraints. The system optimization optimizes toll policy to maximize both net social benefit and equity while the user level determines equilibrium flows and the credit price. Along similar lines, [Wang et al., 2014a] consider capacity enhancements for selected links in addition to link-based credit charges in system optimization to minimize total system costs and a similar equilibrium problem at user level. [Wang et al., 2020] extend the bi-level optimization to consider traffic emission management at the system level.

Recently, [Azevedo et al., 2018] developed a smartphone-based system termed Tripod (an application of the Tri-POP framework) to influence travelers' real time decisions by providing optimized incentives. The involved bi-level optimization framework is novel because it integrates prediction and personalization. The system optimization has the objective to minimize system-wide energy consumption by optimizing energy saved per token and the user optimization has the objective to maximize individual's expected consumer surplus or expected revenue.

5.1.2 Personalized Pricing in Transportation

The main idea of personalization is to tailor a product, service or policy to accommodate specific users in order to improve certain objectives (e.g. revenues, customer satisfaction, conversion, and etc.). In the transportation pricing context, personalization can be achieved via price discrimination based on individual willingness to pay (WTP) for various objectives, including revenue improvement, congestion reduction, and social welfare improvement.

Literature on personalized pricing in the transportation field is scarce. There are some applications in the airline industry to provide personalized fare offers based on estimated WTP [Wittman and Belobaba, 2017]. [Zhang, 2019] develop personalized discounting policies for application on managed lane tolling with an objective of maximizing revenue subject to practical considerations. The demand and supply interactions are captured by a DTA system. A personalized choice model is developed to capture systematic heterogeneity in individual's preferences to managed lanes and sensitivities to toll and travel time.

The Tripod system [Azevedo et al., 2018] mentioned previously uses personalized incentives to encourage altering travel decisions. The personalized menu provided in the Tripod smartphone app includes various travel options and associated incentives. Traveler behavior is modeled by a series of discrete choice models [Xie et al., 2020]. [Zhu et al., 2020] develop a similar personalized system to encourage sustainable travel. The amount of incentives is determined to have the probability of accepting the promoting choice greater than a threshold (e.g. 0.5). The individual preferences are inferred by a particle filter approach. They do not integrate system level goals into the developed personalized system. [Xiong et al., 2020] also develop an integrated and personalized travel information and incentive scheme to minimize system-wide energy consumption. However, they do not consider individual preferences update with real-time data.

5.2 Tri-POP Framework

Tri-POP is an online analytics framework for smart mobility and transportation system management combining capabilities of prediction, optimization, and personalization (POP). Predictions are performed at network and individual levels using real-time data. Based on the latest predictions, optimization algorithms are used to determine the optimal operation parameters at system and user levels. Personalization refers to the user level operation parameters which are determined based on individual preferences considering heterogeneity.

In order to effectively optimize operation parameters, Tri-POP adopts an online bi-level optimization framework shown in Figure 5-1. Periodically, the system level prediction and optimization algorithms determine the system policy (e.g. surge pricing, incentive allocation) for optimizing system-wide objectives in real-time based on predictions. Upon request from a user, the user level optimization determines optimal service (e.g. on-demand) options based on individual preferences (e.g. maximize surplus assortment).

As a proactive operational system, performance of Tri-POP depends on the quality of the predictive models for system dynamics and individual behavior. Predictive models for Tri-POP need to have two key desirable properties. First, as reasonable out-of-sample predictions are critical for meaningful optimization, theory-driven traffic and behavioral models should be preferred compared to black-box models. Second, to address the intrinsic uncertainty of transportation systems, these models need to be estimated or calibrated online to incorporate real-time information (from traffic monitoring system and cellphone sensors).

In summary, the Tri-POP platform is user-oriented such that it considers user behavior in determining the policy and does the best possible for the user. It is also fair as all users are subject to the same policy, which has the potential for wider acceptance. In this study, Tri-POP is applied to provide personalized congestion pricing revenue refunds to improve acceptance of congestion pricing. Broadly, it is applicable to any service that is on-demand and relies on an app (e.g. MOD, ride

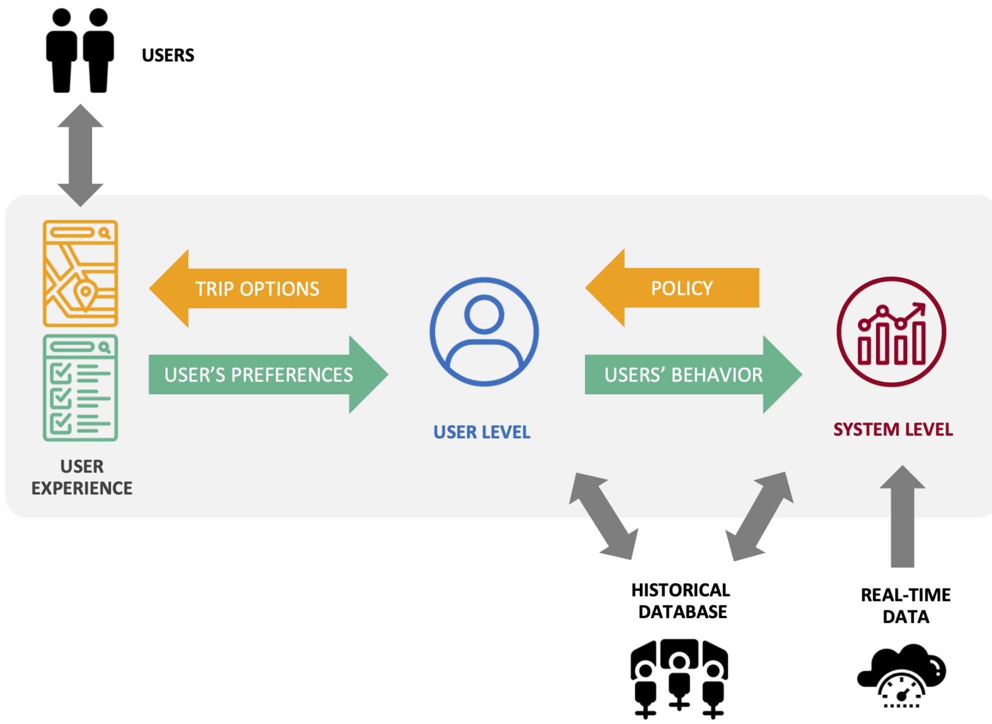


Figure 5-1: Tri-POP Framework

sharing, car sharing, MaaS). More details about Tri-POP can be found in [Atasoy et al., 2020].

5.3 Model Formulation

As discussed in Chapter 4, the context considered in this study is a daily commute problem between a single origin-destination pair for N travelers. Each traveler performs a single morning trip and a single evening trip. Only their morning commute trip is explicitly simulated and their evening trip is assumed to be a mirror of morning trip.

In following sections, the system model (models of demand, supply and day-to-day learning) discussed extensively in Chapter 4 is briefly reviewed and modified to incorporate the effect of revenue refunds. Following this, the optimization problems for uniform and personalized revenue refunds are formulated. Their solution algorithms are discussed at the end.

5.3.1 System model

The only part of the system model described in Chapter 4 that needs to be updated is the demand model to consider the effect of the revenue refund. Let a_n^j denote the amount of refund in dollars received by individual n under instrument j; then her expected remaining income becomes equal to the disposable income I_n minus expected cost $\tilde{c}_{in}^j$ plus the refund a_n^j. Her utility of driving and departing in time interval h (choosing a mobility decision $i \in \{m = C, h | h \in H_n\}$) under instrument j can be written as,

$$\begin{aligned} U_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) =& V_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) + \epsilon_{in} \\ =& -2\alpha_n \tilde{\tau}^j(h) - \beta_{En} SDE\left(h, \hat{t}_n, \tilde{\tau}^j(h)\right) - \beta_{Ln} SDL\left(h, \hat{t}_n, \tilde{\tau}^j(h)\right) \\ & + I_n - 2\tilde{c}_{in}^j + 2a_n^j + \lambda ln\left(\gamma + I_n - 2\tilde{c}_{in}^j + 2a_n^j\right) + \epsilon_{in} \end{aligned} \tag{5.1}$$

where

$$SDE\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) = \max\left(0, \hat{t}_n - \Delta_a - \left(t_h + \tilde{\tau}_i^j\right)\right) \tag{5.2}$$

$$SDL\left(h, \hat{t}_n, \tilde{\tau}_i^j\right) = \max\left(0, \left(t_h + \tilde{\tau}_i^j\right) - \hat{t}_n - \Delta_a\right) \tag{5.3}$$

and individual refund a_n^j is associated with income terms as it is an additional income.

Similarly, the utility of an individual n using PT departing in interval h (choosing a mobility decision $i \in \{m = PT, h | h = \lfloor \hat{t}_n - \tau_{PT} \rfloor\}$) can be written as,

$$\begin{aligned} U_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) =& V_{in}\left(\tilde{\boldsymbol{\phi}}_i^j\right) + \epsilon_{in} \\ =& -2\alpha_n \tau_{PT} - 2\beta_{Wn} W_{PT} \\ & + I_n - 2\tilde{c}_{in}^j + 2a_n^j + \lambda ln\left(\gamma + I_n - 2\tilde{c}_{in}^j + 2a_n^j\right) + \epsilon_{in} \end{aligned} \tag{5.4}$$

As we can see, without considering the nonlinear income effect of remaining income, adding a refund a_n will not change user behavior because a constant a_n will not change utility differences. In contrast, considering the nonlinear income effect, receiving a refund a_n could possibly change user behavior and the degree of change depends on the level on income effect λ and the value of remaining income.

5.3.2 PU Optimization

For pricing with uniform distribution (PU), the main idea is to distribute regulator revenue equally to all users to improve everyone's benefit and make the pricing scheme more politically acceptable. [Small, 1992, Adler and Cetin, 2001] have shown that providing an equal refund can make pricing Pareto-improving (increase everyone's benefit). However, it is not always guaranteed for travelers with heterogeneous values of time [Arnott et al., 1994]. In this study, the PU will be formulated and compared against other instruments.

Similar to the other instruments, the objective of PU is to maximize social welfare, which consists of user benefits and regulator revenue. The total user benefit is equal to the sum of each user's experienced utility difference under PU and NT. Compared to pricing without distribution ($P-$), the distributed refund needs to be considered in the individual's experienced utility under PU. By construction, the refund is the same for everyone, and is denoted by a^{PU}. As a result, the user benefits of PU, Z^{PU} can be written as,

$$Z^{PU} = \sum_{n=1}^{N} \left(\max_{i \in M_n \times H_n^{PU}} \left(U_{in} \left(\boldsymbol{\phi}_i^{PU} \right) \right) - \max_{i \in M_n \times H_n^{NT}} \left(U_{in} \left(\boldsymbol{\phi}_i^{NT} \right) \right) \right) \tag{5.5}$$

For the regulator, its revenue of toll payment and PT fare payment E^{PU} can be written as,

$$E^{PU} = \sum_{n=1}^{N} \sum_{i \in M_n \times H_n^{PU}} \hat{c}_{in}^{PU} \mathbb{I}_n(i|\boldsymbol{T}^{PU}, a^{PU}) \tag{5.6}$$

where i is a mobility decision as a combination of mode and departure time choice; $\mathbb{I}_n(i|\boldsymbol{T}^{PU}, a^{PU})$ is an indicator if traveler n chooses mobility choice i or not given toll vector $\boldsymbol{T}^{PU} = \{T^{PU}(h)|h \in \mathcal{H}\}$ and refund a^{PU}; $\hat{c}_{in}^{PU}$ is equal to the toll payment for driving ($T^{PU}(h)$) or PT fare payment for PT (c_{PT}); and H_n^{PU} might be instrument specific because of budget constraints.

Assume the operation cost is constant across instruments, the surplus of regulator is only represented by the (net) regulator revenue (K). Under pricing with uniform distribution (PU), it can be written as,

$$\begin{aligned} K^{PU} &= E^{PU} - \sum_{n=1}^{N} a^{PU} \\ &= \sum_{n=1}^{N} \left(\sum_{i \in M_n \times H_n^{PU}} \hat{c}_{in}^{PU} \mathbb{I}_n(i|\boldsymbol{T}^{PU}, a^{PU}) - a^{PU} \right) \end{aligned} \tag{5.7}$$

Since the money collected by the regulator needs to be refunded equally to everyone, the refund a^{PU} can be written as,

$$a^{PU} = \frac{1}{N} \delta E^{PU} \tag{5.8}$$

where δ is a parameter representing the percentage of total money collected by the regulator can be distributed. If δ is equal to 1, all regulator revenue will be equally distributed, which achieves revenue neutrality. It can be less than 1 to accommodate the additional operation cost of PU.

The toll profile optimization can be formulated as a simulation-based optimization problem with the objective of maximizing total social welfare (SW) similar to the problem formulated in Chapter 4 to incorporate the equal revenue refund constraint as follows,

$$
\begin{aligned}
\max_{\boldsymbol{T}^{PU}} \quad & Z^{PU} + K^{PU} \\
\text{s.t.} \quad & Z^{PU}, K^{PU}, E^{PU} = SM\left(\boldsymbol{T}^{PU}, a^{PU}, \boldsymbol{\xi}, \boldsymbol{\psi}\right) \\
& a^{PU} = \frac{1}{N}\delta E^{PU} \\
& \boldsymbol{T}^{PU} = \{T^{PU}(h) | h \in \mathcal{H}\} \\
& \boldsymbol{T}^{PU} \geq 0
\end{aligned}
\tag{5.9}
$$

where toll profile $\boldsymbol{T}^{PU}$ is a set of toll values over the entire day; $\boldsymbol{\xi}$ represents all input data for simulation, such as individual income, preferred arrival time, choice attributes; $\boldsymbol{\psi}$ represents all model parameters, such as demand model coefficients, bottleneck capacity, user learning weights, and market parameters for TMC only; $SM(\cdot)$ is the system model covered in Chapter 4; the user benefit Z^{PU} is defined in Equation 5.5; and the regulator revenue K^{PU} is defined in Equation 5.7.

5.3.3 PI Optimization

Along similar lines, the pricing with personalized distribution is formulated as a bi-level optimization problem. The system optimization is to determine the system level policy with the objective to maximize social welfare and the user optimization is to determine the individual refund under suitable objectives (e.g achieve Pareto improvement or maximize social welfare). These two levels are interdependent in that the system optimization depends on the user optimization solution while the user optimization is also dependent on the system optimization solution. Such a bi-level optimization framework reconciles two conflicting objectives and has the potential to improve both efficiency and equity.

System Optimization

For the system optimization, its objective is the same as the objective of PU, which can be formulated as follows,

$$\begin{aligned}
\max_{\boldsymbol{T}^{PI}} \quad & Z^{PI} + K^{PI} \\
\text{s.t.} \quad & Z^{PI}, K^{PI}, E^{PI} = SM\left(\boldsymbol{T}^{PI}, \boldsymbol{a}^{PI}, \boldsymbol{\xi}, \boldsymbol{\psi}\right) \\
& \boldsymbol{a}^{PI} = UO\left(\boldsymbol{T}^{PI}(h)|\boldsymbol{\xi}, \boldsymbol{\psi}\right) \\
& \sum_{n=1}^{N} a_n^{PI} \leq \delta E^{PI} \\
& \boldsymbol{T}^{PI} = \{T^{PI}(h)|h \in \mathcal{H}\} \\
& \boldsymbol{T}^{PI} \geq 0
\end{aligned} \tag{5.10}$$

where $\boldsymbol{a}^{PI}$ represents a set of refunds over the population $\{a_n^{PI}|n = 1, ..., N\}$ determined from the user optimization (UO) given toll policy $\boldsymbol{T}^{PI}$ determined in system optimization. The total revenue refund has to be less or equal to available revenue for distribution.

The user benefits under PI Z^{PI} can be written as,

$$Z^{PI} = \sum_{n=1}^{N} \left(\max_{i \in M_n \times H_n^{PI}} \left(U_{in}\left(\boldsymbol{\phi}_i^{PI}\right)\right) - \max_{i \in M_n \times H_n^{NT}} \left(U_{in}\left(\boldsymbol{\phi}_i^{NT}\right)\right) \right) \tag{5.11}$$

and the total money collected by the regulator E^{PI} can be written as,

$$E^{PI} = \sum_{n=1}^{N} \sum_{i \in M_n \times H_n^{PI}} \hat{c}_{in}^{PI} \mathbb{I}_n\left(i|\boldsymbol{T}^{PI}, a_n^{PI}\right) \tag{5.12}$$

and the (net) regulator revenue K^{PI} can be written as

$$K^{PI} = E^{PI} - \sum_{n=1}^{N} a_n^{PI} \tag{5.13}$$
$$= \sum_{n=1}^{N} \left(\sum_{i \in M_n \times H_n^{PI}} \hat{c}_{in}^{PI} \mathbb{I}_n \left(i | \boldsymbol{T}^{PI}, a_n^{PI} \right) - a_n^{PI} \right)$$

User Optimization

Depending on the objective, the user optimization can be formulated differently. In order to ensure a Pareto improving outcome (i.e. everyone is not worse off but at least one is better off) compared to NT, the revenue can be distributed to make sure every traveler's net experienced utility relative to NT is non-negative. The personalized refunds with this distribution rule is denoted as PI_H.

With a slight abuse of notation, let $z_n \left(I_n - c_{in}^{PI_H} + a_n^{PI_H} \right)$ denote individual n's net experienced utility under PI_H relative to NT as a function of her remaining income $I_n - c_{in}^{PI_H} + a_n^{PI_H}$. The Pareto improving distribution rule can be written as follows,

$$\begin{aligned} &\textbf{if} \quad z_n \left(I_n - c_{in}^{PI_H} \right) \geq 0, \quad \textbf{then} \quad a_n^{PI_H} = 0 \\ &\textbf{else} \quad \textbf{set} \quad a_n^{PI_H} \quad s.t. \quad z_n \left(I_n - c_{in}^{PI_H} + a_n^{PI_H} \right) = 0 \\ &\textbf{where} \\ &\qquad z_n \left(I_n - c_{in}^{PI_H} \right) = \max_{i \in M_n \times H_n^{PI_H}} \left(U_{in} \left(\boldsymbol{\phi}_i^{PI_H} \right) \right) - \max_{i \in M_n \times H_n^{NT}} \left(U_{in} \left(\boldsymbol{\phi}_i^{NT} \right) \right) \end{aligned} \tag{5.14}$$

If the individual net experienced utility is already positive, then she receives no refund; otherwise, she receives the amount of refund that makes her net experienced utility equal to 0. Since the nonlinear income effect of remaining income is modeled by a strictly monotonic quasi-concave function, which is continuous over the interval that $I_n - c_{in}^{PI_H} + a_n^{PI_H} > 0$, Equation 5.14 has a unique solution $a_n^{PI_H}$. Note that the condition $I_n - c_{in}^{PI_H} + a_n^{PI_H} > 0$ is guaranteed because of budget constraint on travel choices.

Note that the premise of achieving a Pareto improving outcome which satisfies the revenue refund constraint in system optimization is that the available revenue for refunds can cover the total losses. For homogeneous travelers, this is guaranteed as long as social welfare of pricing is positive, since homogeneous users face the same generalized cost at equilibrium [Arnott et al., 1994]. However, to the best of our knowledge, this property has not been shown to hold under user heterogeneity, elastic demand, nonlinear income effects in general networks. In the numerical experiments, we will examine this through sensitivity tests.

Alternatively, the objective of user optimization can be to maximize social welfare. Assuming other components in the utility specification do not change (which might not be true as the nonlinear income effect can influence choices), it makes sense to refund revenue to users who value it the most. In other words, low income users receive revenue refunds first, which implicitly also improves equity. The personalized refunds with this distribution rule is denoted as PI_S

An additional policy that needs to be optimized in the system optimization is the revenue distribution control parameter y_d, which determines who is eligible and how much they can get. The corresponding distribution rule can be written as,

$$\begin{aligned}
&\textbf{if} \quad MUI\left(I_n - c_{in}^{PI_S}\right) \leq y_d, \quad \textbf{then} \quad a_n^{PI_S} = 0 \\
&\textbf{else} \quad \textbf{set} \quad a_n^{PI_S} \quad s.t. \quad MUI\left(I_n - c_{in}^{PI_S} + a_n^{PI_S}\right) = y_d \\
&\textbf{where} \\
&\qquad MUI\left(I_n - c_{in}^{PI_S}\right) = \frac{\partial z_n\left(I_n - c_{in}^{PI_S}\right)}{\partial I_n} = 1 + \frac{\lambda}{\gamma + I_n - \bar{c}_{in}^{PI_S}} \\
&\qquad z_n\left(I_n - c_{in}^{PI_S}\right) = \max_{i \in M_n \times H_n^{PI_S}} \left(U_{in}\left(\boldsymbol{\phi}_i^{PI_S}\right)\right) - \max_{i \in M_n \times H_n^{NT}} \left(U_{in}\left(\boldsymbol{\phi}_i^{NT}\right)\right)
\end{aligned} \tag{5.15}$$

where $\bar{c}_{in}^{PI_S}$ represents the cost of the chosen alternative in PI_S (i.e. the cost of the alternative i with the maximum utility); $U_{in}\left(\boldsymbol{\phi}_i^{PI_S}\right)$ is expressed in Equation 5.1 and 5.4; $\max_{i \in M_n \times H_n^{NT}}\left(U_{in}\left(\boldsymbol{\phi}_i^{NT}\right)\right)$ represents the maximum utility of choice i for individual n in NT, which is a constant.

Similarly, it has a unique solution $a_n^{PI_S}$ when the income effect is nonlinear. Note that if the income effect is linear (i.e. utility coefficient of income $\lambda = 0$), this distribution rule does not apply anymore since every traveler has marginal utility of income as 1. Although there is no additional utility gain for poor people from receiving refunds (efficiency gain), it is still possible to improve equity by refunding revenue to poor people according to the individual's remaining income. Let PI_R denote the instrument with the distribution rule based on the remaining income. The corresponding distribution rule can be written as,

$$\begin{aligned} \textbf{if} \quad I_n - \bar{c}_{in}^{PI_R} \geq \hat{y}_d, \quad \textbf{then} \quad a_n^{PI_R} = 0 \\ \textbf{else} \quad \textbf{set} \quad a_n^{PI_R} \quad s.t. \quad I_n - \bar{c}_{in}^{PI_R} + a_n^{PI_R} = \hat{y}_d \end{aligned} \tag{5.16}$$

where $\bar{c}_{in}^{PI_R}$ represents the cost of the chosen alternative in PI_R; the revenue distribution control parameter $\hat{y}_d$ is in unit of dollars and can be optimized in the system optimization. This distribution rule also has an unique solution for everyone because the remaining income $I_n - c_{in}^{PI_R}$ is a strictly monotonic and continuous function.

In fact, it can be shown that the distribution rule in Equation 5.16 is equivalent to the the distribution rule in Equation 5.15 when the income effect is nonlinear given they have the same toll profile (i.e. $\bar{c}^{PI_R} = \bar{c}^{PI_S}$). From the Equation 5.15, the Equation 5.16 can be derived as follows,

1. **if** $MUI\left(I_n - c_{in}^{PI_S}\right) \leq y_d$, **then** $a_n^{PI_S} = 0$

   This is the case in the distribution rule 5.15 that the marginal utility income of an individual n is less or equal to the revenue refund control parameter y_d and she receives zero refund. By expanding $MUI\left(I_n - c_{in}^{PI_S}\right)$, the condition can be expressed as,

$$\begin{aligned} MUI\left(I_n - c_{in}^{PI_S}\right) &= \frac{\partial z_n\left(I_n - c_{in}^{PI_S}\right)}{\partial I_n} \\ &= 1 + \frac{\lambda}{\gamma + I_n - \bar{c}_{in}^{PI_S}} \leq y_d \end{aligned} \tag{5.17}$$

   The Equation 5.17 can be rearranged as follows,

$$\frac{\lambda}{y_d - 1} - \gamma \leq I_n - \bar{c}_{in}^{PI_S} \tag{5.18}$$

Let $\hat{y}_d = \frac{\lambda}{y_d - 1} - \gamma$, the Equation 5.18 implies **if** the remaining income $I_n - \bar{c}_{in}^{PI_R} \geq \hat{y}_d$, **then** $a_n^{PI_R}$ is the same as $a_n^{PI_R}$, which is equal to 0. Note that $\bar{c}_{in}^{PI_S}$ is assumed to be equal to $\bar{c}_{in}^{PI_R}$.

2. **else** **set** $a_n^{PI_S}$ *s.t.* $MUI\left(I_n - c_{in}^{PI_S} + a_n^{PI_S}\right) = y_d$

   This is the case in the distribution rule 5.15 that the marginal utility income of an individual n is greater than the revenue refund control parameter y_d and she receives a refund that makes her MUI equal to y_d. By expanding $MUI\left(I_n - c_{in}^{PI_S} + a_n^{PI_S}\right)$, the condition can be expressed as,

$$\begin{aligned} MUI\left(I_n - c_{in}^{PI_S} + a_n^{PI_S}\right) &= \frac{\partial z_n\left(I_n - c_{in}^{PI_S} + a_n^{PI_S}\right)}{\partial I_n} \\ &= 1 + \frac{\lambda}{\gamma + I_n - \bar{c}_{in}^{PI_S} + a_n^{PI_S}} = y_d \end{aligned} \tag{5.19}$$

   The Equation 5.19 can be rearranged as follows,

$$\frac{\lambda}{y_d - 1} - \gamma = I_n - \bar{c}_{in}^{PI_S} + a_n^{PI_S} \tag{5.20}$$

   Similarly, let $\hat{y}_d = \frac{\lambda}{y_d - 1} - \gamma$, the Equation 5.20 implies that $a_n^{PI_R}$ should be set to make the remaining income $I_n - \bar{c}_{in}^{PI_R} + a_n^{PI_R}$ be equal to $\hat{y}_d$, which is the same as $a_n^{PI_S}$ since $\bar{c}_{in}^{PI_S}$ is equal to $\bar{c}_{in}^{PI_R}$.

Hence, when the income effect is nonlinear, the distribution rule in Equation 5.16 can be derived from the distribution rule in Equation 5.15 by having $\hat{y}_d = \frac{\lambda}{y_d - 1} - \gamma$. Similarly, it is straightforward to show this in the opposite direction since the MUI

function is monotonic and continuous. However, when the income effect is constant, the distribution rule in Equation 5.15 fails as everyone's MUI is equal to 1. Therefore, the distribution rule based on the remaining income in Equation 5.16 dominates the other one. Henceforth, the social welfare maximization distribution rule refers to Equation 5.16 and the corresponding instrument is denoted as PI_S.

Finally, it is possible to combine the Pareto improving distribution rule in Equation 5.14 and the social welfare maximization distribution rule in Equation 5.16 to have a hybrid distribution rule, which ensures Pareto improvement first and if there is some revenue left over, it can be distributed to maximize social welfare. As a result, the performance of this hybrid rule can be expected to dominate the pure Pareto improving distribution rule in terms of both efficiency and equity.

5.3.4 MI Optimization

For every traveler, assuming other components in the utility specification do not change, if the market value of token allocations in TMC is equal to the dollar value of revenue refunds in pricing, she is supposed to have almost the same behavior given the losses due to the transaction fees are minimal. This is because the income effects are similar. This implies that we can utilize the solution of the pricing with personalized refunds PI for the TMC with personalized token allocations MI. Assume the equilibrium token price is \$1, the system optimization solution —toll profile in dollars, can be converted to toll profile in tokens for MI; while the user optimization solution —personalized refunds, can be converted to the personalized token allocations. It is demonstrated in Chapter 6 that MI performs almost the same as PI in terms of the social welfare, equity, behavior change, level of congestion, and distributional impacts. It requires further investigations to solve the optimization problem (i.e. the toll profile and personalized token allocations) of MI directly without relying on the solution of PI. It might not be straightforward since the token price in MI is determined endogenously to balance the demand and supply of tokens, which is dependent on the token allocations and the toll profile. In other words, there is an issue of non-uniqueness (during optimization) of combinations of the personalized token allocations and the toll profile.

5.4 Summary

This chapter first summarizes relevant literature about bi-level optimization and personalized pricing in transportation. Next, the online bi-level optimization framework Tri-POP is introduced and the detailed methodology on bi-level optimization for personalized distribution is developed. Two different distribution rules are developed for personalized distribution; the first maximizes social welfare and the second achieves Pareto improvement first and then maximizes social welfare with the remaining revenue.

Chapter 6

Numerical Experiments

6.1 Experiment Setup

In order to perform credible comparisons of the various tolling instruments, it is important to have realistic parameters and input data. In the following sections, data and parameters of the demand model, supply model, and the day-to-day learning model will be introduced for simulating a realistic base case. Some of the key parameters will be varied in later experiments.

6.1.1 Data and Model Parameters

To begin with, the demand model requires both choice attributes and individual characteristics as input data. For individual characteristics, this includes disposable income I_n and preferred arrival time $\hat{t}_n$. Recall that disposable income I_n defined in this study is personal net income after taxes and subtracting necessary living expenses (e.g. housing, health, food). In other words, it represents the individual's available income for traveling. The individual pre-tax annual income is assumed to be a log-normal distribution fitted using the Integrated Public Use Microdata Series (IPUMS) 2019 census data [Ruggles et al., 2021]. The cumulative distribution function (CDF) of IPUMS data and fitted data are shown in Figure 6-1. The x-axis represents individual pre-tax annual income in units of one thousand dollars and the y-axis represents

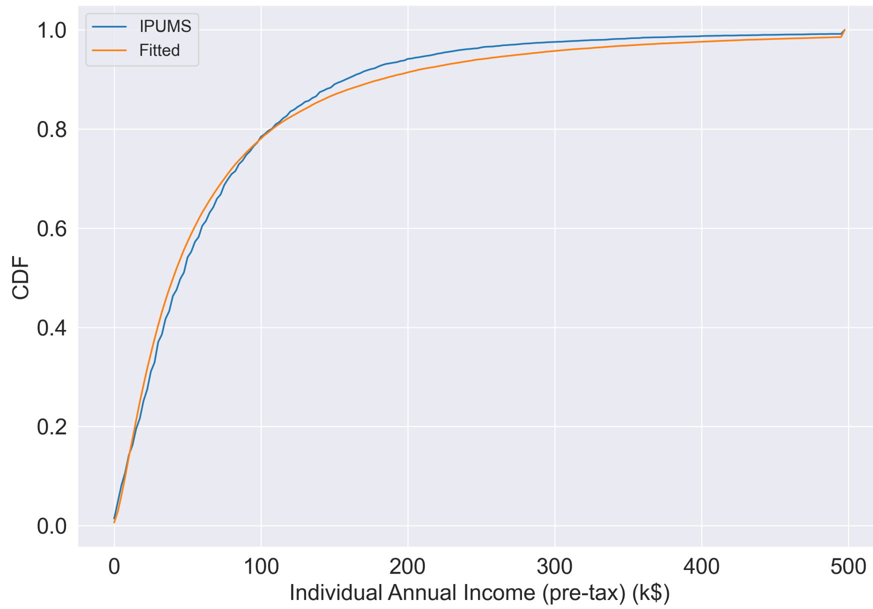


Figure 6-1: Individual pre-tax annual income distribution

cumulative probability values. Note that all annual incomes greater than 500 thousand dollars are grouped together. As we can see, the lognormal distribution fits the income distribution quite well.

Individual daily income can be acquired using annual income divided by 260 working days per year. The individual hourly wage rate can be acquired by dividing daily income by 8 working hours per day. The minimum wage rate is set to $7.25 provided by the U.S. Department of Labor. However, it is not straightforward to get disposable income after taxes because it depends on individual income and other attributes. Further, individual necessary living expenses could vary based on income and there is limited disaggregate data on this. According to data from the Bureau of Labor Statistics, the average pretax household income in the United States in 2019 was $82,852, while average household expenditures except transportation added up to $52,294, which means that the average American has 63% of their pretax or market income available for transportation and other expenditures [U.S. Bureau of Labor Statistics, 2019]. Hence, in this study, it is assumed that each traveler's daily

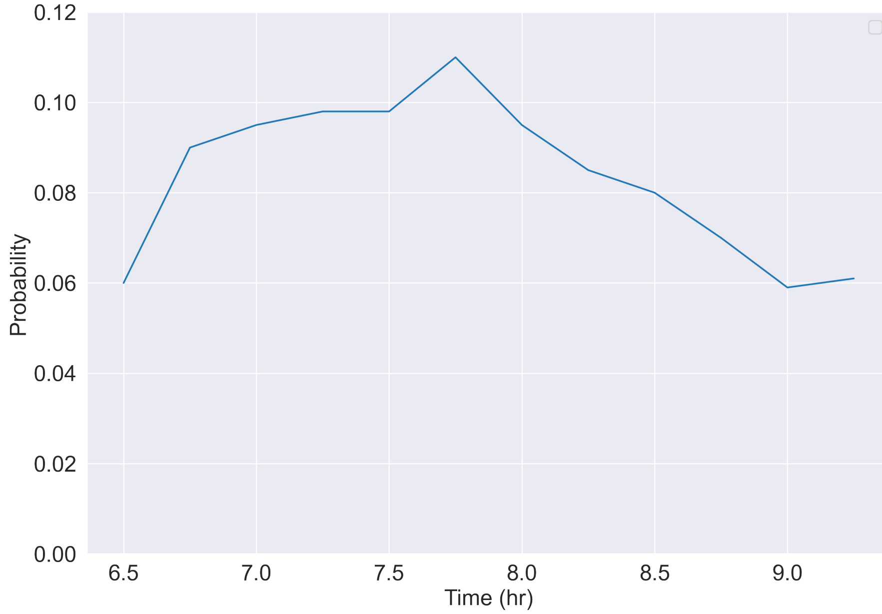


Figure 6-2: Individual preferred departure times distribution

disposable income after taxes and necessary living expenses is equal to 60% of their pre-tax daily income.

The preferred arrival time $\hat{t}_n$ represents the optimal or desired arrival time. The preferred departure time is just a shift of preferred arrival time by free flow travel time. Many researchers have assumed it follows a uniform distribution [De Palma and Lindsey, 2002a]. A more recent study estimated preferred departure of morning road users in Stockholm [Kristoffersson and Engelson, 2018]. The distribution of preferred departure times estimated from their study is adopted in this study and is shown in Figure 6-2. The x-axis represents departure time in hours and the y-axis represents probability values. As we can see, it does not follow the uniform distribution exactly.

For simplicity, the size of the preferred arrival window Δ_a is set as 0, which implies that individuals have a single preferred arrival time as in the standard Vickrey model. The departure time window size parameter η is set to 30, which means the individual departure time choice set H_n ranges over a 60-minute interval.

From empirical evidence (e.g. [Small et al., 2005]), values of time differs substantially over drivers. In this study, individual values of time α_n are calculated as one thirds of their wage rate [White, 2016]. Therefore, they are perfectly correlated with income. However, this will be relaxed and different levels of heterogeneity will be evaluated in experiments. For bottleneck models of congestion, schedule delay costs are another important part of congestion costs. It is likely that values of schedule delay early β_{En} and late β_{Ln} are also distributed across individuals. Because of lack of empirical data, literature on bottleneck models incorporate heterogeneity by making assumptions on ratios between values of schedule early/late to values of time.

As summarized in [Van den Berg and Verhoef, 2011, Van Den Berg and Verhoef, 2011], there are mainly two types of heterogeneity for bottleneck models. The first type of heterogeneity is termed proportional heterogeneity and first considered by [Vickrey, 1973]. It assumes values of time and schedule delays vary proportionally. In other words, the ratios of parameters are identical. A different case studied by [De Palma and Lindsey, 2002b] assumes values of schedule delays are fixed and only values of time are distributed. As a result, ratios of parameters are distributions. This type of heterogeneity is known as ratio heterogeneity. Behavioral interpretations of these two types can be found in [Van Den Berg and Verhoef, 2011]. Further, [Van Den Berg and Verhoef, 2011] considers a more general heterogeneity, assuming the ratio of values of time to values of schedule delay early follows a symmetric triangular distribution from 1 to 3 based on intuition and ratio of values of schedule delay late to values of schedule delay early is a constant 3.9 based on [Arnott et al., 1990b].

In this study, the ratio of values of schedule delay early β_{En} to values of time α_n is assumed to follow a triangular distribution from 0.1 to 1 with a mode at 0.5. The ratio of values of schedule delay late β_{Ln} to α_n is assumed to follow a triangular distribution from 1 to 3 with a mode at 2. The modes are selected as 0.5 and 2 respectively based on empirical relationships that β_{Ln} is twice large as α_n and α_n is twice large as β_{En} [Small, 2012]. The bounds are set based on empirical relationships that $\beta_{En} \leq \alpha_n \leq \beta_{Ln}$ [Small, 2012]. As pointed out by [Small, 2012], waiting times are onerous compared to in vehicle times by multiples of two to three by most assessments. For simplicity, the ratio of values of time α_n to values of waiting time β_{Wn} is assumed to be a constant equal to 3.

Regarding utility coefficients of income effect λ and γ, λ measures the strength of nonlinear income effect and γ ensures the logarithmic function is defined when remaining income is equal to 0. In this study, γ is set to 2 and λ is calibrated to be 3 to have the highest marginal utility of income to be less than 1.34 [Layard et al., 2008].

For the scale parameter μ_n, it is known that it is both confounded with the systematic utility as well as being inversely related to error variance within the choice data [Ben-Akiva et al., 1985]. As pointed out in the literature (e.g. [Louviere and Eagle, 2006]), the modeled heterogeneity can come from heterogeneity in individual coefficients and scale heterogeneity that is shared across coefficients. In this study, the distribution of the scale parameter is assumed to be a lognormal distribution as the scale parameter has to be greater than 0. Also based on realistic judgements, the coefficient of variation of scale parameter is set as 0.5. The mean of scale parameter is calibrated based on price elasticity.

Since coefficients of mobility demand models are distributed, it requires simulation to obtain price elasticity. Each iteration consists of two simulation runs. Run 0 is the base case and run 1 is the case with operation costs and peak hour toll (7 to 8AM) increased by 5% to calculate peak hour price elasticity. Several iterations are done for different initial toll levels. For simplicity, flat toll profiles are used from 6:30 to 9:30AM.

From the literature, the aggregate elasticities of peak hour travel vary greatly from case to case as they are dependent on the model structure, physical environment, activity type, initial toll levels and many other factors. [Ding et al., 2015] estimated the elasticity of departing during the peak in Washington D.C. is -0.0906 for driving alone. [Sasic and Habib, 2013] showed that the elasticity of departing in AM peak for work trips in Toronto is between -0.067 to -0.12 by car mode. [Holguin-Veras et al., 2005] found price elasticity of using crossings (tunnels and bridges) in NYC range from -0.31 to -1.97 for weekdays depending on the time-of-day. When there is no initial toll, the price elasticity just represents fuel elasticity. [Lipow, 2008] estimated

Table 6.1: Price elasticities across income groups by toll levels

Toll	$\leq 25\%$	25% to 50%	50% to 75%	75% to 90%	$90\% \leq$	Total
0	-0.34	-0.29	-0.12	0.00	0.00	-0.19
2.5	-1.14	-0.59	-0.10	-0.04	-0.03	-0.38
5	-1.57	-1.07	-0.20	-0.09	-0.06	-0.53

fuel elasticity as -0.17 and [Gillingham, 2014] estimated fuel elasticity in California as -0.15.

Based on evidence from literature, the mean of scale parameter is determined to be 0.5. The corresponding price elasticities across different income groups and initial toll levels are presented in Table 6.1. As we can see, low income users are more sensitive to price than high income users. Also, when there is no toll, the aggregated price elasticity is similar to empirical fuel elasticity. As the toll level increases, the aggregate price elasticity also increases and becomes closer to empirical values found in [Holguin-Veras et al., 2005].

Regarding the supply model, attributes of car and public transit (PT) have to be specified. The free flow speed of car is set to be 45mph [Ali et al., 2007] and the one way driving distance is assumed to be 18 miles. Thus, the free flow travel time is calculated to be 24 minutes. The operation cost of driving is assumed to be fuel cost only, which is equal to \$3.13 (driving distance (18 miles) times 1/23 gallon per mile times 4 dollars per gallon). For public transit, based on the New York City MTA data, the fare is 2-dollar; average speed is 25mph; and headway is 10 minutes. The PT distance is also assumed to be 18 miles, and the resulting PT travel time is 43 minutes since both headway and travel time of PT are constant. The expected waiting time is 5 minutes.

The bottleneck capacity s is determined based on travel time index (TTI) calibration. The TTI represents the ratio between actual travel time and free flow travel time. For instance, a TTI value of 1.2 means actual travel time is 20% longer than free flow travel time. In this study, the base case capacity is selected as 2340 vehicles per hour to have a reasonable level of congestion with TTI as 1.68 [Chen, 2010] under the no toll (NT) scenario.

As mentioned in Chapter 4, an exponential smoothing filter is adopted to update travel time information and individual departure time. The greater the learning weights are, the more unstable the system becomes [Cantarella and Cascetta, 1995]. In this study, the learning weights θ_τ and $\theta_{\tilde{t}}$ are selected as 0.1.

Recall that we focus on the morning commute and hence, we simulate half of a day (12 hours) with a simulation time interval Δ_t of 1-minute, yielding 720 time intervals, $\bar{t} = 0 \ldots 719$. The market elements (allocation, expiration, and price adjustment) and trading behavior are also simulated for the first half. The second half of a day is assumed to be a mirror of the first half. The departure time interval (Δ_h) is assumed to be 5 minutes. The total population number N is 7500. Descriptions and values of key parameters are summarized in Table 6.2.

Table 6.2: Model and simulation parameters

Variables	**Description**	**Values**
N	Population	$7,500$
Δ_t	Duration of a simulation time step	1 min
Δ_h	Duration of a departure time interval	5 min
Δ_a	Size of desired arrival window	0 min
η	Departure time window size parameter	30
λ	Coefficient of nonlinear income effect	3
γ	Nonlinear income effect adjustment parameter	2
s	Bottleneck capacity (per min)	39
t_0	Free flow travel time	24 mins
c_f	Operation cost of car	\$3.13
τ_{PT}	PT travel time	43 mins
W_{PT}	Expected waiting time	5 mins
c_{PT}	Operation cost of PT	\$2
$\theta_\tau/\theta_{\tilde{t}}$	Learning weights	0.1

6.1.2 Existence and Uniqueness of Equilibrium

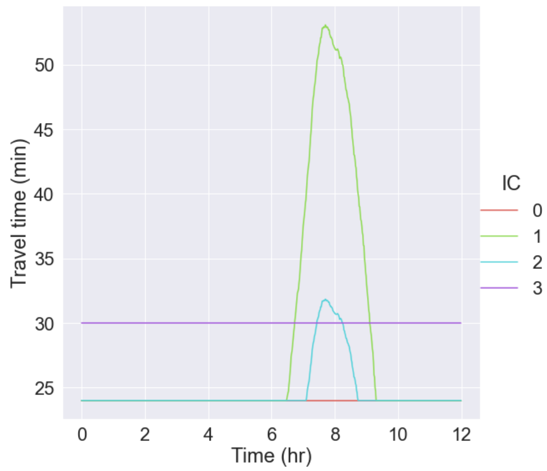
In this section, we examine existence and uniqueness of properties of equilibrium in the day-to-day dynamic model. It is important to check whether an equilibrium exists since the analysis and comparisons are meaningless when system states are still changing. Once equilibrium existence is confirmed, it is important to confirm uniqueness of equilibrium to avoid issues with multiple equilibria. As pointed out by [Lindsey, 2004], some problems include the questionable validity of comparative statistics under multiple equilibria, the possible instability of equilibrium, the difficulty of determining which equilibrium will prevail, and sensitivity to starting values.

In the literature, existence and uniqueness of equilibrium in the bottleneck model have been widely studied (e.g. [Hendrickson and Kocur, 1981, Hurdle et al., 1983, Smith, 1984, Daganzo, 1985]). However, they limit the scope to heterogeneity in desired arrival time. [Lindsey, 2004] considers a more general heterogeneity in values of time, desired arrival times, and values of schedule delay. He established conditions for existence and uniqueness of a deterministic departure time user equilibrium in the bottleneck model. [De Palma et al., 1983] extend considerations to stochastic user equilibrium but travelers are assumed to have identical systematic travel cost functions. They prove the equilibrium departure rate is unique but their findings cannot be directly applied to our model as we consider heterogeneity.

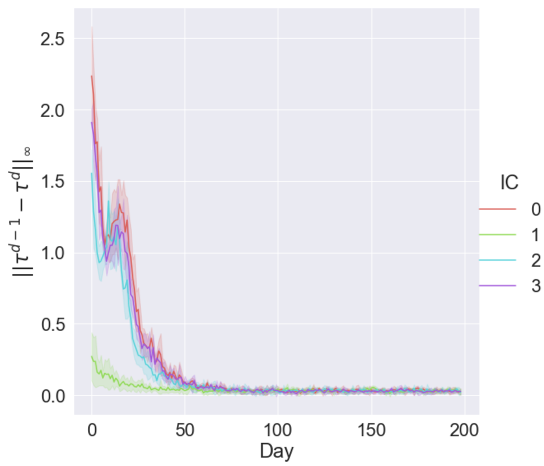
[Miyao and Shapiro, 1981] establish conditions for the existence, uniqueness and stability of equilibrium for discrete choice models. Although their framework is relatively general, it is once again not compatible with our model because they consider individuals with the same choice set whereas in our case, choice sets (departure time choices) are individual specific. To the best of our knowledge, there are no analytical results on uniqueness of the equilibrium for our model. However, simulations with different initial conditions suggest that the equilibrium exists and is unique.

Using data and parameters discussed in the previous section, a No Toll (NT) scenario with different sets of initial travel time information of driving is simulated. The population size N is 7500. Because the simulation is stochastic, five replications with different initial seeds for random number generation are performed for each initial travel time of driving condition.

The four different sets of initial travel time information of driving are plotted in Figure 6-3. Note that the initial travel time information of driving serves as the basis for the departure time decisions in day 0 of the simulation. The initial travel time



(a) Initial travel time of driving information



(b) Convergence of travel time of driving

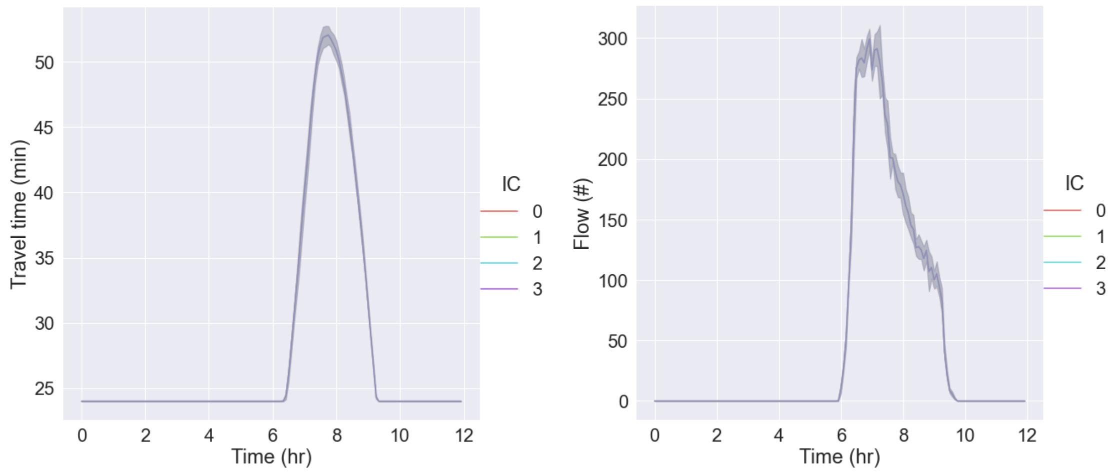
Figure 6-3: Various initial travel time information of driving and corresponding convergence

set 0 of driving represents free flow travel times of driving across the entire day; the initial travel time set 1 of driving is an equilibrium travel time of driving from the simulation using the initial travel time set 0 for a particular random seed; the initial travel time set 2 of driving is generated as 0.6 of the travel time set 1; finally, the initial travel time set 3 of driving represents a constant 30 minutes travel time of driving across the entire day.

Travel time of driving is used as a representative measure of system state partly because it is central to the day-to-day learning process of travelers (alternatively, departure flows could also be used). The infinity norm (also known as supremum norm) of day $d-1$'s travel time vector of driving and day d's travel time vector of driving is calculated and used as a measure of convergence across days. The expression of this norm can be written as follows:

$$||\boldsymbol{\tau}^{d-1} - \boldsymbol{\tau}^{d}||_\infty = \sup\{|\tau_i^{d-1} - \tau_i^{d}| : i \in \{m = C, h|h \in \mathcal{H}\}\} \tag{6.1}$$

where $\boldsymbol{\tau}^{d-1}$ is the vector of travel times of day $d-1$ defined over a set of time intervals $\mathcal{H}$. The corresponding convergence of travel time of driving is plotted in

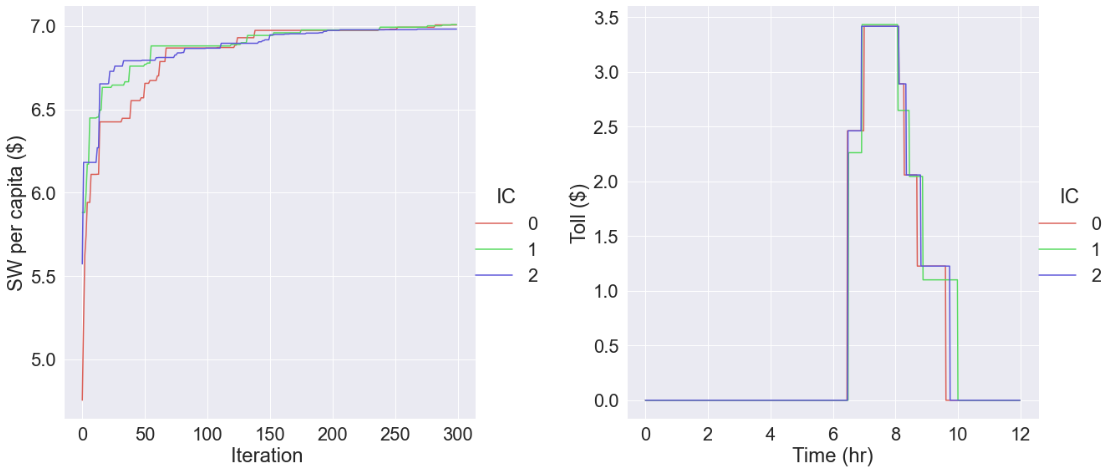


(a) Travel time of driving at equilibrium (b) Flow rate of driving at equilibrium

Figure 6-4: Travel time and flow of driving at equilibrium with different sets of initial travel time information

Figure 6-3. The lines in the plot represent averages over different random seeds and bands are plotted based on the standard deviations. As we can see from the plot, the infinity norm converges to zero by about 50 days, which implies the point wise maximum difference between the travel time vector of day $d-1$ and travel time vector of day d converges to 0 according to the definition of infinity norm. This suggests that simulations with different initial travel time sets have all converged in terms of travel times. The convergence of green line does not start from 0 because the initial travel time set 1 of driving is an equilibrium travel time of a particular random seed.

However, this still does not establish whether the equilibrium is unique or not. Thus, the travel time vectors and flow vectors of driving at equilibrium are plotted in Figure 6-4. As we can see, simulations with different initial travel times of driving converge to the same travel times and departure flow rate at equilibrium. As only two mode choices are considered and the travel time and headway of PT are constant, once the travel time of driving converges, the resulting departure flow of PT also converges.



(a) Convergence of the social welfare (b) Optimal toll profile

Figure 6-5: Convergence of the social welfare and optimal time-of-day step toll profile of $P-$ by initial values

6.1.3 Optimization Performance

As discussed in Chapter 4, the system optimization for the time-of-day step toll profile is solved using a type of metaheuristic algorithm known as Differential Evolution (DE). In the literature, metaheuristic algorithms have been shown to work well for nonconvex and nonlinear toll design problems (e.g. [Shepherd and Sumalee, 2004, Zhang and Yang, 2004]). They are fairly robust to initial values because of mutation, crossover and selection steps in the algorithm.

In this study, the population size of the DE algorithm NP is set to 15, which means 15 candidates of the time-of-day step toll profile are evaluated in one iteration. Taking advantage of parallelization, they can be simulated and evaluated concurrently. With data and parameters discussed in the previous section, a pricing without distribution ($P-$) case is used to test the performance of the optimization algorithm. Three different initial populations (with 15 candidates) are used for the algorithm to optimize the time-of-day step toll profile.

The convergence of the optimization objective (social welfare) along with the optimal time-of-day step toll profile with three different initial populations are shown in Figure 6-5. The total number of iterations is 300 but considering 15 candidates are evaluated in each iteration, a total of 4500 candidates are evaluated. Recall that social welfare is equal to the sum of user benefits relative to NT and regulator revenue. As we can see, although the starting social welfare is very different because of different initial values, they converge to the same social welfare (within a tolerance of $0.01) after about 150 iterations. The optimal toll profiles are also very similar with small differences.

6.2 Market Design for Tradable Mobility Credits

In Chapter 3, we discuss various market elements (e.g. allocation, expiration, transaction fees, price adjustment and market rules governing trading) and market behavior for TMC to manage travel demand and avoid profiteering. In this section, numerical experiments are conducted to demonstrate the performance of TMC and compare it to pricing. We first introduce market setup in this study. Next, we demonstrate the existence and uniqueness of TMC system equilibrium. Then, we inspect the market behavior of travelers. Finally, we examine the robustness of the two instruments under sub-optimal toll profiles.

6.2.1 Market Setup

The token allocation rate is an exogenous variable determined by the regulator. Given a toll profile, the more tokens the regulator distributes, the smaller the token market price is at equilibrium and vice versa. For an extreme case that everyone has enough tokens to pay the toll, the token price would be zero as there is no demand for buying tokens. Hence, it is important that the regulator specifies an appropriate allocation amount.

We determine the amount of total daily allocation by dividing the regulator revenue of pricing without distribution (K^{P-}) by 1 dollar assuming the token market price at equilibrium is 1 dollar. Next, we re-optimize the toll profile in tokens given the determined allocation. The optimization performances with different allocation rates are examined in the next section.

Experiments in this section assume tokens are distributed uniformly to everyone (MU) while experiments in next section looks into personalized distribution. Also, continuous allocation is the default method for token allocation because of its advantages discussed in Chapter 3, which is also demonstrated later in the case of sub-optimal toll profiles. Every traveler is assumed to have a random account balance at the beginning of the simulation. The effect of initial account balance is examined in the following section 6.2.4.

Since the base case $P-$ has revenue as 4.1040 dollars per capita over the entire day, the individual allocation rate r is 0.00285 tokens per minute. The token lifetime L should be one day since the context we consider is a daily commute problem and the frequency of trip making is one day. However, since we only simulate half day, the lifetime L is 720 minutes.

The effect of both proportional and fixed transaction fees are investigated later in this section. In all prior experiments they are assumed to be 0. The token price is adjusted daily by a constant based on the demand and supply of tokens, which is represented by regulator revenue of MU directly K^{MU}. The greater the revenue is, the more demand for buying tokens and price increases and vice versa. The initial price (p^0) is by default set at 1 dollar but is varied in later experiments. The constant price change Δp is 5 cents (0.05 dollars). To avoid unnecessary price fluctuations for small amounts of regulator revenue, the price does not change when revenue is in the interval $[-K_t, K_t]$, where K_t is equal to 300 dollars in this study (based on preliminary experiments). In addition, the price cannot fall below 0 but there is no maximum cap for it.

Descriptions and values of key parameters are summarized in Table 6.3.

6.2.2 Optimization Performance

The convergence of the optimization objective values (social welfare) along with the optimal time-of-day step toll profile in tokens and in dollars (tokens times token price) with three different allocation rates varied from 15% less than the baseline to 15% more than the baseline are shown in Figure 6-6.

As we can see, although the starting social welfare are very different because of

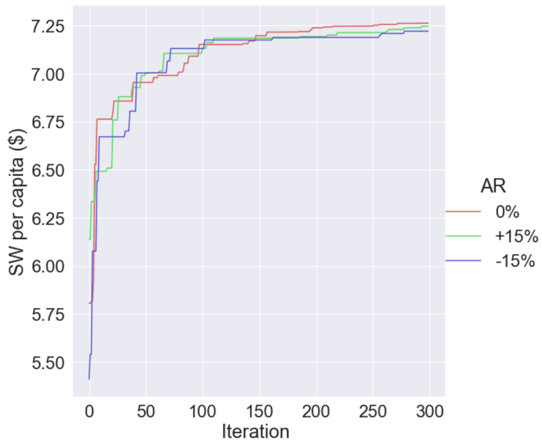
Table 6.3: Market parameters for tradable mobility credits (base case)

Variables	**Description**	**Values**
r	Token allocation rate	0.00285 per min
L	Token lifetime	720 mins
F_B^F/F_S^F	Proportional transaction fee of buying/selling	0
F_B^P/F_S^P	Fixed transaction fee of buying/selling	\$0
p^0	Initial token price	\$1
Δp	Price change	\$0.05
K_t	Regulator revenue threshold	\$300

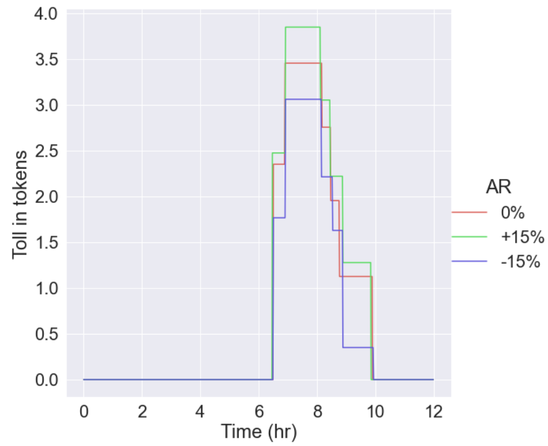
different allocation rates, they converge to the same social welfare (within a tolerance of \$0.02) at the end of 300 iterations. Also, the social welfare of TMC is greater than that of pricing without distribution.

The token price of the baseline allocation rate is \$1 while the lower allocation rate has the token price equal to \$1.1 and the higher allocation rate has the token price equal to \$0.9. This is consistent with our expectations that the lower allocation rate leads to the higher token price due to less supply and vice versa. The optimal toll profiles in tokens in Figure 6-6b show that the lower allocation rate leads to the overall higher toll profile in tokens and vice versa. This is because as the token price increases (decreases) the toll in tokens has to decrease (increase) to maintain the product of the price and toll in tokens similar to the optimal toll in dollars.

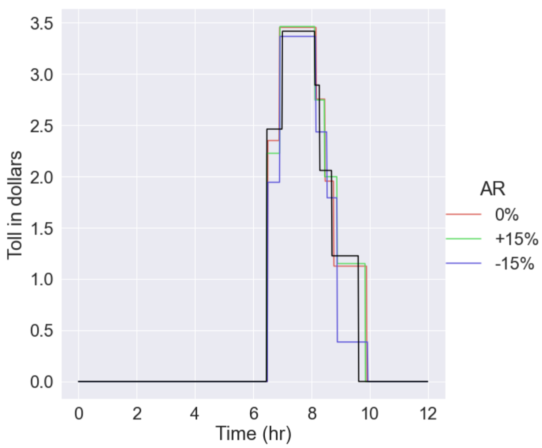
The product of toll profiles in tokens and token prices are shown in Figure 6-6c. The black line represents the optimal toll profile in dollars for $P-$, which is also the red line shown in Figure 6-5a. In the peak hour (from 7AM to 8:30AM) when the most people are traveling, the toll profiles in dollars are very similar with small differences. The differences among the toll profiles in the off peak indicate that the objective function (social welfare function) might have the relatively flat landscape near the optimal values.



(a) Convergence of the social welfare



(b) Optimal toll profile in tokens

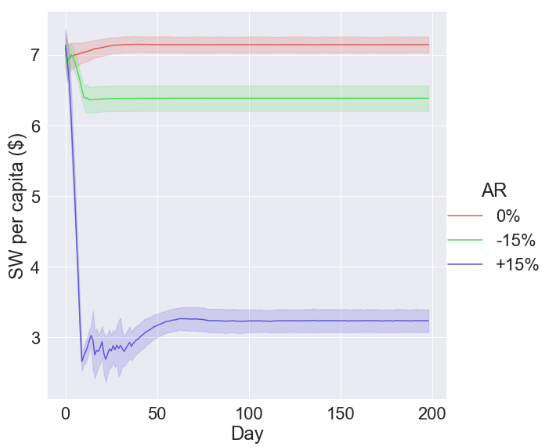


(c) Optimal toll profile in dollars

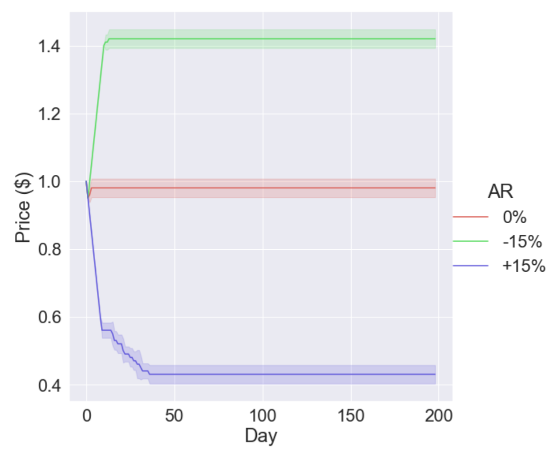
Figure 6-6: Convergence of the social welfare, optimal time-of-day step toll profile in tokens and in dollars of MU by three allocation rates

6.2.3 Existence and Uniqueness of Equilibrium

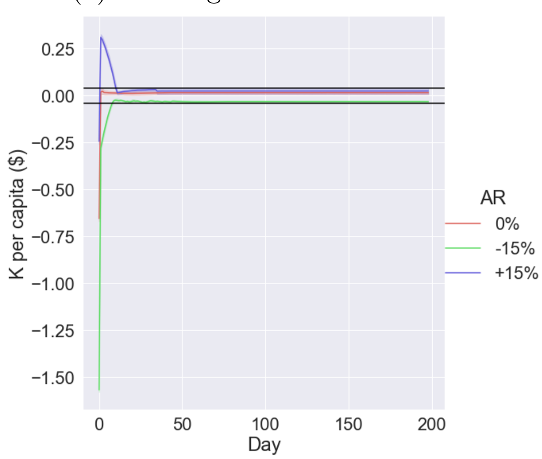
We first present the effect of various allocation rates r on convergence of token price, social welfare, regulator revenue, and travel time of driving in Figure 6-7. As we can see, different allocation rates lead to different social welfare and price values. At the baseline allocation rate (recall that this is computed based on an optimized toll in dollars from pricing without distribution), token price of MU (TMC with uniform allocation) converges to \$1. With smaller allocation rate, token price converges to be higher than \$1 due to more demand and less supply; with greater allocation rate, token price converges to be less than \$1 due to less demand and more supply. The social welfare of greater or smaller allocation rates are less than that of the baseline allocation rate. The social welfare of the baseline allocation rate is greater than that of pricing without distribution. The regulator revenues under the three allocation rates converge to be within the regulator revenue threshold band (the black lines) as shown in Figure 6-7c. Travel times of driving under the three allocation rates converge too as shown in Figure 6-7d.



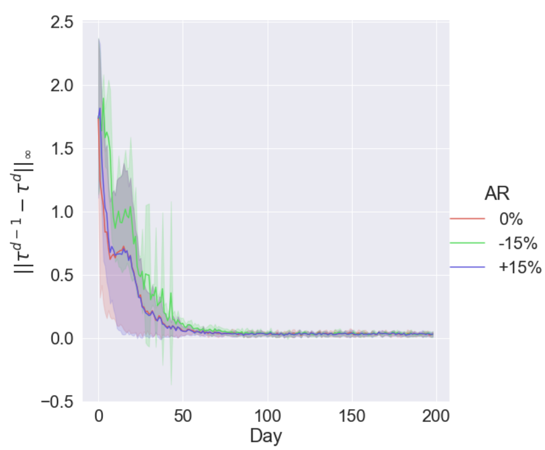
(a) Convergence of social welfare



(b) Convergence of token price



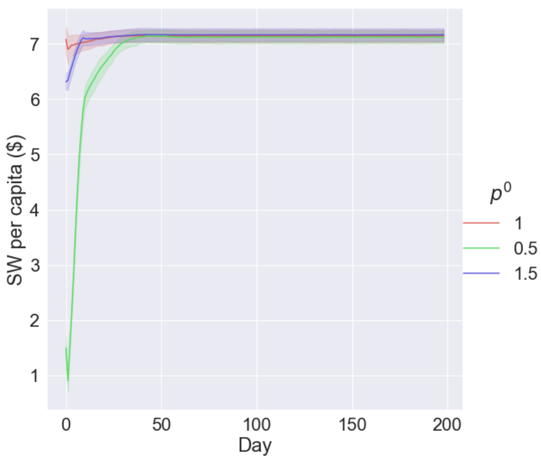
(c) Convergence of regulator revenue



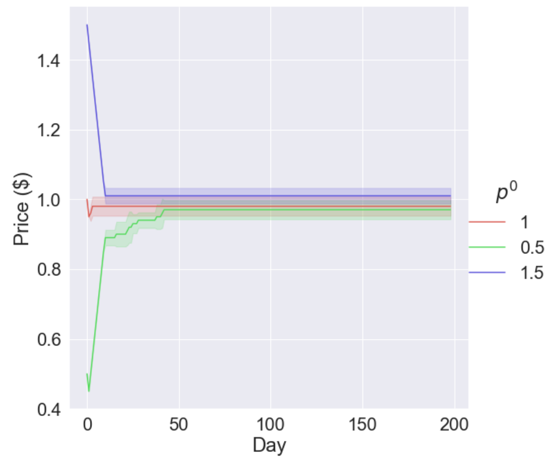
(d) Convergence of travel time of driving

Figure 6-7: Convergence of social welfare, token price, regulator revenue, and travel time of driving by allocation rate r

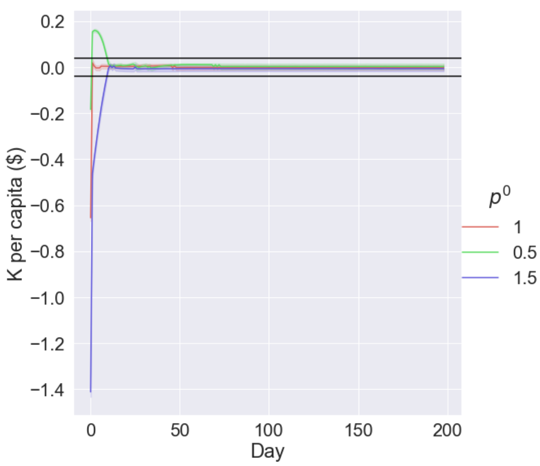
Next, the effect of various initial market prices p^0 on convergence of token price, social welfare, regulator revenue, and travel time are examined in Figure 6-8. The price and social welfare converge to values that are not statistically significantly different at a significance level of 0.05 no matter what the initial price is. The regulator revenues under the three initial prices converge to be within the regulator revenue threshold band (the black lines) as shown in Figure 6-8c. Travel times under the three initial prices converge too as shown in Figure 6-8d.



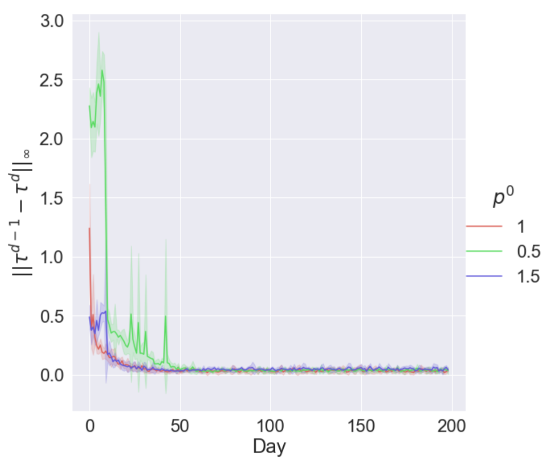
(a) Convergence of social welfare



(b) Convergence of token price



(c) Convergence of regulator revenue



(d) Convergence of travel time of driving

Figure 6-8: Convergence of social welfare, token price, regulator revenue, and travel time of driving by initial price p^0

6.2.4 Trading behavior

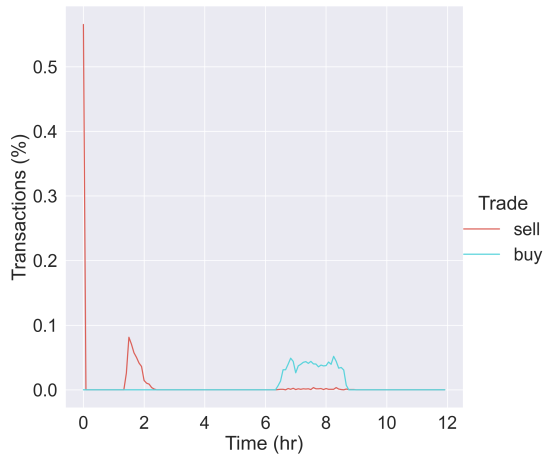
We present the effect of full initial account balances and random initial account balances on the transaction numbers and amount by time-of-day at equilibrium. The plots are based on simulations with a particular random seed because stochasticity can make visualizations hard to interpret but findings and insights from the plots are general.

In Figure 6-9a, the numbers of buying and selling transactions in percentage of corresponding total numbers by time-of-day at equilibrium for full initial account balances are plotted. As we can see, buying transactions only happen in the peak hour because travelers can only buy tokens at time of traveling if they are short of tokens, which is consistent with the buying rule. While selling transactions happen at the beginning of the day, in the early morning, and peak hour, which can be explained by the plot of the average transaction amount by time-of-day for full initial account balances in Figure 6-9b. For travelers sell at the beginning of the day, all of them sell at full wallets as shown in Figure 6-9b (2.052 tokens as equilibrium token price is $1) because they travel in the off peak and do not need to use tokens; for travelers sell in the early morning (around 2AM), their account balances at time of selling are not full because their future token allocations until their departure times can cover their toll and it is optimal for them to sell now; for travelers sell in the peak period, they sell at full wallets because their account balances reach full after paying small toll charges. The selling behavior is consistent with the derived selling strategy but the excessive trading the beginning of the day contradicts the motivation of continuous allocation.

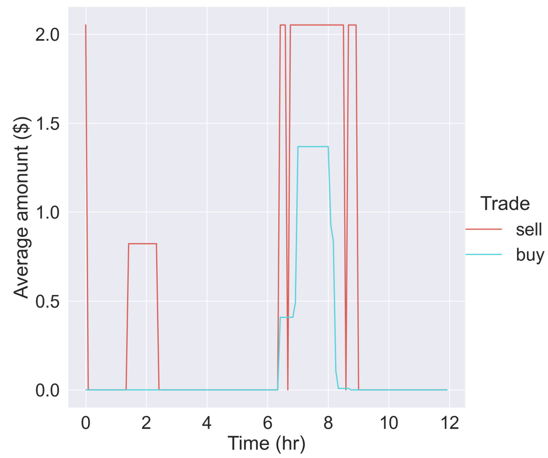
In practice, travelers will participate in the program in different times. As a result, their account balances at the beginning of the day will be different. Assume their initial account balances are distributed uniformly between 0 and the maximum account balance (2.052 tokens). As shown in Figure 6-9c, the selling transactions are spread across the day with a peak in the early morning (around 2AM) because of the same reason explained above. Apart from these travelers who sell in the early morning not at full wallets, other travelers sell only at full wallets as shown in in Figure 6-9d, which are desirable.

6.2.5 Effect of Transaction Fees

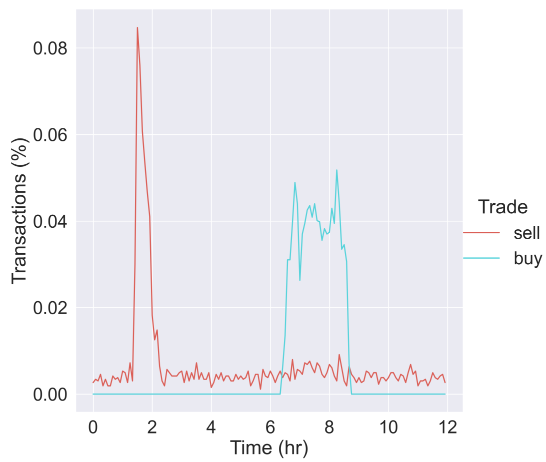
In the literature, a transaction fee is argued to prevent undesirable market behavior like frequent selling. For example, [Brands et al., 2020] apply a small transaction fee



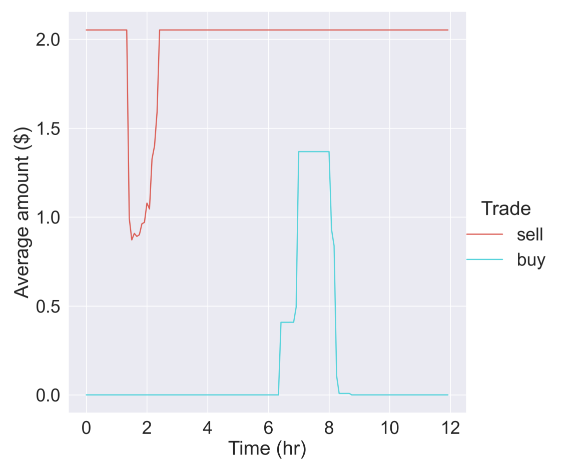
(a) Transaction numbers of full wallets



(b) Avg. trading amount of full wallets



(c) Transaction numbers of random wallets



(d) Avg. trading amount of random wallets

Figure 6-9: The effects of full and random initial account balances on the transaction numbers and amount by time-of-day at equilibrium

of 0.01 euro to prevent frequent selling in their experiment. However, it has also been shown that transaction fees could reduce system efficiency [Nie, 2012].

Our analysis in Chapter 3 shows that the effect of a fixed transaction fees is to prevent multiple transactions while the effect of a proportional transaction fees is to make one sell as soon as possible when the conditional profit is positive (if buying is required at the time of the next trip).

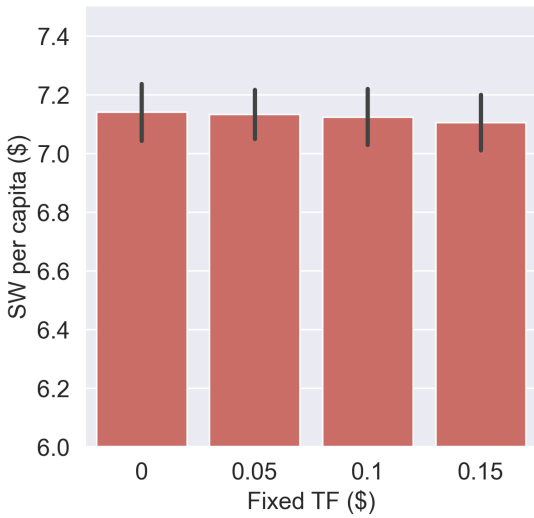
Numerical experiments in this section examine the effect of proportional and fixed transaction fees on social welfare and undesirable behavior. Specifically, undesirable behavior is defined as buying back tokens sold previously. This is because we would like to have users strictly being either sellers or buyers (not both). Sellers are the ones who travel in the off peak and sell their tokens while buyers are the ones who are willing to pay a high toll to travel in the peak. Optimal transaction fees are determined to eliminate buyback behavior while yielding the least efficiency loss.

For simplicity, the fixed transaction fees of buying and selling are varied together with the proportional transaction fees set to zero and vice versa. The effects of fixed and proportional transaction fees on social welfare and buyback behavior are shown in Figure 6-10. From the simulation experiments, a small fixed transaction fees (5 cents in this study) is seen to be able to eliminate buyback behavior in Figure 6-10c and reduce welfare only slightly in Figure 6-10a. While there are higher social welfare losses in the case of the proportional transactions fees in Figure 6-10b, which is also less effective in reducing buyback behavior in Figure 6-10d.

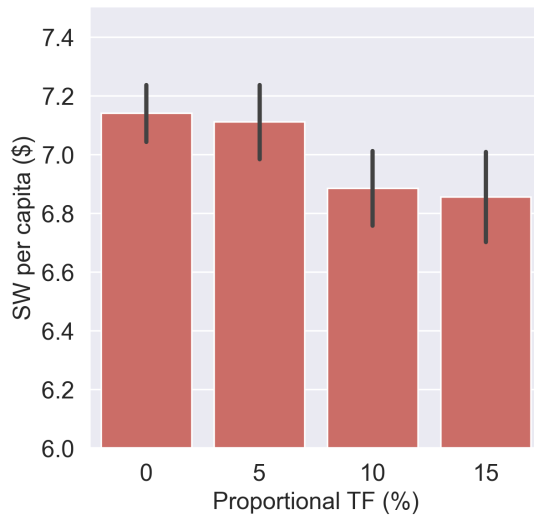
6.2.6 Sub-optimal Toll Profile

In practice, toll profiles may often be sub-optimal because of changing conditions, forecast errors and uncertainty. It is difficult to update these toll profiles (especially at the network level) regularly in practice since travelers need time for planning. For example, Singapore updates the ERP scheme once every three months. However, some market elements of the TMC scheme (e.g. allocation rate) are easier to change and also have the potential to influence travelers' behavior to recover efficiency losses. In this section, two scenarios of a sub-optimal toll profile are investigated, including forecast error and non-recurrent events.

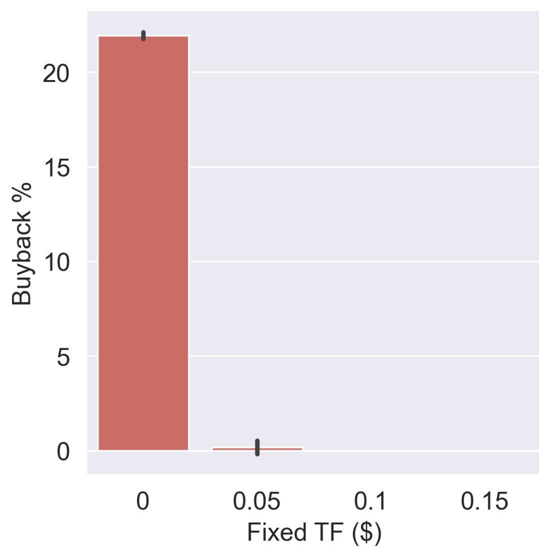
The first scenario is forecast error wherein actual road capacity is assumed to be 15% less than the anticipated road capacity used to optimize the toll profile. The social welfare of pricing and TMC with this sub-optimal toll profile (based on anticipated road capacity) are plotted in Figure 6-11 and denoted by $P{-}_S$ and MU_S. The social welfare of pricing and TMC with optimal toll profiles are also plotted and denoted as $P{-}_O$ and MU_O. As we can see, MU_S (TMC with sub-optimal tolls) is able to recover efficiency loss by reducing the allocation rate, which reduces token supply and increases token price. The optimal allocation rate is determined using a



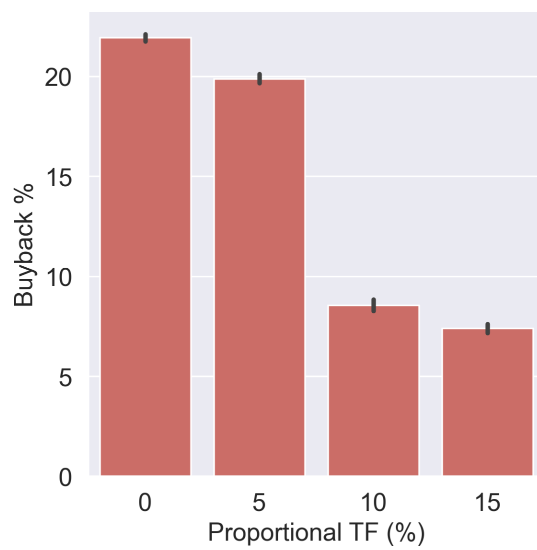
(a) Effect of Fixed TF on SW



(b) Effect of Proportional TF on SW



(c) Effect of Fixed TF on Buyback



(d) Effect of Proportional TF on Buyback

Figure 6-10: The effect of fixed and proportional transaction fees on social welfare and buyback behavior

grid search, which is found to be 15% lower than the original allocation rate.

The second scenario is a non-recurrent event. Specifically, it is assumed that there is a sudden within-day capacity drop by 15% (e.g. due to an accident or incident) from 7AM to 8:30AM on the 10th day after the system has reached an equilibrium. The social welfare across days for the three instruments are plotted in Figure 6-12. The first instrument is pricing without distribution and denoted as $P-$. The second

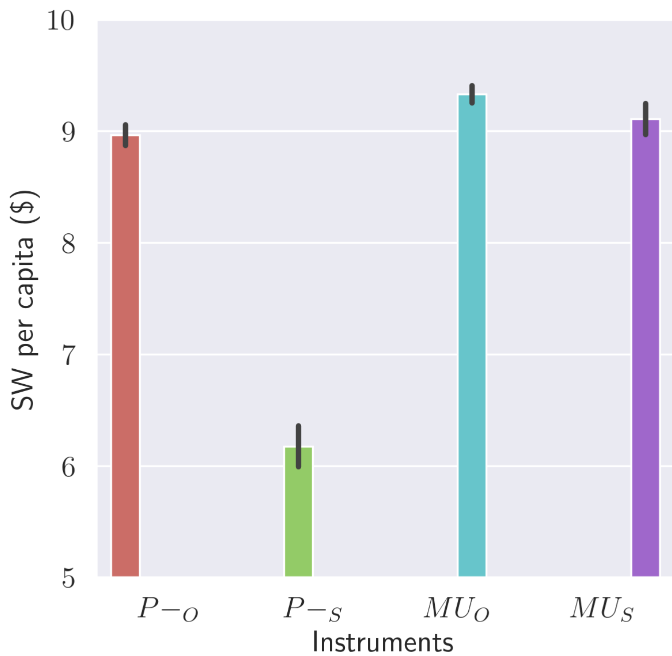


Figure 6-11: The social welfare of pricing and TMC with sub-optimal and optimal toll profiles

one is TMC with lump-sum allocation and denoted as MU_L. The third instrument is TMC with continuous allocation and denoted as MU_C.

Under MU_L, travelers receive the entire day's token allocation at the beginning of the day in the form of a 'lump-sum' allocation. This form of token allocation is the standard design of TMC schemes in the literature (e.g. [Yang and Wang, 2011a, Brands et al., 2020]). Regarding trading, they can buy additional tokens at the time of traveling for immediate use and redeem unused tokens at the end of the day. Since trading is automated, there is no transaction fee considered under the lump-sum allocation. The regulator has three market parameters to control, including within-day token price, regulation starting time and ending time. She cannot control allocation rate as all the tokens have already been allocated at the beginning of the day. Users who have not traveled yet before the new token price takes effect can update their plans according to the new information. Using the DE algorithm, we determine that it is optimal for the regulator to increase the token price to $1.8 between 6:55AM and 9:15AM. As shown in Figure 6-12, it performs better than $P-$.

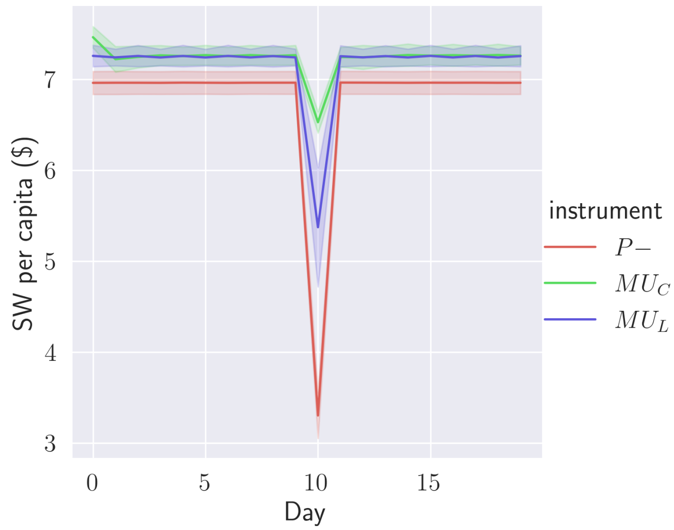


Figure 6-12: Social welfare of pricing and TMC with sub-optimal and optimal toll profiles

For MU_C, the regulator not only can control token price, regulation starting and ending time, she can also control allocation rate and transaction fees. We ignore proportional transaction fees and optimize fixed transaction fees of buying and selling together. Through optimization, between 7:05AM and 8:50AM, the regulator should set token price equal to \$1.25, allocation rate r equal to 0 and fixed transaction fee equal to \$0.5. It performs better than the lump-sum allocation MU_L as shown in Figure 6-12. This is intuitive because the cost term in the behavioral model under continuous allocation depends on allocation rate and transaction fees, which provides the regulator more degrees of freedom to intervene. This demonstrates the advantages of a continuous allocation of tokens over a lump-sum allocation of tokens.

6.3 Personalization for Pricing and TMC

6.3.1 Experimental Design

Three important factors: capacity, income effect and heterogeneity are varied one at a time across three levels as presented in Table 6.4. Values used in the base case are highlighted in red. With regard to the capacity factor, bottleneck capacity s is varied

Table 6.4: Factor levels for experiments

Factor	Level 1	Level 2	Level 3
Capacity (s)	-15%	0%	15%
Income Effect (λ)	0	3	6
Heterogeneity (c.o.v)	0.2	0.9	1.6

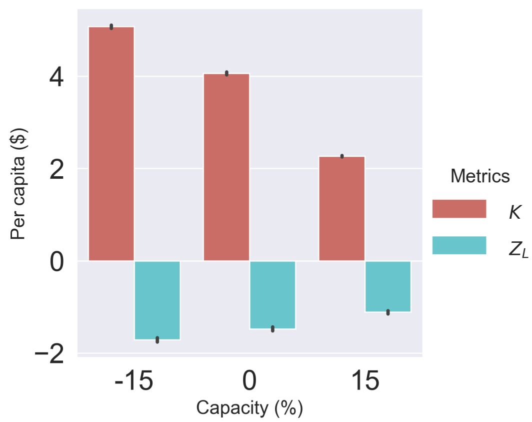
from 15% less capacity than the baseline to 15% more capacity than the baseline; for the income effect, the nonlinear income effect coefficient in the utility specification λ is varied from 0 to 6; for heterogeneity, the coefficient of variation of value of time α_n is varied from 0.2 to 1.6.

For each scenario in the experimental design, six instruments and NT are simulated with five different random seeds until convergence. The selected instruments are pricing without distribution ($P-$), pricing with uniform distribution (PU), pricing with personalized social welfare maximization distribution rule (PI_S), pricing with personalized hybrid distribution rule (PI_H), TMC with uniform distribution (MU), and TMC with personalized distribution (MI) (with hybrid distribution rule as default). Pricing with the personalized Pareto improving distribution rule is not presented since it is dominated by the hybrid distribution rule. Also, it is assumed all regulator revenue are available for distribution (i.e. $\delta = 1$).

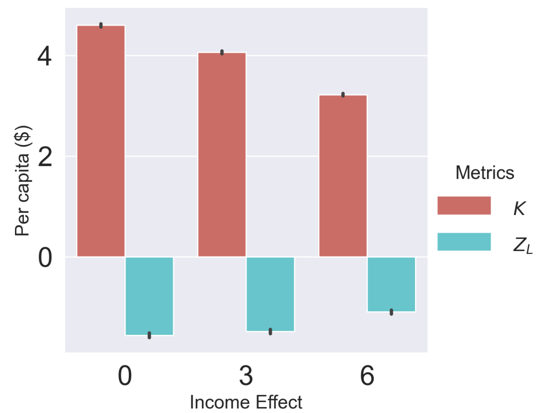
6.3.2 Results

As discussed in Chapter 5, the premise of achieving Pareto improvement from distribution is that available revenue for distribution can cover total user losses. Although it has not been proved theoretically for heterogeneous travelers, this holds for the pricing without distribution instrument $P-$ across all of our simulation experiments as shown in Figure 6-13, where Z_L represents the aggregate user losses and K represents the regulator revenue.

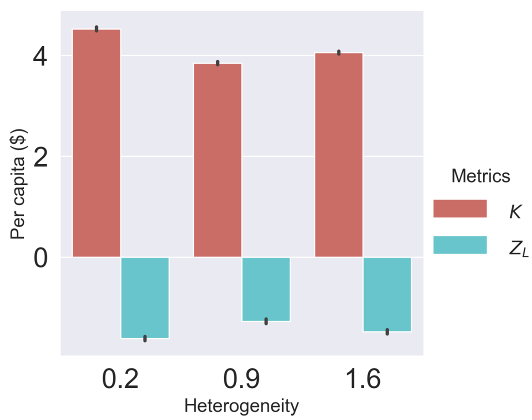
Also, we can observe the linear trend that as capacity level increases, regulator revenue decreases due to less congestion, and as income effect increases, regulator revenue decreases due to increasing cost sensitivity. However, the trend is not linear and not monotonic across the selected three levels of heterogeneity.



(a) Varying capacity



(b) Varying income effect

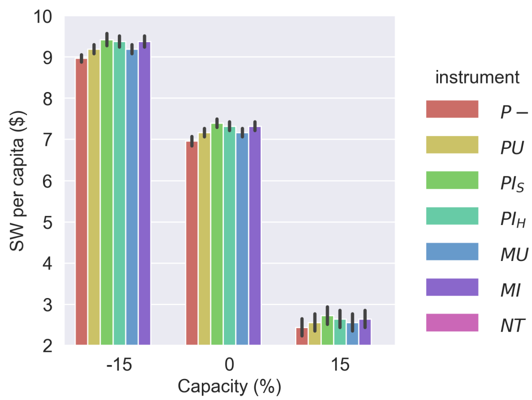


(c) Varying heterogeneity

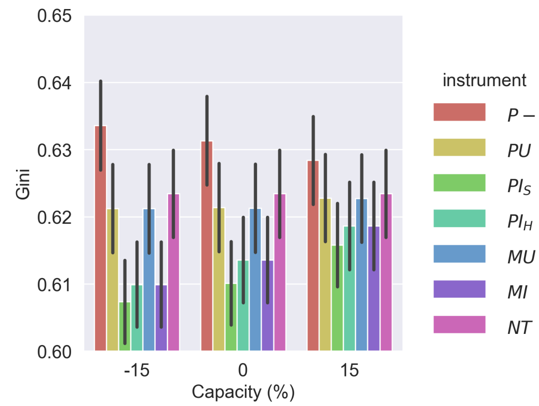
Figure 6-13: Regulator revenue versus user loss for the pricing without distribution ($P-$) by three factors

The comparative performance of the various instruments under varying levels of capacity in terms of social welfare, Gini coefficient, PT share and travel time index (TTI) are shown in Figure 6-14. A similar plot of the ratios of these four metrics to the baseline values under varying levels of capacity can be found in Figure A-1 in Appendix A. The social welfare is computed relative to the NT and consists of the user benefit and regulator revenue. As it is revenue neutral for all instruments with distribution ($\delta = 1$), their social welfare are also their user benefits. For the pricing without distribution $P-$, since regulator revenue is not neutral under $P-$, it is expected that the user benefit of $P-$ is less than other instruments. This is confirmed by the following plots on distributional impacts.

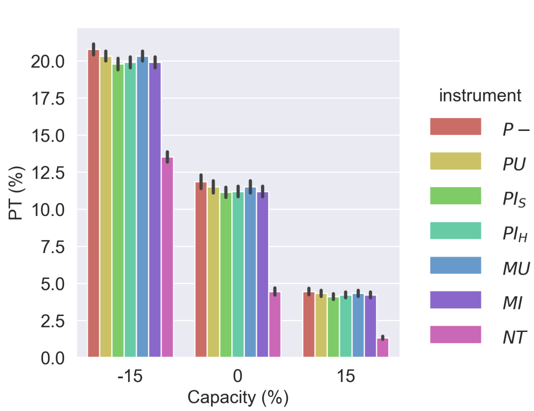
As capacity level decreases, congestion increases and the social welfare of all in-



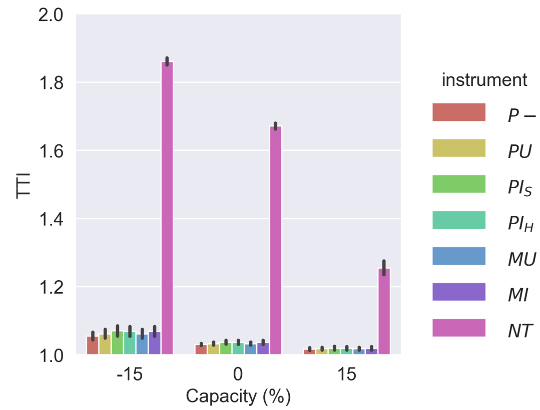
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments



(c) PT share of various instruments



(d) TTI of various instruments

Figure 6-14: Social welfare, Gini coefficient, PT share and Travel time index (TTI) of various instruments by capacity levels

struments increase relative to NT. Among all instruments, PI_S achieves the highest social welfare as its distribution rule is to maximize social welfare directly. The pricing with hybrid distribution rule PI_H has social welfare less than that of PI_S as its revenue is distributed to compensate all users' losses (not only low-income users) to ensure Pareto improvement. The pricing with uniform distribution PU has social welfare less than that of PI_H as its revenue is distributed uniformly to all users including those who do not have losses.

We can also observe than TMC with uniform token allocation MU performs the same as PU and TMC with personalized token allocation MI performs the same as PI given the effect of transaction fess are minimal. This is because the market value of token allocation is roughly equal to the dollar value of the corresponding refund, which causes similar behavior changes as the income effects are similar.

It is worth noting that in the absence of the income effects, all instruments are expected to have almost identical social welfare given that the losses due to the TMC transaction fees are minimal. This is because the distribution does not change the user behavior as utility differences are the same. The distribution also does not bring additional benefit to low-income users because their marginal utilities of income are 1. This is demonstrated in the following experiments on varying levels of the income effects. Considering the nonlinear income effects, the results indicate that the behavior changes and the marginal utility benefit of income due to the refunding and the credit allocation lead to small gains in welfare. This is promising given the other benefits, especially the distributional benefits, shown in next.

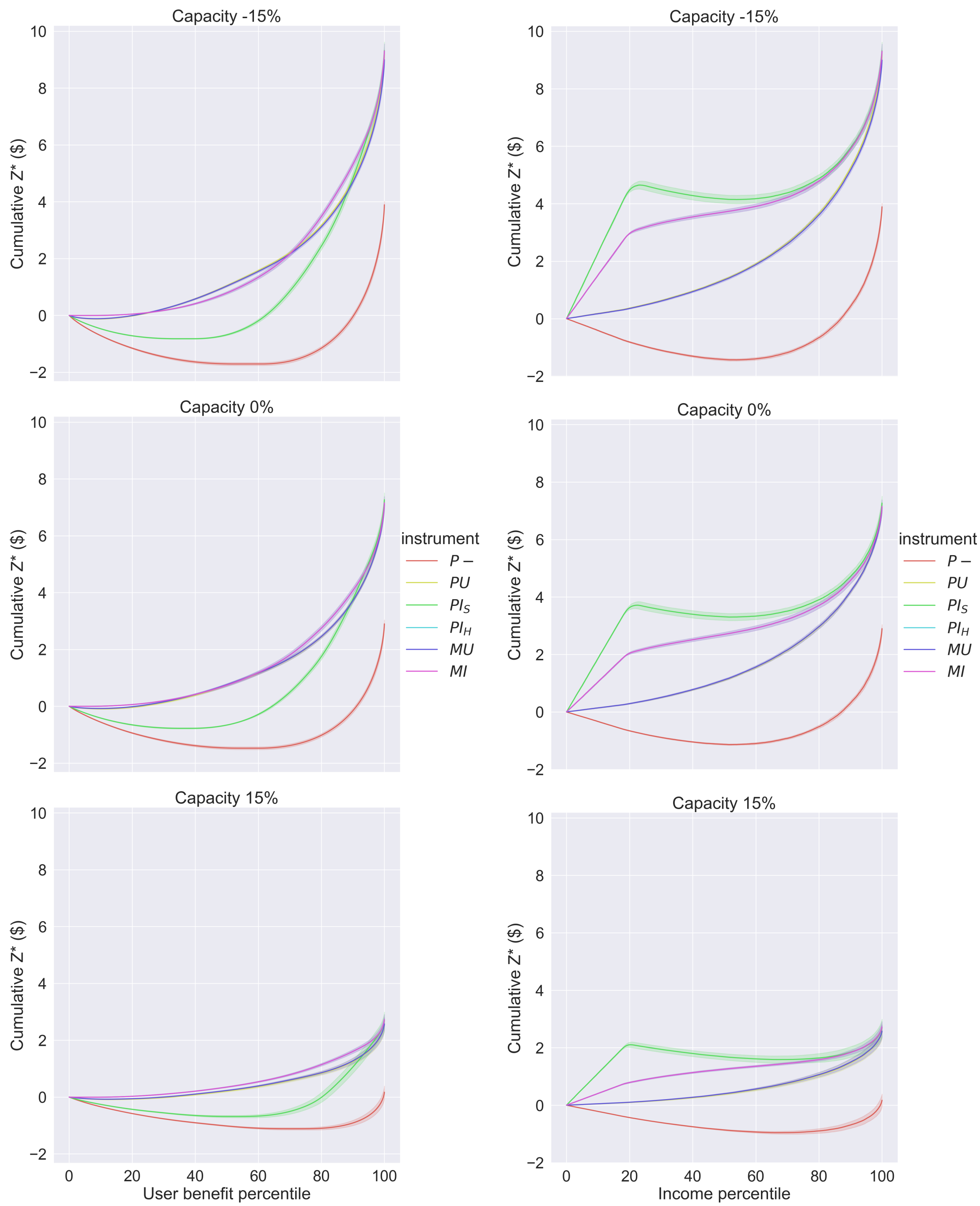
The Gini coefficient is calculated using the individual disposable income I_n plus her benefit z_n. The Gini coefficient of $P-$ increases as capacity level decreases, which implies that $P-$ becomes less equitable. This is because as capacity level decreases, the toll has to increase to deal with the increasing congestion leading to the greater losses of low income users. Among all instruments, PI_S is the most equitable because the social welfare maximization distribution rule also directly benefits low income users. It becomes more equitable as capacity level decreases because more revenue can be distributed to benefit low income users. PI_H is less equitable than PI_S because its revenue is distributed to compensate all users' losses. PU is less equitable than PI_H because its revenue is distributed uniformly to all users. The TMC instruments have the same equity as the corresponding pricing instruments with refunding because of similar reasons mentioned in social welfare discussion. A uniform refund or token allocation always improves the Gini coefficient because it increases the proportion of user benefits obtained by the lower income segments.

Regarding the PT share in NT, it increases as capacity level decreases because as capacity decreases, congestion on the road network increases leading to more users switching to PT. $P-$ has much more PT users than that of NT because of tolling. The difference between PT share in $P-$ and NT also increases as capacity decreases because as capacity decreases, congestion on the road network increases leading to higher toll, which causes more users switching to PT. The rest of instruments have similar PT shares because their toll profiles are constrained to be similar to toll profile of $P-$ in toll optimization to avoid significant behavior changes. Without this constraint, the optimal toll of instruments with distribution could increase relative to $P-$, causing the greater behavior changes. In this case, it is more users to switch to PT as we only consider two modes and such changes might be desirable if we want to promote PT. However, broadly, this could lead to some undesirable impacts on the economy also. For example, the trip cancellation might increase due to revenue refunding and higher toll charges, which prevents economic growth if the extent is too large. The issue of behavior change is subtle and needs further investigations taking more aspects of travel behavior into considerations.

In Figure 6-14d, the travel time index (TTI) of NT increases as capacity level decreases because congestion on the road network increases. Clearly, all instruments including both refunding and TMC attain the desired improvements in network performance as shown in Figure 6-14d that their TTI are close to 1.

The distributional impacts of instruments are shown in Figure 6-15. The y-axis is cumulative user benefits normalized by population. The x-axis of Figure 6-15a (left column) is user benefit percentile while the x-axis 6-15b (right column) is income percentile. If an instrument is Pareto improving, then its line should not go below 0 in Figure 6-15a. In addition, if an instrument is progressive, then its line is supposed to increase fast initially in Figure 6-15b, which means low income users have the most user benefits.

We can observe that $P-$ is regressive and not Pareto improving because low income users have losses in Figure 6-15a. When capacity is high (15%), the user benefit per capita of $P-$ is close to 0 as shown in Figure 6-15, meaning that the most of social welfare gain is due to regulator revenue. Uniform distribution (PU and MU) cannot eliminate "losers" because every user receives the same amount of distribution. Although PI_S benefits low income users significantly compared to other instruments as shown in Figure 6-15b, the mid income users still have losses. PI_H and MI are not only progressive that they benefit low income more than other users as shown



(a) Lorenz curve of user benefits (b) Distribution of user benefits by income

Figure 6-15: Distributional impacts of various instruments by capacity levels

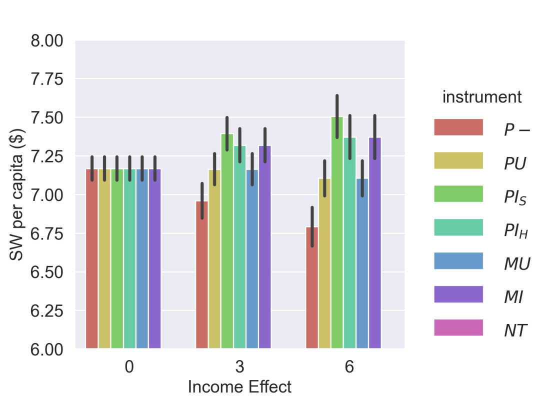
in Figure 6-15b, but also Pareto improving that no user has loss as shown in Figure 6-15a.

The results demonstrate that the developed bi-level optimization framework can significantly improve the distributional impacts to achieve progressive Pareto improvement without having significant behavior changes and deteriorating network performance. This is promising to improve the public acceptance of congestion tolling as it addresses the important equity issue.

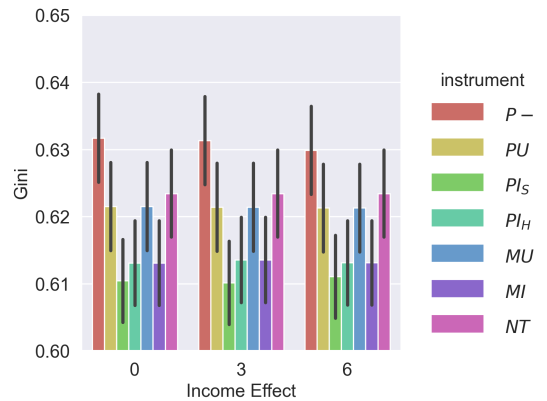
In order to operationalize this bi-level optimization framework, it is important to measure and infer user preferences and behavior accurately. We can take advantage of the existing development of the Tri-POP framework. In the example of Tripod application reviewed in Chapter 5, user behavior is modeled by a discrete choice model with inter- and intra-consumer heterogeneity using stated preferences (SP) data [Danaf et al., 2019]. The SP data collection can leverage the revealed preferences (RP) data (e.g. the departure and arrival times, origin and destination, trip mode and purpose, and activity duration) obtained using the smartphone-based sensing app [You et al., 2019] and pre-survey data on the user's characteristics (e.g. income, age, car owernship, etc.) to generate choice tasks. User's preferences can be updated online after each choice and offline periodically. The developed online estimation methodology by [Danaf et al., 2019] is also demonstrated to be scalable.

Next, the comparative performance of the various instruments under varying levels of income effects in terms of social welfare, Gini coefficient, PT share and travel time index (TTI) are shown in Figure 6-16. A similar plot of the ratios of these four metrics to the baseline values under varying levels of income effects can be found in Figure A-2 in Appendix A. As we can see, in the absence of nonlinear income effect ($\lambda = 0$), all instruments perform the same in terms of all four metrics because there is no impact on behavior as we have explained already.

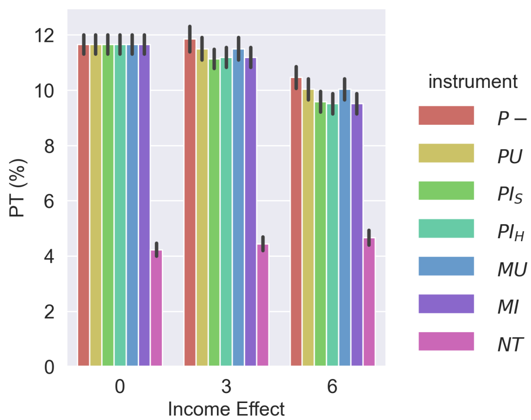
Considering the nonlinear income effects, similar to previous cases, PI_S achieves the highest social welfare and lowest Gini coefficient among the considered instruments. Its social welfare increases as the income effect increases due to the higher marginal utilities of income of low income users. In contrast, the social welfare of $P-$ decreases as the income effect increases since users are more sensitive to tolls. PI_H performs the same as MI, which perform slightly worse than PI_S in terms of efficiency and equity. PU and MU also perform the same, which are worse than



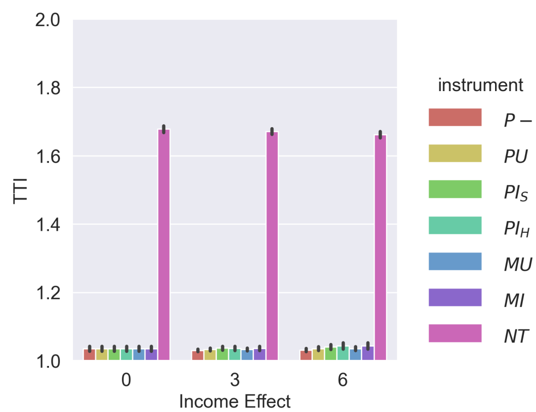
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments



(c) PT share of various instruments



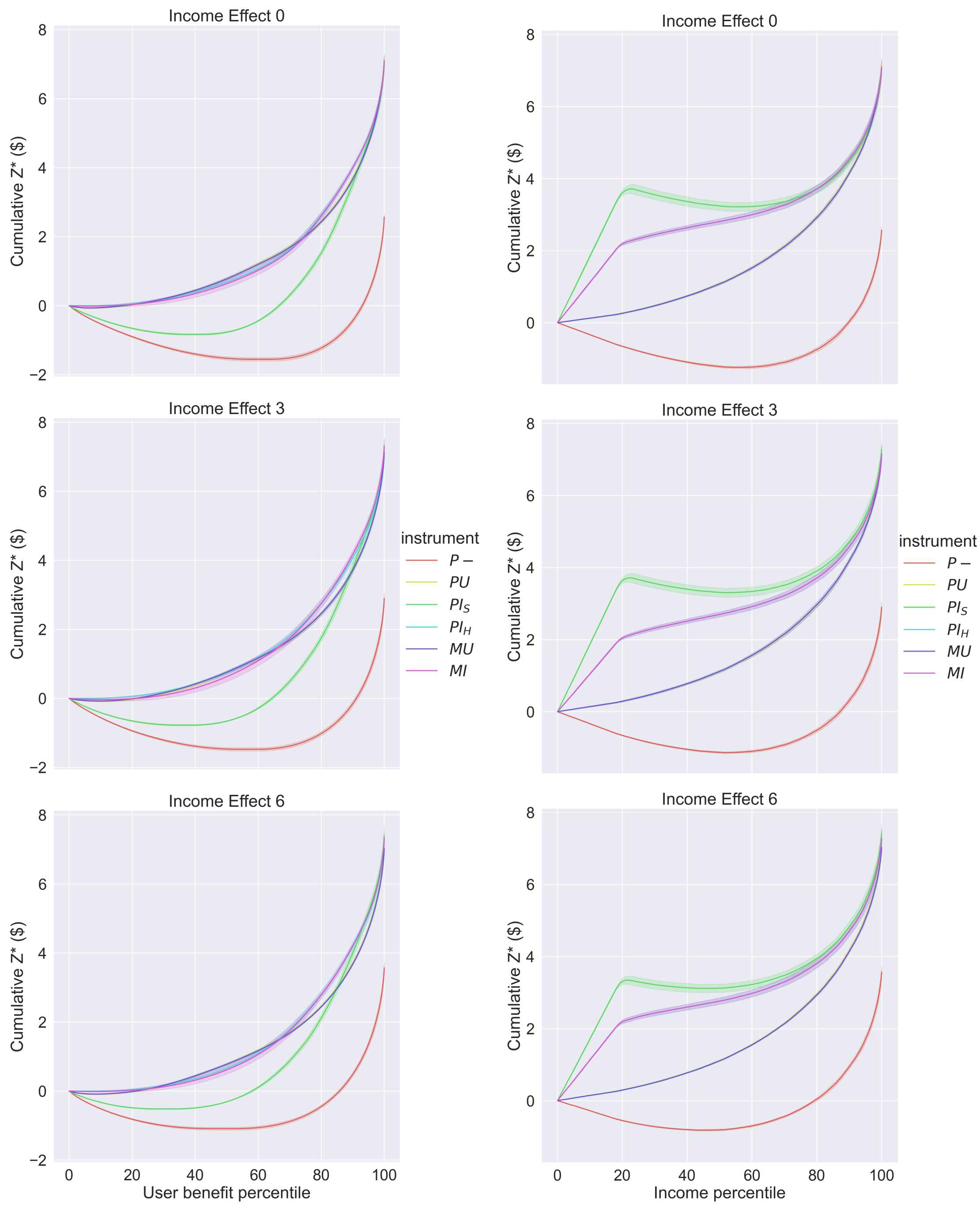
(d) TTI of various instruments

Figure 6-16: Social welfare, Gini coefficient, PT share and Travel time index (TTI) of various instruments by income effect levels

instruments with the hybrid distribution rule.

Regarding the PT share in NT, it increases as the income effect level increases because the PT fare is cheaper than the fuel cost of driving leading to more users switching to PT. Similar to varying capacity levels, all instruments have similar PT shares because of their constrained toll profiles. This implies that they do not cause significant behavior changes. In addition, all instruments have similar TTI as that of $P-$ that are close to 1 as shown in Figure 6-16d, which implies they all attain the desired improvements in network performance.

The distributional impacts of instruments across three levels of income effect are shown in Figure 6-17. Clearly, $P-$ is regressive and not Pareto improving; PI_S benefits low income users without paying attention to mid income users; PU and

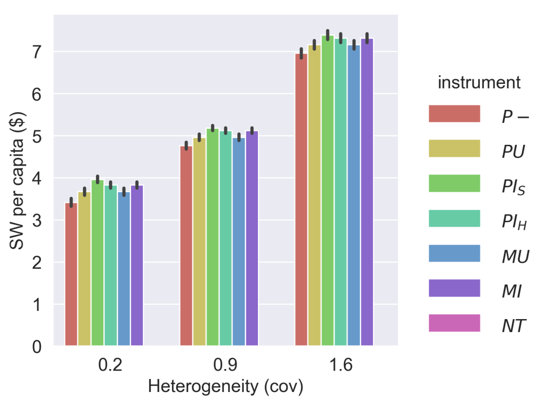


(a) Lorenz curve of user benefits (b) Distribution of user benefits by income

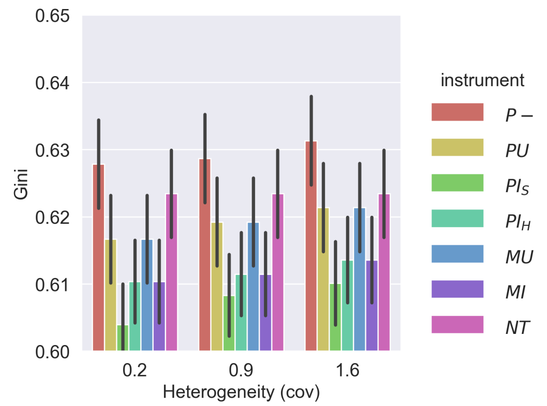
Figure 6-17: Distributional impacts of various instruments by income effect levels

MU fail to benefit all users; and PI_H and MI are both progressive and Pareto improving.

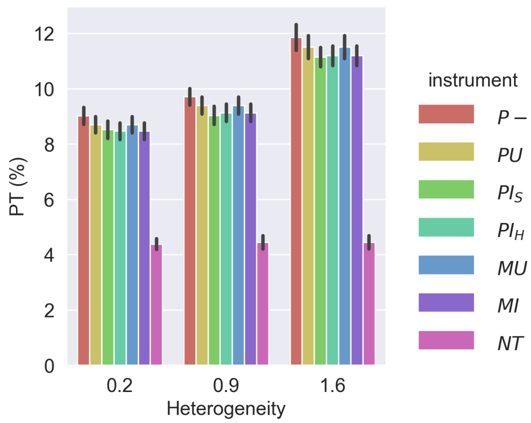
Finally, the social welfare, Gini coefficient, PT share and travel time index (TTI)



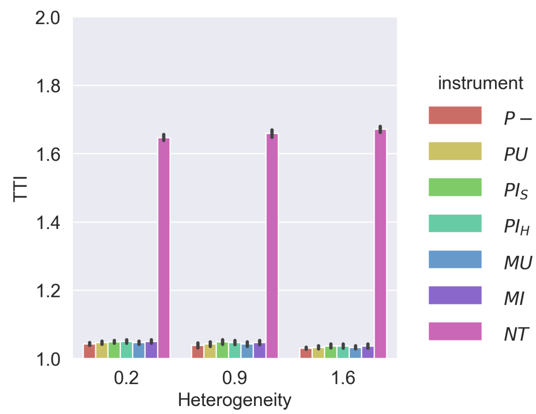
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments



(c) PT share of various instruments



(d) TTI of various instruments

Figure 6-18: Social welfare, Gini coefficient, PT share and Travel time index (TTI) of various instruments by heterogeneity levels

of various instruments across three levels of heterogeneity are shown in Figure 6-18. A similar plot of the ratios of these four metrics to the baseline values under varying levels of heterogeneity can be found in Figure A-3 in Appendix A. As the heterogeneity level increases, the social welfare of $P-$ increases, which is consistent with findings in the literature (e.g. [Van Den Berg and Verhoef, 2011]). This shows the importance of incorporating heterogeneity into the analysis to avoid underestimating the benefits of congestion tolling instruments.

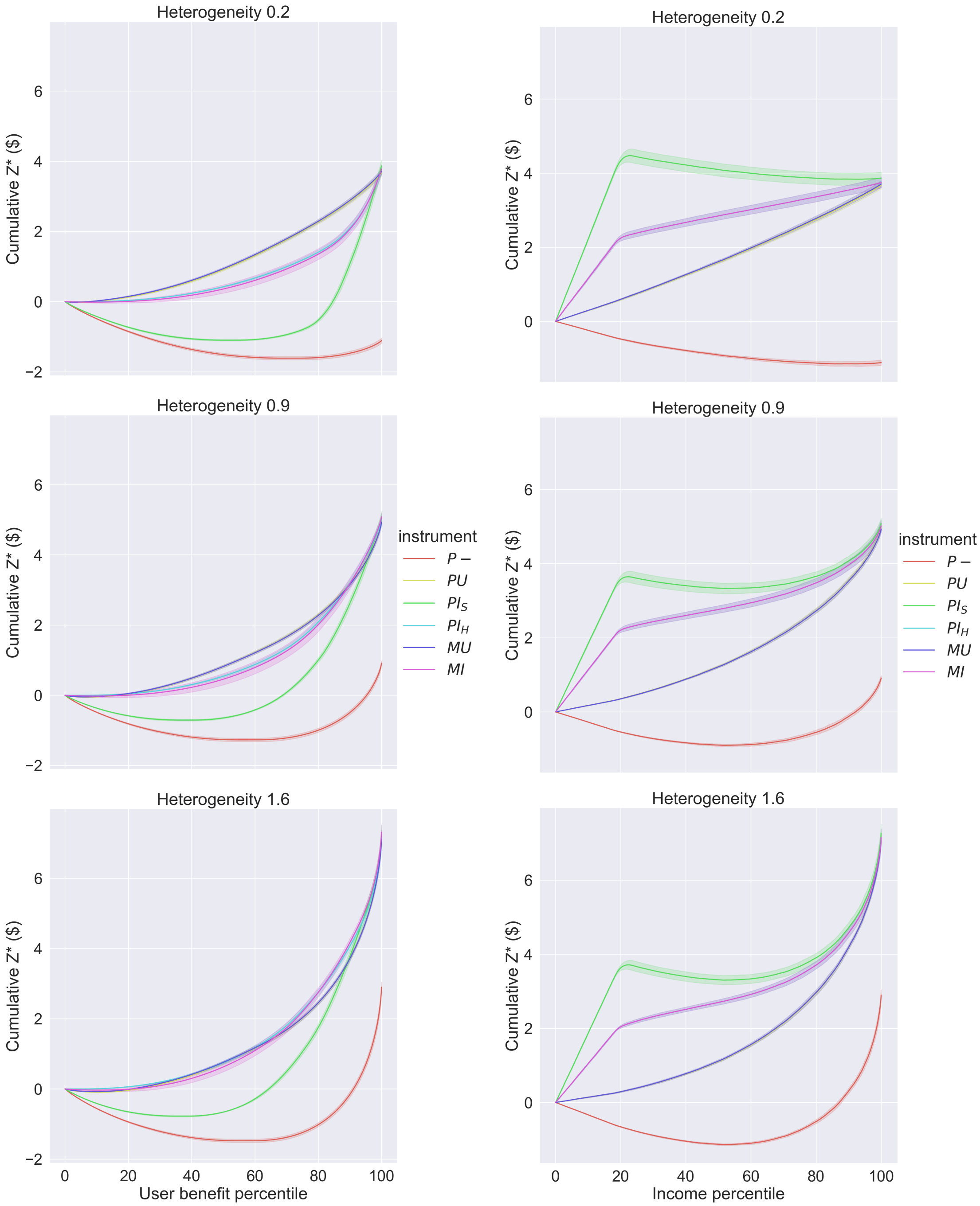
Similar to the previous two sets of experiments, PI_S achieves the highest social welfare and lowest Gini coefficient among the considered instruments. PI_H performs the same as MI, which perform slightly worse than PI_S in terms of efficiency and equity. PU and MU also perform the same, which are worse than instruments with hybrid distribution rule.

As shown in Figure 6-18c, all instruments including both revenue refunding and TMC have similar PT shares to that of $P-$, which means they do not cause significant behavior changes. As shown in Figure 6-18d, all instruments have similar TTI, which means they all attain the desired improvements in network performance.

The distributional impacts of instruments across three levels of heterogeneity are shown in Figure 6-19. Similarly, $P-$ is regressive and not Pareto improving; PI_S benefits low income users without paying attention to mid income users; PU and MU fail to benefit all users; and PI_H and MI are both progressive and Pareto improving. For the lowest heterogeneity level, the user benefit per capita of $P-$ is negative (reflected by the negative cumulative user benefit at the end of x-axis as shown in Figure 6-19). This implies that social welfare gain of $P-$ is due to regulator revenue only, which causes equity issues and reduces public acceptance.

6.4 Summary

This chapter first introduces the setup for numerical experiments. Next, the existence and uniqueness of equilibrium and the performance of the DE optimization algorithm are demonstrated numerically. Next, the market behavior of travelers is investigated to demonstrate the market design for TMC prevents undesirable behavior like excessive activities at the boundary and frequent selling. The advantages of the TMC instruments over pricing under sub-optimal toll profiles are also demonstrated. Finally, the comparative performance of the various instruments across different levels of three factors are examined. It is shown that with realistic capacity, income effects, and heterogeneity, personalization for both pricing and TMC can significantly improve distributional impacts of congestion tolling, achieving progressive Pareto improvement while maintaining efficiency without causing significant behavior changes, and deteriorating network performance.



(a) Lorenz curve of user benefits (b) Distribution of user benefits by income

Figure 6-19: Distributional impacts of various instruments by heterogeneity levels

Chapter 7

Conclusions

7.1 Contributions and Findings

This thesis has two major contributions. First, we develop a market design for TMC schemes that ensures TMC is used for mobility management and avoids undesirable behavior such as hoarding, frequent selling and speculation, excessive activity at boundary (of token expiration), and negotiation cost. The developed design considers all aspects of the market including token allocation, expiration, transaction fee, price adjustment and market rules governing trading. In addition, a heuristic approach to model disaggregate selling behavior is developed and the resulting simple selling strategy is derived. The effect of proportional and fixed transaction fees on selling behavior are discussed analytically. The developed market design addresses a growing and imminent need to develop methodologies to realistically model TMC schemes that are suited for real-world deployments and can help us better understand the performance of these systems – and the impact in particular, of market dynamics.

Second, we develop a bi-level optimization framework for personalized distribution to make congestion tolling (both price and quantity controls) efficient, equitable, and Pareto improving. It is an application of an online analytics framework for transportation system management and smart mobility termed Tri-POP, which combines prediction, optimization and personalization (POP). The system optimization determines toll policy with the objective to maximize social welfare while the user optimization can be formulated with different objectives (e.g. maximize social welfare or achieve Pareto improvement) to determine individual distribution of revenue for pricing or tokens for TMC. The system optimization is solved using a metaheuristic approach termed Differential Evolution (DE) while the user optimization is solved analytically. The developed personalized congestion tolling is promising to improve the public acceptance as it addresses the important equity issue.

The performance of the designed instruments are demonstrated via microsimulation in a daily commute context between a single origin-destination pair. Under congestion tolling, travelers are subject to a time-of-day toll profile in units of dollars (for pricing) or tokens (for TMC). The simulation experiments employ a day-to-day assignment framework wherein transportation demand is modeled using a logit-mixture model with the nonlinear income effects and supply is modeled using a standard bottleneck model. The evaluation framework includes four main categories: social welfare, distributional impacts, behavior change, and level of congestion.

Regarding market design for TMC, the results indicate that small fixed transaction fees can effectively mitigate undesirable behavior in the market without a significant loss in efficiency (social welfare) whereas proportional transaction fees are less effective both in terms of efficiency and in achieving desirable market behavior. Further, the TMC scheme is more robust in the presence of forecasting errors and non-recurrent events due to the adaptiveness of the market.

For the personalized pricing and TMC instruments, we vary three factors (capacity, income effect, and heterogeneity) across three levels, one at a time. We examine the comparative performance of the personalized congestion tolling (both pricing and TMC) relative to the traditional congestion pricing without revenue refunds and congestion tolling with uniform distribution. First, we show that the regulator revenue of the pricing without revenue refunds $P-$ is always greater than the user losses in all experiments we try, which is the premise of achieving Pareto improvement.

Second, we find it is important to consider the nonlinear income effect, otherwise, all instruments have almost the identical social welfare given that the losses due to the TMC transaction fees are minimal. This is because the distribution does not change the user behavior as utility differences are the same.

Finally, with the realistic capacity, heterogeneity and income effect, personalization with the hybrid distribution rule for both pricing and TMC (PI_H and MI) can make congestion tolling more efficient, equitable and achieve Pareto improvement without having significant behavior changes and deteriorating network performance. Although pricing with social welfare maximization distribution rule (PI_S) is able to achieve the highest social welfare and equity among the considered instruments, it fails to benefit all users such that the mid income users still have losses.

7.2 Limitations and Future Work

We suggest the following future research directions to improve and extend methodologies developed in this study.

1. There are a few aspects of the market design for TMC that need further investigation. First, the token price is adjusted daily in this thesis but it is worth investigating the within day price adjustment also, which may balance demand and supply faster and improve efficiency. Second, the user selling strategy is derived using a heuristic approach, which needs to be validated with empirical data. Third, it is worth considering users' loss aversion and novelty effect.

2. This thesis mainly focuses on the departure time choice and two-mode choice in the context of a daily commute. More dimensions of travelers' behavior (e.g. route choice, other modes, parking, and other travel purposes) can be incorporated. The toll optimization then needs to be extended to capture such behaviors. In addition, it would be more realistic to relax the assumption that the evening trip and morning trip are symmetric.

3. The time-of-day pricing or token charging schemes considered in this thesis are for a single link. For real-world implementations, it is meaningful to extend current work to accommodate tolling on general networks and investigate the

implications of different forms of tolling (e.g. link-based, cordon-based, area-based, and distance-based). It is also possible to consider multiple objectives in the optimization (e.g. reduce emission or energy consumption). Also, for general networks, both social and spatial equity issues exist, which may need different solutions to offset the disadvantage.

4. It is shown that the real-time optimization can further improve efficiency (e.g. [de Palma and Lindsey, 2011]). It is meaningful to investigate adaptive tolling and market parameters for TMC in real-time based on predictions (e.g. speed and flow). Travelers' preferences can also be updated both online and offline to improve optimization performance.

5. This thesis develops and demonstrates the personalized congestion tolling can be more efficient, equitable and benefit all travelers (Pareto improving), which likely increases its public acceptance. However, as reviewed in Chapter 2, the public acceptance is also affected by factors such as privacy, complexity and uncertainty. For practical applications of the personalized pricing or TMC, it is important for policy makers to address these concerns. Also, it is important to investigate the interactions of personalized congestion tolling with other sectors of the economy (e.g. labor supply, freight).

6. The premise of achieving Pareto improvement from personalization is that available revenue for distribution can cover total user losses. Although this holds for all experiments conducted in this study, it is important to mathematically derive conditions for this to hold and investigate whether this holds or not in general networks considering heterogeneity, elastic demand and nonlinear income effect. On the other hand, it is possible to include a constraint in toll optimization to ensure the optimized toll profile allows to achieve Pareto improvement, otherwise it is not sensible to implement congestion tolling.

7. The scope of thesis focuses on the short-term effects of congestion tolling assuming that the long-term decisions such as residential location choice, work

location choice, vehicle ownership, network infrastructures and services, and consumption are not affected. While in the long-term, congestion tolling could have substantial impacts to travelers and different among instruments. For example, it is found that the demand is more elastic in the long-run as travelers have more time to react and find other options (e.g. living or working at places where they can avoid toll charges).

8. In the numerical experiments, we do not have reliable data sources for some parameters. For example, because we do not find empirical data about the distribution of values of schedule delay early and late, we assume their distribution based on reasonable justification. Similarly, we assume a constant necessary living expense for all travelers to convert their market income to disposable income for transportation. The individual specific random component term is calibrated based on price elasticity but is assumed to be perfectly correlated under all instruments, which might not be true.

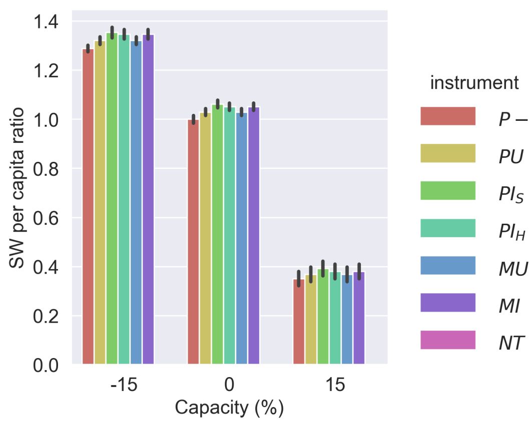
9. More extensive testing and analysis of the developed market design for TMC and personalization framework are necessary. More levels of factors and full factorial experiments can be conducted to compare various instruments. For example, it is important to investigate the influence of different distributions of disposable income. Large-scale simulation experiments on a realistic network are helpful to provide insights for real-world applications. Field experiments can provide valuable empirical data for model estimation and validation, performance evaluation, and insights on users' attitudes towards new instruments.

Appendix A

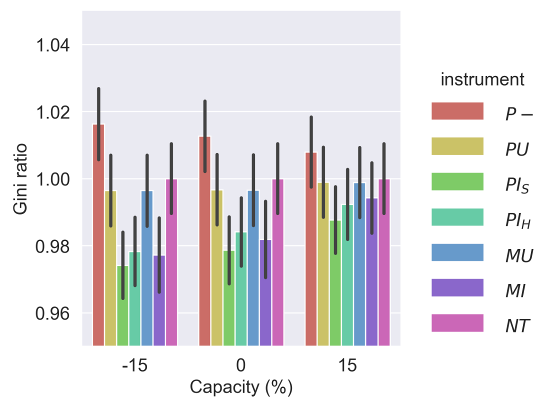
Figures

The ratios of social welfare, Gini coefficient, PT share and travel time index to the baseline values under varying levels of capacity are shown in Figure A-1. Specifically, in Figure A-1a, the ratios of social welfare of each instrument to the social welfare of $P-$ with the baseline capacity are plotted; in Figure A-1b, the ratios of Gini coefficient of each instrument to the Gini coefficient of NT with the baseline capacity are plotted; in Figure A-1c, the ratios of PT share of each instrument to the PT share of NT with the baseline capacity are plotted; in Figure A-1d, the ratios of TTI of each instrument to the TTI of NT with the baseline capacity are plotted.

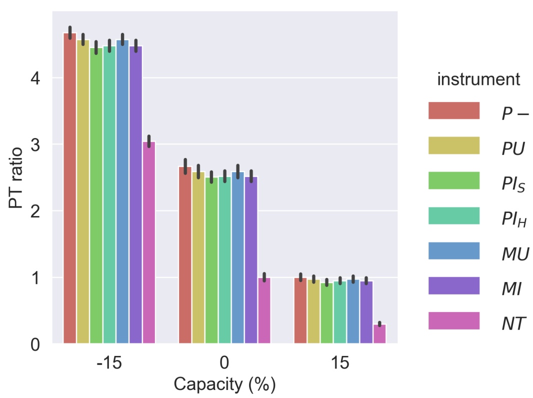
Similarly, the ratios of social welfare, Gini coefficient, PT share and travel time index to the baseline values under varying levels of income effects are shown in Figure A-2; the ratios of social welfare, Gini coefficient, PT share and travel time index to the baseline values under varying levels of heterogeneity are shown in Figure A-3. The findings and insights from these three plots are the same as those discussed in Chapter 6.



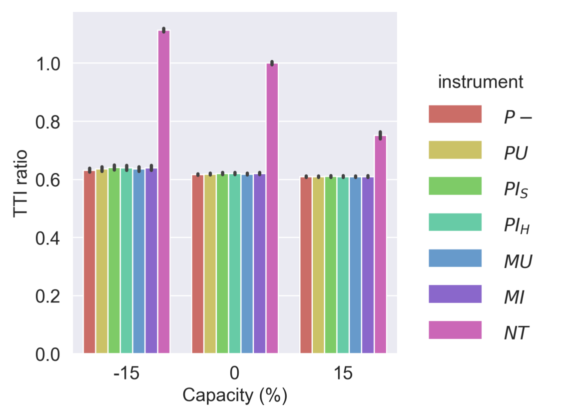
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments

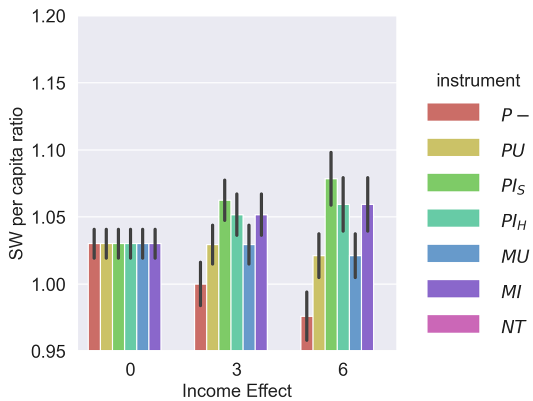


(c) PT share of various instruments

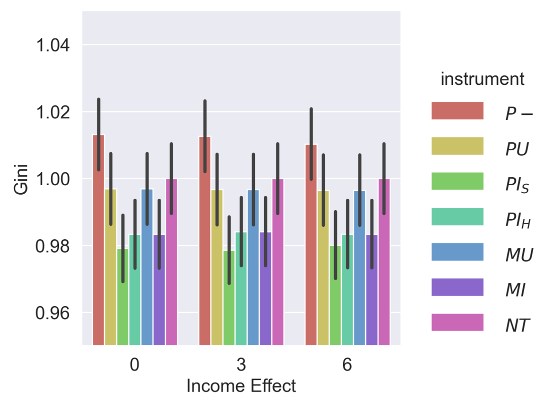


(d) TTI of various instruments

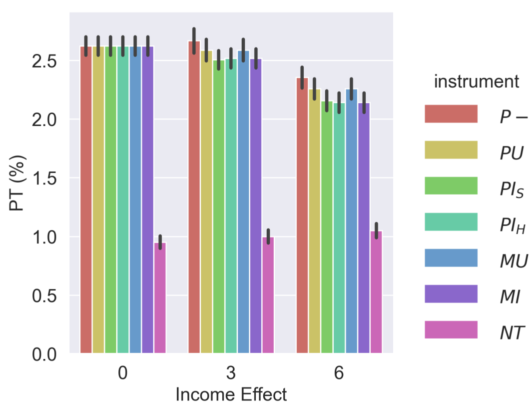
Figure A-1: The ratios of social welfare, Gini coefficient, PT share and Travel time index (TTI) to the baseline values of various instruments by capacity levels



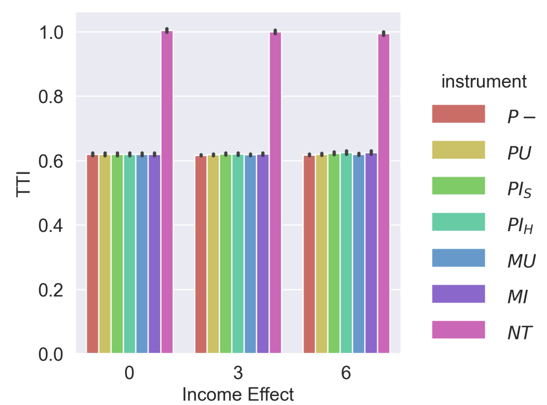
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments

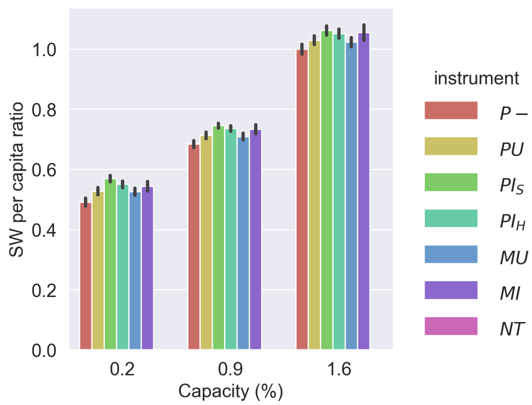


(c) PT share of various instruments

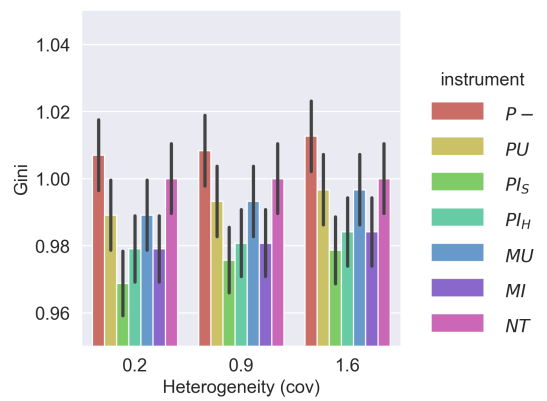


(d) TTI of various instruments

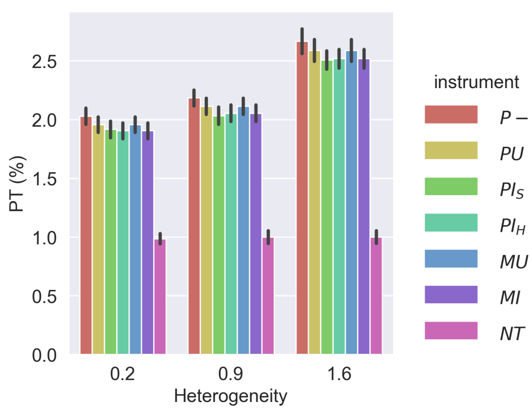
Figure A-2: The ratios of social welfare, Gini coefficient, PT share and Travel time index (TTI) to the baseline values of various instruments by income effect levels



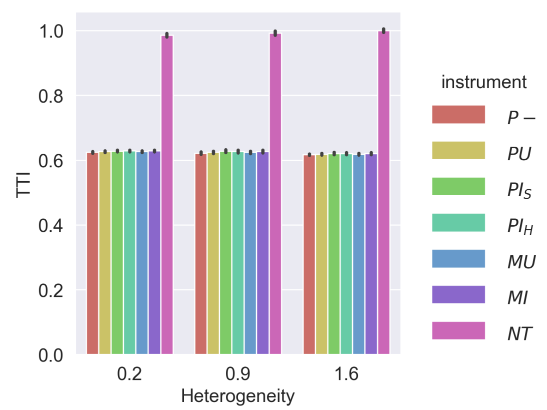
(a) Social welfare of various instruments



(b) Gini coefficients of various instruments



(c) PT share of various instruments



(d) TTI of various instruments

Figure A-3: The ratios of social welfare, Gini coefficient, PT share and Travel time index (TTI) to the baseline values of various instruments by heterogeneity levels

Bibliography

[Adler and Cetin, 2001] Adler, J. L. and Cetin, M. (2001). A direct redistribution model of congestion pricing. *Transportation research part B: methodological*, 35(5):447–460.

[Akamatsu and Wada, 2017] Akamatsu, T. and Wada, K. (2017). Tradable network permits: A new scheme for the most efficient use of network capacity. *Transportation Research Part C: Emerging Technologies*, 79:178–195.

[Ali et al., 2007] Ali, A. T., Flannery, A., and Venigalla, M. M. (2007). Prediction models for free flow speed on urban streets. Technical report.

[Allen et al., 2006] Allen, S., Gaunt, M., and Rye, T. (2006). An investigation into the reasons for the rejection of congestion charging by the citizens of edinburgh.

[Altshuler, 2010] Altshuler, A. (2010). Equity, pricing, and surface transportation politics. *Urban Affairs Review*, 46(2):155–179.

[Anas, 2020] Anas, A. (2020). The cost of congestion and the benefits of congestion pricing: A general equilibrium analysis. *Transportation Research Part B: Methodological*, 136:110–137.

[Araldo et al., 2019] Araldo, A., Gao, S., Seshadri, R., Azevedo, C. L., Ghafourian, H., Sui, Y., Ayaz, S., Sukhin, D., and Ben-Akiva, M. (2019). System-level optimization of multi-modal transportation networks for energy efficiency using personalized incentives: Formulation, implementation, and performance. *Transportation Research Record*, page 0361198119864906.

[Arnott et al., 1990a] Arnott, R., de Palma, A., and Lindsey, R. (1990a). Departure time and route choice for the morning commute. *Transportation Research Part B: Methodological*, 24(3):209–228.

[Arnott et al., 1990b] Arnott, R., De Palma, A., and Lindsey, R. (1990b). Economics of a bottleneck. *Journal of urban economics*, 27(1):111–130.

[Arnott et al., 1994] Arnott, R., De Palma, A., and Lindsey, R. (1994). The welfare effects of congestion tolls with heterogeneous commuters. *Journal of Transport Economics and Policy*, pages 139–161.

[Atasoy et al., 2020] Atasoy, B., de Azevedo, C. L., Akkinepally, A. P., Seshadri, R., Zhao, F., Abou-Zeid, M., and Ben-Akiva, M. (2020). Smart mobility via prediction, optimization and personalization. In *Demand for Emerging Transportation Systems*, pages 227–265. Elsevier.

[Azevedo et al., 2018] Azevedo, C. L., Seshadri, R., Gao, S., Atasoy, B., Akkinepally, A. P., Christofa, E., Zhao, F., Trancik, J., and Ben-Akiva, M. (2018). Tripod: sustainable travel incentives with prediction, optimization, and personalization. In *Proceedings of the Transportation Research Record 97th Annual Meeting.*

[Aziz et al., 2015] Aziz, H. A., Ukkusuri, S. V., and Romero, J. (2015). Understanding short-term travel behavior under personal mobility credit allowance scheme using experimental economics. *Transportation Research Part D: Transport and Environment*, 36:121–137.

[Bao et al., 2014] Bao, Y., Gao, Z., Xu, M., and Yang, H. (2014). Tradable credit scheme for mobility management considering travelers' loss aversion. *Transportation Research Part E: Logistics and Transportation Review*, 68:138–154.

[Bao et al., 2019] Bao, Y., Verhoef, E. T., and Koster, P. (2019). Regulating dynamic congestion externalities with tradable credit schemes: does a unique equilibrium exist? *Transportation Research Part B: Methodological*, 127:225–236.

[Ben-Akiva et al., 1984] Ben-Akiva, M., Cyna, M., and De Palma, A. (1984). Dynamic model of peak period congestion. *Transportation Research Part B: Methodological*, 18(4-5):339–355.

[Ben-Akiva et al., 1985] Ben-Akiva, M. E., Lerman, S. R., Lerman, S. R., et al. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

[Blackorby and Donaldson, 1990] Blackorby, C. and Donaldson, D. (1990). A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *canadian Journal of Economics*, pages 471–494.

[Braid, 1996] Braid, R. M. (1996). Peak-load pricing of a transportation route with an unpriced substitute. *Journal of Urban Economics*, 40(2):179–197.

[Brands et al., 2021] Brands, D., Verhoef, E., and Knockaert, J. (2021). Pcoins for parking: a field experiment with tradable mobility permits.

[Brands et al., 2020] Brands, D. K., Verhoef, E. T., Knockaert, J., and Koster, P. R. (2020). Tradable permits to manage urban mobility: market design and experimental implementation. *Transportation Research Part A: Policy and Practice*, 137:34–46.

[Buitelaar et al., 2007] Buitelaar, E., Van der Heijden, R., and Argiolu, R. (2007). Managing traffic by privatization of road capacity: A property rights approach. *Transport Reviews*, 27(6):699–713.

[Bull et al., 2003] Bull, A., CEPAL, N., et al. (2003). *Traffic Congestion: The Problem and how to Deal with it.* ECLAC.

[Cantarella and Cascetta, 1995] Cantarella, G. E. and Cascetta, E. (1995). Dynamic processes and equilibrium in transportation networks: towards a unifying theory. *Transportation Science*, 29(4):305–329.

[Cascetta and Cantarella, 1991] Cascetta, E. and Cantarella, G. E. (1991). A day-to-day and within-day dynamic stochastic assignment model. *Transportation Research Part A: General*, 25(5):277–291.

[Chen, 2010] Chen, M. (2010). Travel time based congestion measures for freeway corridors.

[Chu, 1999] Chu, X. (1999). Alternative congestion pricing schedules. *Regional Science and Urban Economics*, 29(6):697–722.

[Constantin and Florian, 1995] Constantin, I. and Florian, M. (1995). Optimizing frequencies in a transit network: a nonlinear bi-level programming approach. *International Transactions in Operational Research*, 2(2):149–164.

[Daganzo, 1985] Daganzo, C. F. (1985). The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, 19(1):29–37.

[Danaf et al., 2019] Danaf, M., Becker, F., Song, X., Atasoy, B., and Ben-Akiva, M. (2019). Online discrete choice models: Applications in personalized recommendations. *Decision Support Systems*, 119:35–45.

[De Borger and Proost, 2012] De Borger, B. and Proost, S. (2012). A political economy model of road pricing. *Journal of Urban Economics*, 71(1):79–92.

[De Borger et al., 2008] De Borger, B., Proost, S., and Van Dender, K. (2008). Private port pricing and public investment in port and hinterland capacity. *Journal of Transport Economics and Policy (JTEP)*, 42(3):527–561.

[de Cea Ch et al., 2009] de Cea Ch, J., Valverde, G. G., et al. (2009). Effect of advanced traveler information systems and road pricing in a network with non-recurrent congestion. *Transportation Research Part A: Policy and Practice*, 43(5):481–499.

[De Palma et al., 1983] De Palma, A., Ben-Akiva, M., Lefevre, C., and Litinas, N. (1983). Stochastic equilibrium model of peak period traffic congestion. *Transportation Science*, 17(4):430–453.

[de Palma and Kilani, 2005] de Palma, A. and Kilani, K. (2005). Switching in the logit. *Economics Letters*, 88(2):196–202.

[De Palma et al., 2005] De Palma, A., Kilani, M., and Lindsey, R. (2005). Congestion pricing on a road network: A study using the dynamic equilibrium simulator metropolis. *Transportation Research Part A: Policy and Practice*, 39(7-9):588–611.

[De Palma and Lefevre, 1983] De Palma, A. and Lefevre, C. (1983). Individual decision-making in dynamic collective systems. *Journal of Mathematical Sociology*, 9(2):103–124.

[De Palma and Lindsey, 1998] De Palma, A. and Lindsey, R. (1998). Information and usage of congestible facilities under different pricing regimes. *Canadian Journal of Economics*, pages 666–692.

[De Palma and Lindsey, 2000] De Palma, A. and Lindsey, R. (2000). Private toll roads: Competition under various ownership regimes. *The Annals of Regional Science*, 34(1):13–35.

[De Palma and Lindsey, 2002a] De Palma, A. and Lindsey, R. (2002a). Comparison of morning and evening commutes in the vickrey bottleneck model. *Transportation Research Record*, 1807(1):26–33.

[De Palma and Lindsey, 2002b] De Palma, A. and Lindsey, R. (2002b). Congestion pricing in the morning and evening peaks: A comparison using the bottleneck model. In *Transportation Visioning-2002 and Beyond (Vision d'avenir des transports-2002 et au-dela), Canadian Transportation Research Forum, Proceedings of the 37th Annual ConferenceCANADIAN TRANSPORTATION RESEARCH FORUM.*

[De Palma and Lindsey, 2004] De Palma, A. and Lindsey, R. (2004). Congestion pricing with heterogeneous travelers: A general-equilibrium welfare analysis. *Networks and spatial economics*, 4(2):135–160.

[de Palma and Lindsey, 2011] de Palma, A. and Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6):1377–1399.

[de Palma et al., 2018] de Palma, A., Proost, S., Seshadri, R., and Ben-Akiva, M. (2018). Congestion tolling-dollars versus tokens: A comparative analysis. *Transportation Research Part B: Methodological*, 108:261–280.

[Delle Site and Salucci, 2013] Delle Site, P. and Salucci, M. V. (2013). Transition choice probabilities and welfare analysis in random utility models with imperfect before–after correlation. *Transportation Research Part B: Methodological*, 58:215–242.

[Deng, 1998] Deng, X. (1998). Complexity issues in bilevel linear programming. In *Multilevel optimization: Algorithms and applications*, pages 149–164. Springer.

[Dial, 1999] Dial, R. B. (1999). Minimal-revenue congestion pricing part i: A fast algorithm for the single-origin case. *Transportation Research Part B: Methodological*, 33(3):189–202.

[Ding et al., 2015] Ding, C., Mishra, S., Lin, Y., and Xie, B. (2015). Cross-nested joint model of travel mode and departure time choice for urban commuting trips: Case study in maryland–washington, dc region. *Journal of Urban Planning and Development*, 141(4):04014036.

[Dogterom et al., 2017] Dogterom, N., Ettema, D., and Dijst, M. (2017). Tradable credits for managing car travel: a review of empirical research and relevant behavioural approaches. *Transport Reviews*, 37(3):322–343.

[Duranton and Turner, 2011] Duranton, G. and Turner, M. A. (2011). The fundamental law of road congestion: Evidence from us cities. *American Economic Review*, 101(6):2616–52.

[Eliasson, 2001] Eliasson, J. (2001). Road pricing with limited information and heterogeneous users: A successful case. *The annals of regional science*, 35(4):595–604.

[Fan and Jiang, 2013] Fan, W. and Jiang, X. (2013). Tradable mobility permits in roadway capacity allocation: Review and appraisal. *Transport Policy*, 30:132–142.

[Farrell and Saleh, 2005] Farrell, S. and Saleh, W. (2005). Road-user charging and the modelling of revenue allocation. *Transport Policy*, 12(5):431–442.

[Ferrari, 1999] Ferrari, P. (1999). A model of urban transport management. *Transportation Research Part B: Methodological*, 33(1):43–61.

[Fiorello et al., 2010] Fiorello, D., Fermi, F., Maffi, S., and Martino, A. (2010). Mobility rights for urban road pricing: a modelling analysis with a system dynamics approach. In *12th World Conference on Transport Research, July*, pages 11–15.

[Francke and Kaniok, 2013] Francke, A. and Kaniok, D. (2013). Responses to differentiated road pricing schemes. *Transportation Research Part A: Policy and Practice*, 48:25–30.

[Gastwirth, 1972] Gastwirth, J. L. (1972). The estimation of the lorenz curve and gini index. *The review of economics and statistics*, pages 306–316.

[Gaunt et al., 2007] Gaunt, M., Rye, T., and Allen, S. (2007). Public acceptability of road user charging: the case of edinburgh and the 2005 referendum. *Transport Reviews*, 27(1):85–102.

[Gillingham, 2014] Gillingham, K. (2014). Identifying the elasticity of driving: evidence from a gasoline price shock in california. *Regional Science and Urban Economics*, 47:13–24.

[Global, 2018] Global, I. (2018). Traffic scorecard ranking.

[Goddard, 1997] Goddard, H. C. (1997). Using tradeable permits to achieve sustainability in the world's large cities: policy design issues and efficiency conditions for controlling vehicle emissions, congestion and urban decentralization with an application to mexico city. *Environmental and Resource Economics*, 10(1):63–99.

[Goodwin and Noland, 2003] Goodwin, P. and Noland, R. B. (2003). Building new roads really does create extra traffic: a response to prakash et al. *Applied Economics*, 35(13):1451–1457.

[Goodwin, 1996] Goodwin, P. B. (1996). Empirical evidence on induced traffic. *Transportation*, 23(1):35–54.

[Grant-Muller and Xu, 2014] Grant-Muller, S. and Xu, M. (2014). The role of tradable credit schemes in road traffic congestion management. *Transport Reviews*, 34(2):128–149.

[Gu et al., 2018] Gu, Z., Liu, Z., Cheng, Q., and Saberi, M. (2018). Congestion pricing practices and public acceptance: A review of evidence. *Case Studies on Transport Policy*, 6(1):94–101.

[Gulipalli and Kockelman, 2008] Gulipalli, P. K. and Kockelman, K. M. (2008). Credit-based congestion pricing: A dallas-fort worth application. *Transport Policy*, 15(1):23–32.

[Guo et al., 2012] Guo, X. et al. (2012). Pareto-improving congestion pricing and revenue refunding with elastic demand.

[Guo and Yang, 2010] Guo, X. and Yang, H. (2010). Pareto-improving congestion pricing and revenue refunding with multiple user classes. *Transportation Research Part B: Methodological*, 44(8-9):972–982.

[Hansen et al., 1992] Hansen, P., Jaumard, B., and Savard, G. (1992). New branch-and-bound rules for linear bilevel programming. *SIAM Journal on scientific and Statistical Computing*, 13(5):1194–1217.

[Harsanyi, 1955] Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, 63(4):309–321.

[Hau, 1990] Hau, T. D. (1990). Electronic road pricing: developments in hong kong 1983-1989. *Journal of Transport Economics and Policy*, 24(2):203–214.

[He et al., 2013] He, F., Yin, Y., Shirmohammadi, N., and Nie, Y. M. (2013). Tradable credit schemes on networks with mixed equilibrium behaviors. *Transportation Research Part B: Methodological*, 57:47–65.

[Hearn and Ramana, 1998] Hearn, D. W. and Ramana, M. V. (1998). Solving congestion toll pricing models. In *Equilibrium and advanced transportation modelling*, pages 109–124. Springer.

[Hendrickson and Kocur, 1981] Hendrickson, C. and Kocur, G. (1981). Schedule delay and departure time decisions in a deterministic model. *Transportation science*, 15(1):62–77.

[Hensher and Li, 2013] Hensher, D. A. and Li, Z. (2013). Referendum voting in road pricing reform: A review of the evidence. *Transport Policy*, 25:186–197.

[Hepburn, 2006] Hepburn, C. (2006). Regulation by prices, quantities, or both: a review of instrument choice. *Oxford review of economic policy*, 22(2):226–247.

[Holguin-Veras et al., 2005] Holguin-Veras, J., Ozbay, K., de Cerreno, A. L., et al. (2005). Evaluation study of port authority of new york and new jersey's time of day pricing initiative. Technical report, University Transportation Research Center.

[Hurdle et al., 1983] Hurdle, V., Hauser, E., and Fargier, P. (1983). Effects of the choice of departure time on road traffic congestion. theoretical approach. *Transportation and traffic theory*, 8:223–263.

[Ison and Rye, 2005] Ison, S. and Rye, T. (2005). Implementing road user charging: the lessons learnt from hong kong, cambridge and central london. *Transport reviews*, 25(4):451–465.

[Jaensirisak et al., 2003] Jaensirisak, S., May, A. D., and Wardman, M. (2003). Acceptability of road user charging: the influence of selfish and social perspectives. In *Acceptability of transport pricing strategies*. Emerald Group Publishing Limited.

[Jaensirisak et al., 2005] Jaensirisak, S., Wardman, M., and May, A. (2005). Explaining variations in public acceptability of road pricing schemes. *Journal of Transport Economics and Policy (JTEP)*, 39(2):127–154.

[Jiang et al., 2011] Jiang, L., Mahmassani, H. S., and Zhang, K. (2011). Congestion pricing, heterogeneous users, and travel time reliability: Multicriterion dynamic user equilibrium model and efficient implementation for large-scale networks. *Transportation research record*, 2254(1):58–67.

[Kaneko and Nakamura, 1979] Kaneko, M. and Nakamura, K. (1979). The nash social welfare function. *Econometrica: Journal of the Econometric Society*, pages 423–435.

[Kirk, 2004] Kirk, D. E. (2004). *Optimal control theory: an introduction*. Courier Corporation.

[Knight, 1924] Knight, F. H. (1924). Some fallacies in the interpretation of social cost. *The Quarterly Journal of Economics*, 38(4):582–606.

[Kockelman and Kalmanje, 2005] Kockelman, K. M. and Kalmanje, S. (2005). Credit-based congestion pricing: a policy proposal and the public's response. *Transportation Research Part A: Policy and Practice*, 39(7-9):671–690.

[Kristoffersson and Engelson, 2018] Kristoffersson, I. and Engelson, L. (2018). Estimating preferred departure times of road users in a large urban network. *Transportation*, 45(3):767–787.

[Kuhn et al., 2017] Kuhn, B., Balke, K., Wood, N., and Colyar, J. (2017). Active traffic management (atm) implementation and operations guide. Technical report, United States. Federal Highway Administration.

[Labbé et al., 1998] Labbé, M., Marcotte, P., and Savard, G. (1998). A bilevel model of taxation and its application to optimal highway pricing. *Management science*, 44(12-part-1):1608–1622.

[Laih, 1994] Laih, C.-H. (1994). Queueing at a bottleneck with single-and multi-step tolls. *Transportation Research Part A: Policy and Practice*, 28(3):197–208.

[Lawphongpanich and Yin, 2010] Lawphongpanich, S. and Yin, Y. (2010). Solving the pareto-improving toll problem via manifold suboptimization. *Transportation Research Part C: Emerging Technologies*, 18(2):234–246.

[Layard et al., 2008] Layard, R., Mayraz, G., and Nickell, S. (2008). The marginal utility of income. *Journal of Public Economics*, 92(8-9):1846–1857.

[Li et al., 2020] Li, Z.-C., Huang, H.-J., and Yang, H. (2020). Fifty years of the bottleneck model: A bibliometric review and future research directions. *Transportation research part B: methodological*, 139:311–342.

[Lindsey et al., 2012] Lindsey, C. R., Van den Berg, V. A., and Verhoef, E. T. (2012). Step tolling with bottleneck queuing congestion. *Journal of Urban Economics*, 72(1):46–59.

[Lindsey, 2004] Lindsey, R. (2004). Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. *Transportation science*, 38(3):293–314.

[Lindsey and Verhoef, 2001] Lindsey, R. and Verhoef, E. (2001). Traffic congestion and congestion pricing. in: Handbook of transport systems and traffic control. *Publication of: Elsevier Scientific Publishing Company.*

[Lindsney and Verhoef, 2001] Lindsney, R. and Verhoef, E. (2001). *Traffic congestion and congestion pricing.* Emerald Group Publishing Limited.

[Lipow, 2008] Lipow, G. W. (2008). Price–elasticity of energy demand: A bibliography. *Carbon Tax Center (www. carbontax. org).*

[Liu et al., 2009] Liu, Y., Guo, X., and Yang, H. (2009). Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks. *NETNOMICS: Economic Research and Electronic Networking*, 10(1):123–140.

[Liu and Yang, 2021] Liu, Y. and Yang, Z. (2021). Information provision and congestion pricing in a risky two-route network with heterogeneous travelers. *Transportation Research Part C: Emerging Technologies*, 128:103083.

[Liu et al., 2014] Liu, Z., Wang, S., and Meng, Q. (2014). Toll pricing framework under logit-based stochastic user equilibrium constraints. *Journal of Advanced Transportation*, 48(8):1121–1137.

[Louviere and Eagle, 2006] Louviere, J. J. and Eagle, T. (2006). Confound it! that pesky little scale constant messes up our convenient assumptions. In *Sawtooth software conference*. Sawtooth Software Inc.

[Maher et al., 2005] Maher, M., Stewart, K., and Rosa, A. (2005). Stochastic social optimum traffic assignment. *Transportation Research Part B: Methodological*, 39(8):753–767.

[Mayeres and Proost, 2001] Mayeres, I. and Proost, S. (2001). Marginal tax reform, externalities and income distribution. *Journal of public economics*, 79(2):343–363.

[McFadden, 2001] McFadden, D. (2001). Economic choices. *American economic review*, 91(3):351–378.

[McFadden, 2017] McFadden, D. (2017). Foundations of welfare economics and product market applications. Technical report, National Bureau of Economic Research.

[McNeil et al., 2021] McNeil, N., MacArthur, J., and Tan, H. (2021). New mobility for all: Evaluation of a transportation incentive program for residents of affordable housing in portland, or.

[Migdalas, 1995] Migdalas, A. (1995). Bilevel programming in traffic planning: Models, methods and challenge. *Journal of global optimization*, 7(4):381–405.

[Miralinaghi and Peeta, 2016] Miralinaghi, M. and Peeta, S. (2016). Multi-period equilibrium modeling planning framework for tradable credit schemes. *Transportation Research Part E: Logistics and Transportation Review*, 93:177–198.

[Miralinaghi et al., 2019] Miralinaghi, M., Peeta, S., He, X., and Ukkusuri, S. V. (2019). Managing morning commute congestion with a tradable credit scheme under commuter heterogeneity and market loss aversion behavior. *Transportmetrica B: Transport Dynamics*, 7(1):1780–1808.

[Miyao and Shapiro, 1981] Miyao, T. and Shapiro, P. (1981). Discrete choice and variable returns to scale. *International Economic Review*, pages 257–273.

[Newell et al., 2005] Newell, R. G., Sanchirico, J. N., and Kerr, S. (2005). Fishing quota markets. *Journal of environmental economics and management*, 49(3):437–462.

[Nie, 2012] Nie, Y. M. (2012). Transaction costs and tradable mobility credits. *Transportation Research Part B: Methodological*, 46(1):189–203.

[Nie and Liu, 2010] Nie, Y. M. and Liu, Y. (2010). Existence of self-financing and pareto-improving congestion pricing: Impact of value of time distribution. *Transportation Research Part A: Policy and Practice*, 44(1):39–51.

[Noordegraaf et al., 2014] Noordegraaf, D. V., Annema, J. A., and van Wee, B. (2014). Policy implementation lessons from six road pricing cases. *Transportation Research Part A: Policy and Practice*, 59:172–191.

[Parry and Bento, 2001] Parry, I. W. and Bento, A. (2001). Revenue recycling and the welfare effects of road pricing. *Scandinavian Journal of economics*, 103(4):645–671.

[Pigou, 1920] Pigou, A. C. (1920). *The economics of welfare.* Macmillan.

[Price, 2013] Price, K. V. (2013). Differential evolution. In *Handbook of optimization*, pages 187–214. Springer.

[Rapoport et al., 2014] Rapoport, A., Gisches, E. J., Daniel, T., and Lindsey, R. (2014). Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment. *Transportation Research Part B: Methodological*, 68:154–172.

[Raux, 2007] Raux, C. (2007). Tradable driving rights in urban areas: their potential for tackling congestion and traffic-related pollution. *The Implementation and Effectiveness of Transport Demand Management Measures*, page 95.

[Rawls, 2020] Rawls, J. (2020). *A theory of justice.* Harvard university press.

[Rotaris et al., 2010] Rotaris, L., Danielis, R., Marcucci, E., and Massiani, J. (2010). The urban road pricing scheme to curb pollution in milan, italy: Description, impacts and preliminary cost–benefit analysis assessment. *Transportation Research Part A: Policy and Practice*, 44(5):359–375.

[Ruggles et al., 2021] Ruggles, S., Flood, S., Foster, S., Goeken, R., Pacas, J., Schouweiler, M., and Sobek, M. (2021). Ipums usa: Version 11.0. Minneapolis, MN: IPUMS, 2021. https://doi.org/10.18128/D010.V11.0.

[Saleh and Sammer, 2016] Saleh, W. and Sammer, G. (2016). *Travel demand management and road user pricing: success, failure and feasibility.* Routledge.

[Santos and Fraser, 2006] Santos, G. and Fraser, G. (2006). Road pricing: lessons from london. *Economic Policy*, 21(46):264–310.

[Sasic and Habib, 2013] Sasic, A. and Habib, K. N. (2013). Modelling departure time choices by a heteroskedastic generalized logit (het-genl) model: An investigation on home-based commuting trips in the greater toronto and hamilton area (gtha). *Transportation Research Part A: Policy and Practice*, 50:15–32.

[Schaller, 2010] Schaller, B. (2010). New york city's congestion pricing experience and implications for road pricing acceptance in the united states. *Transport Policy*, 17(4):266–273.

[Schrank et al., 2015] Schrank, D., Eisele, B., Lomax, T., and Bak, J. (2015). 2015 urban mobility scorecard.

[Schuitema et al., 2010] Schuitema, G., Steg, L., and Rothengatter, J. (2010). The acceptability, personal outcome expectations, and expected effects of transport pricing policies. *Journal of Environmental Psychology*, 30(4):587–593.

[Seale et al., 2003] Seale, C., Gobo, G., Gubrium, J. F., and Silverman, D. (2003). *Qualitative research practice*. Sage.

[Sen, 2009] Sen, A. K. (2009). *The idea of justice*. Harvard University Press.

[Shepherd and Sumalee, 2004] Shepherd, S. and Sumalee, A. (2004). A genetic algorithm based approach to optimal toll level and location problems. *Networks and Spatial Economics*, 4(2):161–179.

[Sinha et al., 2017] Sinha, A., Malo, P., and Deb, K. (2017). A review on bilevel optimization: from classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295.

[Small, 1982] Small, K. A. (1982). The scheduling of consumer activities: work trips. *The American Economic Review*, 72(3):467–479.

[Small, 1992] Small, K. A. (1992). Using the revenues from congestion pricing. *Transportation*, 19(4):359–381.

[Small, 2012] Small, K. A. (2012). Valuation of travel time. *Economics of transportation*, 1(1-2):2–14.

[Small, 2015] Small, K. A. (2015). The bottleneck model: An assessment and interpretation. *Economics of Transportation*, 4(1-2):110–117.

[Small et al., 2007] Small, K. A., Verhoef, E. T., and Lindsey, R. (2007). *The economics of urban transportation*. Routledge.

[Small et al., 2005] Small, K. A., Winston, C., and Yan, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4):1367–1382.

[Small and Yan, 2001] Small, K. A. and Yan, J. (2001). The value of "value pricing" of roads: Second-best pricing and product differentiation. *Journal of Urban Economics*, 49(2):310–336.

[Smith, 1984] Smith, M. J. (1984). The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation science*, 18(4):385–394.

[Song et al., 2018] Song, X., Danaf, M., Atasoy, B., and Ben-Akiva, M. (2018). Personalized menu optimization with preference updater: a boston case study. *Transportation Research Record*, 2672(8):599–607.

[Sørensen et al., 2014] Sørensen, C. H., Isaksson, K., Macmillen, J., Åkerman, J., and Kressler, F. (2014). Strategies to manage barriers in policy formation and implementation of road pricing packages. *Transportation Research Part A: Policy and Practice*, 60:40–52.

[Tian et al., 2019] Tian, Y., Chiu, Y.-C., and Sun, J. (2019). Understanding behavioral effects of tradable mobility credit scheme: An experimental economics approach. *Transport Policy*, 81:1–11.

[Ubbels and Verhoef, 2006] Ubbels, B. and Verhoef, E. (2006). Behavioural responses to road pricing. empirical results from a survey among dutch car owners.

[United Nations Department of Economical and Social Affairs (DESA), 2018] United Nations Department of Economical and Social Affairs (DESA) (2018). World urbanization prospects 2018.

[U.S. Bureau of Labor Statistics, 2019] U.S. Bureau of Labor Statistics (2019). Consumer expenditures - 2019, "https://www.bls.gov/news.release/pdf/cesan.pdf". Page 5. Accessed Dec. 29, 2021.

[Van den Berg and Verhoef, 2011] Van den Berg, V. and Verhoef, E. T. (2011). Congestion tolling in the bottleneck model with heterogeneous values of time. *Transportation Research Part B: Methodological*, 45(1):60–78.

[Van Den Berg and Verhoef, 2011] Van Den Berg, V. and Verhoef, E. T. (2011). Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay. *Journal of Public Economics*, 95(7-8):983–992.

[Van Dender, 2003] Van Dender, K. (2003). Transport taxes with multiple trip purposes. *Scandinavian journal of economics*, 105(2):295–310.

[Verhoef et al., 1995] Verhoef, E., Nijkamp, P., and Rietveld, P. (1995). Second-best regulation of road transport externalities. *Journal of transport economics and policy*, pages 147–167.

[Verhoef et al., 1997] Verhoef, E., Nijkamp, P., and Rietveld, P. (1997). Tradeable permits: their potential in the regulation of road transport externalities. *Environment and Planning B: Planning and Design*, 24(4):527–548.

[Verhoef, 2002a] Verhoef, E. T. (2002a). Second-best congestion pricing in general networks. heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B: Methodological*, 36(8):707–729.

[Verhoef, 2002b] Verhoef, E. T. (2002b). Second-best congestion pricing in general networks. heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B: Methodological*, 36(8):707–729.

[Verhoef et al., 1996] Verhoef, E. T., Emmerink, R. H., Nijkamp, P., and Rietveld, P. (1996). Information provision, flat and fine congestion tolling and the efficiency of road usage. *Regional Science and Urban Economics*, 26(5):505–529.

[Verhoef and Small, 2003] Verhoef, E. T. and Small, K. A. (2003). Product differentiation on roads: Constrained congestion pricing with heterogeneous users.

[Verhoef and Small, 2004] Verhoef, E. T. and Small, K. A. (2004). Product differentiation on roads. *Journal of Transport Economics and Policy (JTEP)*, 38(1):127–156.

[Vicente et al., 1994] Vicente, L., Savard, G., and Júdice, J. (1994). Descent approaches for quadratic bilevel programming. *Journal of optimization theory and applications*, 81(2):379–399.

[Vickrey, 1973] Vickrey, W. (1973). *Pricing, metering, and efficiently using urban transportation facilities*. Number 476.

[Vickrey, 1969] Vickrey, W. S. (1969). Congestion theory and transport investment. *The American Economic Review*, 59(2):251–260.

[Wada and Akamatsu, 2013] Wada, K. and Akamatsu, T. (2013). A hybrid implementation mechanism of tradable network permits system which obviates path enumeration: An auction mechanism with day-to-day capacity control. *Procedia-Social and Behavioral Sciences*, 80:304–326.

[Wang et al., 2014a] Wang, G., Gao, Z., Xu, M., and Sun, H. (2014a). Models and a relaxation algorithm for continuous network design problem with a tradable credit scheme and equity constraints. *Computers & Operations Research*, 41:252–261.

[Wang et al., 2020] Wang, G., Xu, M., Grant-Muller, S., and Gao, Z. (2020). Combination of tradable credit scheme and link capacity improvement to balance economic growth and environmental management in sustainable-oriented transport development: A bi-objective bi-level programming approach. *Transportation Research Part A: Policy and Practice*, 137:459–471.

[Wang et al., 2018] Wang, J.-P., Liu, T.-L., and Huang, H.-J. (2018). Tradable od-based travel permits for bi-modal traffic management with heterogeneous users. *Transportation Research Part E: Logistics and Transportation Review*, 118:589–605.

[Wang et al., 2014b] Wang, J. Y., Ehrgott, M., Dirks, K. N., and Gupta, A. (2014b). A bilevel multi-objective road pricing model for economic, environmental and health sustainability. *Transportation Research Procedia*, 3:393–402.

[Wang et al., 2012] Wang, X., Yang, H., Zhu, D., and Li, C. (2012). Tradable travel credits for congestion management with heterogeneous users. *Transportation Research Part E: Logistics and Transportation Review*, 48(2):426–437.

[Wardrop, 1952] Wardrop, J. G. (1952). Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362.

[Weitzman, 1974] Weitzman, M. L. (1974). Prices vs. quantities. *The review of economic studies*, 41(4):477–491.

[White, 2016] White, V. (2016). Revised departmental guidance on valuation of travel time in economic analysis. *Office of the Secretary of Transportation, US Department of Transportation, Available at: transportation. gov/sites/dot. gov/files/docs/2016% 20Revised% 20Value% 20of% 20Travel% 20Time% 20Guidance. pdf.*

[Wittman and Belobaba, 2017] Wittman, M. D. and Belobaba, P. P. (2017). Personalization in airline revenue management–heuristics for real-time adjustment of availability and fares. *Journal of Revenue and Pricing Management*, 16(4):376–396.

[Wong, 1997] Wong, J.-T. (1997). Basic concepts for a system for advance booking for highway use. *Transport Policy*.

[Wu et al., 2012] Wu, D., Yin, Y., Lawphongpanich, S., and Yang, H. (2012). Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. *Transportation Research Part B: Methodological*, 46(9):1273–1287.

[Xiao et al., 2019] Xiao, F., Long, J., Li, L., Kou, G., and Nie, Y. (2019). Promoting social equity with cyclic tradable credits. *Transportation Research Part B: Methodological*, 121:56–73.

[Xiao et al., 2012] Xiao, F., Shen, W., and Zhang, H. M. (2012). The morning commute under flat toll and tactical waiting. *Transportation Research Part B: Methodological*, 46(10):1346–1359.

[Xiao and Zhang, 2014] Xiao, F. and Zhang, H. (2014). Pareto-improving toll and subsidy scheme on transportation networks. *European Journal of Transport and Infrastructure Research*, 14(1).

[Xie et al., 2020] Xie, Y., Zhang, Y., Akkinepally, A. P., and Ben-Akiva, M. (2020). Personalized choice model for managed lane travel behavior. *Transportation research record*, 2674(7):442–455.

[Xiong et al., 2020] Xiong, C., Shahabi, M., Zhao, J., Yin, Y., Zhou, X., and Zhang, L. (2020). An integrated and personalized traveler information and incentive scheme for energy efficient mobility systems. *Transportation Research Part C: Emerging Technologies*, 113:57–73.

[Yan and Lam, 1996] Yan, H. and Lam, W. H. (1996). Optimal road tolls under conditions of queueing and congestion. *Transportation Research Part A: Policy and Practice*, 30(5):319–332.

[Yang, 1999] Yang, H. (1999). System optimum, stochastic user equilibrium, and optimal link tolls. *Transportation science*, 33(4):354–360.

[Yang and Bell, 1997] Yang, H. and Bell, M. G. (1997). Traffic restraint, road pricing and network equilibrium. *Transportation Research Part B: Methodological*, 31(4):303–314.

[Yang and Wang, 2011a] Yang, H. and Wang, X. (2011a). Managing network mobility with tradable credits. *Transportation Research Part B: Methodological*, 45(3):580–594.

[Yang and Wang, 2011b] Yang, H. and Wang, X. (2011b). Managing network mobility with tradable credits. *Transportation Research Part B: Methodological*, 45(3):580–594.

[Ye and Yang, 2013] Ye, H. and Yang, H. (2013). Continuous price and flow dynamics of tradable mobility credits. *Procedia-Social and Behavioral Sciences*, 80:61–78.

[Yin, 2002] Yin, Y. (2002). Multiobjective bilevel optimization for transportation planning and management problems. *Journal of advanced transportation*, 36(1):93–105.

[You et al., 2019] You, L., Zhao, F., Cheah, L., Jeong, K., Zegras, P. C., and Ben-Akiva, M. (2019). A generic future mobility sensing system for travel data collection, management, fusion, and visualization. *IEEE Transactions on Intelligent Transportation Systems*, 21(10):4149–4160.

[Zhang and Yang, 2004] Zhang, X. and Yang, H. (2004). The optimal cordon-based network congestion pricing problem. *Transportation Research Part B: Methodological*, 38(6):517–537.

[Zhang, 2019] Zhang, Y. (2019). *Real-time personalized toll optimization based on traffic predictions*. PhD thesis, Massachusetts Institute of Technology.

[Zhao et al., 2008] Zhao, Y., Kockelman, K. M., and Karlstrom, A. (2008). Welfare calculations in discrete choice settings: Role of error term correlation. In *Proceedings of the transportation research board 87th annual meeting*. Citeseer.

[Zhu et al., 2015] Zhu, D.-L., Yang, H., Li, C.-M., and Wang, X.-L. (2015). Properties of the multiclass traffic network equilibria under a tradable credit scheme. *Transportation Science*, 49(3):519–534.

[Zhu et al., 2020] Zhu, X., Wang, F., Chen, C., and Reed, D. D. (2020). Personalized incentives for promoting sustainable travel behaviors. *Transportation Research Part C: Emerging Technologies*, 113:314–331.