Real-time Personalized Toll Optimization based on Traffic Predictions

by

Yundi Zhang

B.Eng., Hong Kong University of Science and Technology (2014)

M.S., Massachusetts Institute of Technology (2017)

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Signature of Author: Signature redacted
Department of Civil and Environmental Engineering
August 16, 2019

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

Certified by: Signature redacted
Arun Akkinepally
Research Scientist
Thesis Co-Supervisor

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

Real-time Personalized Toll Optimization based on Traffic Predictions

by

Yundi Zhang

Submitted to the Department of Civil and Environmental Engineering
on Aug 16, 2019, in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Transportation

Abstract

Road pricing is a traffic congestion management strategy that alters traffic demand and raises funds for transportation supply improvements. Compared to static pricing and reactive dynamic pricing, proactive dynamic pricing is most effective in achieving traffic management objectives, as the toll is based on traffic predictions that incorporate real-time information. We investigate a proactive toll pricing framework where the toll is optimized in real time based on traffic predictions generated by a dynamic traffic assignment (DTA) system. Toll optimization performance relies on accurate predictions, which is backed by the online calibration of the DTA system. We develop enhanced online calibration methodologies featuring a heuristic technique to calibrate supply parameters and improve the prediction accuracy of traffic speed. We test online calibration using real data from a real network consisting of managed lanes and general-purpose lanes. We find the methodologies improve estimation and prediction accuracies of flow and speed. We then formulate toll pricing as an optimization problem to maximize expected revenue, subject to network condition requirements and tolling regulations. We test the proactive toll pricing system in a closed-loop evaluation framework where a microscopic simulator is used to mimic the real network. We perform tests in multiple demand and supply scenarios and find that the system generates higher revenue when online calibration is enabled.

Growing applications of electronic toll collection enrich disaggregate trip data, making it possible to improve traffic management by personalized toll pricing. We develop a personalized toll pricing system by extending the original system to a two-level framework, where a new personalized discount module generates discount offers for a subset of individuals, while the original optimization module optimizes the displayed toll rate and a control parameter that affects how much discount to offer. Discount also depends on individual traveler's choice behavior, represented by an enhanced route choice model that captures heterogeneities among individuals. We use real personalized trip records to estimate the choice model. We find that variables generated from individuals' trip history are capable of capturing heterogeneities among individuals. We test the personalized toll pricing system and find it improves optimization objective compared to non-personalized pricing.

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

Thesis Co-Supervisor: Arun Akkinepally
Title: Research Scientist

Acknowledgments

First, I would like to express my sincere gratitude to my advisor Prof. Moshe Ben-Akiva. Your insights have greatly inspired my research, and your support has encouraged me to overcome challenges in my PhD life. I am very honored to be your student.

I would like to thank my thesis co-supervisor Dr. Arun Akkinepally and other committee members: Prof. Carolina Osorio and Prof. Bilge Atasoy, for your insightful suggestions on my thesis. I would also thank Prof. Bilge Atasoy, Dr. Arun Akkinepally, and Dr. Ravi Seshadri, for working closely with me throughout my Master and PhD life at MIT.

I am very grateful to Katherine Rosa and Eunice Kim, who have been taking care of the ITS lab and creating a loving atmosphere. I also thank Kiley Clapper and Max Martelli for making the administrative process smooth and efficient.

Cintra/Ferrovial offered important supports for this research. Thanks to Ricardo Sanchez, John Brady, Thu Hoang, Wei He, and Ning Zhang for their professional insights and feedback on this research.

My thanks also go to current and former members of the ITS lab: Haizheng Zhang, Xiang Song, Mazen Danaf, Shi Wang, Peiyu Jing, Yifei Xie, Siyu Chen, Youssef Aboutaleb, Isabel Viegas, Eytan Gross, Yihang Sui, Ajinkya Ghorpade, Kenneth Koh, Samarth Gupta, Carlos Azevedo, Andrea Araldo, Kyungsoo Jeong, as well as my friends and lovely neighbors Jordan Goldstein and Qin Zhang.

Last but not least, I want to express my wholehearted thanks to my parents Prof. Jian Zhang and Prof. Fan He. With the two of you being the strongest support, I can always freely explore the fantastic world and pursue happiness in my life.

Table of Contents

Chapter 1 Introduction ..... - 11 -

1.1 Background and Motivation ..... - 11 -

1.1.1 Congestion and Road Pricing ..... - 11 -

1.1.2 Toll Managed Lanes ..... - 12 -

1.1.3 Proactive Pricing based on Dynamic Traffic Assignment System ..... - 13 -

1.1.4 Personalized Pricing ..... - 15 -

1.2 Proposed Framework ..... - 17 -

1.2.1 The Managed Lane Network ..... - 18 -

1.2.2 Proposed Real-time Personalized Toll Pricing System ..... - 21 -

1.3 Summary of Findings ..... - 25 -

1.4 Thesis Outline ..... - 25 -

Chapter 2 Literature Review ..... - 27 -

2.1 Calibration of the Dynamic Traffic Assignment System ..... - 27 -

2.1.1 Offline Calibration ..... - 28 -

2.1.2 Online Calibration ..... - 29 -

2.2 Road Pricing and Toll Managed Lanes ..... - 31 -

2.2.1 Reactive Toll Pricing Schemes ..... - 31 -

2.2.2 Proactive Toll Pricing Schemes ..... - 32 -

2.3 Discrete Choice Models ..... - 33 -

2.3.1 Modeling of Systematic and Random Heterogeneities ..... - 33 -

2.3.2 Discrete Choice Models for Route Choice ..... - 34 -

2.3.3 Discrete Choice Models for Managed Lanes ..... - 34 -

2.4 Personalization in Transportation and Price Discrimination ..... - 35 -

2.4.1 Price Discrimination and Personalized Pricing ..... - 35 -

2.4.2 Personalized Pricing/Discounting/Incentives in Transportation ..... - 37 -

2.5 Summary ..... - 38 -

Chapter 3 Real-time Toll Optimization ..... - 41 -
3.1 Introduction ..... - 41 -
3.2 Optimization Formulation ..... - 43 -
3.3 DTA-based Solution Framework ..... - 45 -
3.4 Calibration and Prediction Models for the DTA System ..... - 48 -
3.4.1 State Estimation ..... - 48 -
3.4.2 Prediction Model ..... - 49 -
3.4.3 Evaluation Metric for Calibration and Prediction ..... - 50 -
3.4.4 Online Calibration of Supply Model ..... - 50 -
3.5 Closed-loop Evaluation Framework ..... - 52 -
3.6 Case Study ..... - 53 -
3.6.1 Network and Data ..... - 53 -
3.6.2 Offline Calibration ..... - 55 -
3.6.3 Online Calibration and Prediction ..... - 57 -
3.6.4 Toll Optimization ..... - 62 -
3.6.5 Toll Optimization under Multiple Scenarios ..... - 65 -
3.7 Conclusion ..... - 68 -

Chapter 4 Real-time Personalized Toll Optimization ..... - 71 -
4.1 Introduction ..... - 71 -
4.2 Optimization Formulation ..... - 72 -
4.2.1 Naïve Formulation of Personalized Pricing ..... - 73 -
4.2.2 Two-level Formulation of Personalized Pricing ..... - 74 -
4.2.3 Personalized Pricing Policy ..... - 77 -
4.2.4 Formulation of Personalized Discounting ..... - 82 -
4.2.5 Alternative Personalized Discount Policies ..... - 83 -
4.2.6 Generalization to a Full Managed Lane Network ..... - 85 -
4.3 DTA-based Solution Framework ..... - 86 -
4.4 Personalized Choice Model ..... - 87 -
4.4.1 Model Specification ..... - 88 -
4.4.2 Model Estimation and Application in the Toll Pricing System ..... - 92 -

4.5 Case Study ..... - 93 -
4.5.1 Network and Data ..... - 93 -
4.5.2 Personalized Choice Model ..... - 94 -
4.5.3 Offline Calibration ..... - 100 -
4.5.4 Toll Optimization ..... - 101 -
4.6 Conclusion ..... - 108 -

Chapter 5 Conclusions ..... - 111 -
5.1 Contributions ..... - 111 -
5.2 Findings ..... - 113 -
5.3 Future Research Directions ..... - 114 -

Bibliography ..... - 118 -

List of Figures

Figure 1.1: Illustration of a toll managed lane network ..... - 19 -
Figure 1.2: The proposed real-time personalized toll pricing system ..... - 21 -
Figure 1.3: The non-personalized toll pricing system ..... - 24 -
Figure 3.1: Overview of the real-time non-personalized toll pricing system ..... - 46 -
Figure 3.2: Calibration of DynaMIT ..... - 51 -
Figure 3.3: The North Tarrant Express network ..... - 54 -
Figure 3.4: The data points and the estimated supply curve for selected road segments ..... - 56 -
Figure 3.5: Calibration results on Day 6 ..... - 60 -
Figure 3.6: Traffic prediction results on Day 6 ..... - 61 -
Figure 3.7: Comparison between optimized and base toll and corresponding revenue ..... - 63 -
Figure 3.8: Flow on ML ..... - 64 -
Figure 3.9: Speed on GPL ..... - 64 -
Figure 3.10: Toll rates in different scenarios ..... - 66 -
Figure 4.1: The system optimization module ..... - 87 -
Figure 4.2: Time-of-day-specific ASC ..... - 96 -
Figure 4.3: Relationship between toll rate and scaled toll ..... - 97 -
Figure 4.4: ML market share before and after choice model calibration ..... - 101 -
Figure 4.5: Comparison of toll rates between Test A and Test B ..... - 103 -
Figure 4.6: Comparison of toll rates between Test B and Test C ..... - 103 -
Figure 4.7: Comparison of toll rates between Test C and Test D ..... - 105 -
Figure 4.8: Comparison of revenue in each test ..... - 105 -
Figure 4.9: Distribution of the toll rates presented to travelers in each test ..... - 106 -
Figure 4.10: Distribution of the toll charged to ML users ..... - 106 -
Figure 4.11: Distribution of travel time ..... - 107 -

List of Tables

Table 3-1: Statistics of offline estimated supply parameters ..... - 55 -
Table 3-2: Calibration and prediction accuracies ..... - 58 -
Table 3-3: Experimental scenarios ..... - 65 -
Table 3-4: Revenue improvement compared to the base toll under experimental scenarios ..... - 65 -
Table 4-1: Estimate of parameters related to ASC ..... - 96 -
Table 4-2: Estimate of parameters related to sensitivities to toll and traffic ..... - 98 -
Table 4-3: Estimate of parameters related to ISC ..... - 99 -
Table 4-4: Statistics measuring heterogeneity in travelers' sensitivities and preferences ..... - 99 -
Table 4-5: Statistics of the model's performance on training data and testing data ..... - 100 -
Table 4-6: Toll rate generation method in each test ..... - 102 -
Table 4-7: Summary statistics of the four tests ..... - 108 -

Chapter 1 Introduction

This chapter starts with the background for this research and the motivation of the proposed toll pricing framework. Then we introduce the non-personalized and personalized toll pricing framework we propose.

1.1 Background and Motivation

In this section, we introduce road pricing and managed lanes as an application of road pricing. We then discuss different pricing strategies and the motivation for proposing a proactive pricing strategy. We also talk about the possibility and potential benefit of personalized pricing, as well as the challenges of applying personalized pricing in road pricing.

1.1.1 Congestion and Road Pricing

Traffic congestion is a long-standing and ever-growing issue in modern travelers' lives. Growth of economy and population, increase in car ownership, and urban sprawl are driving congestion to grow rapidly in recent years. According to statistics from the FHWA (Taylor, 2009; Federal Highway Administration, 2017), from 2009 to 2017, the average duration of daily congestion slightly reduces from 4.33 hours to 4.28 hours, while the travel time index measuring peak versus off-peak travel times increased from 1.18 to 1.33, and the variability of travel time increased from 1.43 to 2.15. The negative externalities from congestion are multifaced. One externality is the loss of productivity, which costs US drivers a total of nearly 87 billion dollars in 2018 (INRIX, 2019). Another externality is the waste of energy. In 2014, congestion caused an extra expenditure of 3.1 billion gallons of fuel in the US (Schrank et al., 2015). Other externalities include pollution, emissions, road rage, etc.

Congestion management aims to improve transportation system performance and reduce traffic congestion by altering traffic demand and transportation supply. Among congestion management schemes, road pricing (i.e., tolling) is a commonly applied strategy, which may aim to generate revenue to recover road construction and maintenance costs, thus incentivize improvement to transportation supply, as well as managing congestion by altering temporal and spatial dimensions of travel behaviors, and travelers' decisions on mode choice or whether to travel (Saleh and Sammer, 2009).

Traditionally a fixed toll rate is charged on each road segment, and the price is usually based on distance and cost of construction and maintenance. Time-of-day toll pricing is also commonly used to control the temporal distribution of demand. By charging a higher price during peak periods, traffic congestion can be reduced. In recent years, dynamic toll pricing has been extensively studied (de Palma and Lindsey, 2011), which aims to achieve more efficient traffic management by considering real-time changes in traffic demand and supply. Thanks to traffic sensing technologies, the road operator is able to monitor traffic conditions in real-time and adjust the toll rate accordingly. Applications of dynamic toll pricing arise in many cities, mostly on toll managed lane networks.

1.1.2 Toll Managed Lanes

Managed lane is a type of freeway lane that is operated with some traffic management schemes. Access to managed lanes is often restricted to certain vehicles, for example, high occupancy vehicles (HOV), buses, or vehicles traveling on the peak direction. In parallel to managed lanes are general-purpose lanes that are accessible to all vehicles. Some managed lanes are separated from general-purpose lanes with solid white lines, while others are isolated with physical separation. A managed lane network refers to a freeway system that consists of managed lanes and general-purpose lanes in parallel to each other.

Toll managed lanes are accessible to travelers who pay a toll. Toll collection relies on an electronic toll collection (ETC) system so that the traffic does not slow down. Traveler with an ETC account will be identified with the toll transponder mounted on the vehicle and get charged automatically

when passing through a toll gantry. The gantry is also equipped with cameras that take photos for any vehicle without a transponder. The license plate is then identified, and a bill is mailed to the address of vehicle registration. For toll managed lanes on which dynamic toll pricing is applied, the toll rate is controlled by an operator in real-time based on traffic conditions measured by flow and speed sensors along the managed lanes and general-purpose lanes. All travelers on the network can see the real-time toll rate displayed on a sign before they decide whether to take toll managed lanes or general-purpose lanes. Price for larger vehicles is higher, while HOVs usually get a discount and emergency vehicles may be free. This thesis does not involve other types of managed lanes, so the term managed lanes always refers to toll managed lanes.

In many applications, toll managed lanes arise from an infrastructure expansion project, in which a private company invests in constructing those lanes in parallel to an existing freeway corridor. The company is then licensed to collect tolls to pay off construction and maintenance costs and generate additional revenue. Such a project benefits the society by expanding network capacity while benefiting the investing company with toll revenue. Under such a business model, revenue maximization is often an important consideration to the operator, and the government usually imposes tolling regulations for the welfare of society. The operator decides what price to charge given regulatory requirements on traffic flow, speed, and toll rate.

1.1.3 Proactive Pricing based on Dynamic Traffic Assignment System

In current practices, dynamic toll pricing is mostly based on reactive algorithms, so that toll prices are adjusted in reaction to observed traffic conditions. For example, the toll rate is raised when managed lanes are congested, and vice versa. While a negative feedback control algorithm can be designed to maintain traffic flow and speed effectively, it is difficult to design a control algorithm that aligns with an objective of boosting revenue.

On the contrary, we advocate proactive decision making, which means, toll prices are determined based on predicted traffic conditions instead of observed ones. Under the proactive scheme, it is possible to formulate an optimization problem that explicitly maximizes certain metrics measured in the future, subject to constraints on expected traffic conditions and other metrics. We propose,

develop, implement, and test a real-time toll pricing system, where toll pricing is formulated as an optimization problem to maximize expected revenue in the near future subject to regulations on toll and network conditions. The expected revenue is measured based on traffic predictions. It is worth mentioning that while this research focuses on a revenue maximization objective given regulations, the system we design, after modest modification, can accommodate alternative objectives and constraints.

We apply a dynamic traffic assignment (DTA) system for traffic prediction. A DTA system consists of a demand model and a supply model. For a simulation-based DTA system, the demand model simulates individual travelers' route choice based on traffic conditions and assigns the origin-destination (OD) demand to flow on each path. The supply model simulates traffic flow on each road segment based on the path flow and estimates traffic speed. A macroscopic supply model applies a macroscopic traffic dynamics model (e.g., the Greenshields model) to estimate traffic speed, while a microscopic model simulated individual vehicles' movements to obtain traffic speed. As the simulated traffic conditions affect route choice and are also conditional on route choice, traffic simulation involves interactions between the demand and supply models.

The DTA system we apply runs in rolling horizons. For each rolling time period (e.g., 5 minutes), the system receives new real-time information (e.g., sensor flow and speed measurements) from the network and performs traffic simulations in two stages: state estimation and traffic prediction. State estimation is the process of estimating the current traffic state. State refers to the demand and supply conditions and is represented with demand and supply parameters in the DTA system. State estimation relies on the online calibration module to find the best demand and supply parameters that match simulated traffic (flow and speed) with real sensor measurements. Then the DTA system predicts demand and supply parameters for the near future and predicts traffic conditions by simulating the future.

Proactive toll pricing relies on the DTA system's capability to understand and predict traffic conditions. Thus, it is vital to obtain an accurate estimation of demand and supply parameters through online calibration. While calibration of demand to match flow measurements has been successfully conducted in many studies using the generalized least squares (GLS) algorithm, there

is limited research on calibrating supply parameters and accurately matching speed measurements. We develop enhanced online calibration methodologies for the DTA system, featuring a heuristic to calibrate supply parameters online.

1.1.4 Personalized Pricing

The ETC system, together with automatic vehicle identification (AVI) sensors that are available in many networks, in addition to flow and speed sensors, make available individual-level trip data. Such data include an individual's choice between managed lanes and general-purpose lanes in each trip, as well as the attributes of each alternative and scenario-specific variables at the time when the choice is made. Each individual's characteristics may be generated with an abundance of trip data. Such data allow researchers to develop personalized route choice models. A personalized choice model makes it possible to identify the heterogeneity among individuals and choice scenarios, and thus to apply personalized toll pricing for further improvement of traffic management.

Price discrimination has been researched extensively and affects many aspects of our lives. Pigou (1920) proposes the definitions of three levels of price discrimination, which are, personalized pricing, product versioning, and group pricing. Among them, personalized pricing (i.e., first-degree price discrimination or perfect price discrimination) is the most flexible and most effective in boosting revenue. With an objective of maximizing revenue, the best pricing strategy, in general, is to set the price for each individual at the point where we expect maximum revenue from that individual.

An individual's decision to purchase a product (or to take the managed lanes) may be modeled with a binary choice model, where Alternative 1 represents the option to purchase, and Alternative 2 stands for the option not to buy. A utility function is specified for Alternative 1, which measures the benefits and costs of choosing that alternative over the other. The utility of Alternative 2 is thus normalized to zero. An individual chooses the alternative with higher utility.

A fixed utility model specifies the utility function in a way such that each component of the utility function is fixed. Based on this model, an individual makes the purchase if the price does not exceed a certain value (defined as his/her willingness to pay, WTP) and does not make the purchase if the price is higher than that value. WTP represents the portion of utility except for price, or in other words, the net benefit of Alternative 1 without consideration of price. Thus, the personalized pricing policy that maximizes revenue is to charge each individual at his/her WTP.

On the other hand, a random utility model (McFadden, 1975) captures the stochasticity in an individual's decision-making process by decomposing the utility into systematic and random components, where the random component follows a distribution. The logit model is a commonly used random utility model. For a binary choice logit model, the random component in the utility of Alternative 1 is assumed to follow a logistic distribution, while the total utility of Alternative 2 is normalized to zero. Due to the stochasticity in utility, an individual's decision is random, and the probability to choose Alternative 1 is given by a logistic function of the systematic utility. Based on a random utility choice model, the personalized pricing policy that maximizes revenue is to charge each individual at the price where price elasticity of purchase probability is -1.

However, the personalized pricing policy discussed above is only optimal when individuals' decisions to make the purchase is independent of each other, which is valid only when demand-supply interactions are not in effect. In applications where inventory limit is fixed, some individuals' decisions to make the purchase may render other individuals unable to make a purchase. In such cases, dynamic programming is often applied to evaluate the expected revenue from each unit of inventory (i.e., value of inventory), and take it into consideration when determining the best pricing strategy. In the field of road pricing, demand-supply interactions through traffic dynamics are extremely complicated, and there is no literature on personalized road pricing that explicitly takes into consideration the complex and dynamic demand-supply interactions. Dynamic programming requires the inventory limit to be well-defined and known, while there is not a well-defined limit on how many cars may take managed lanes during a given time period, but clearly some individuals' decisions to take managed lanes will affect traffic conditions on the network thus other individuals' utility of choosing managed.

While product versioning and group pricing have been extensively applied in practice, personalized pricing is not as popular due to equity concerns and lack of public acceptance. Charging a higher price for the same product to an individual just because he/she eagerly needs the product is sometimes considered immoral. Individuals who are charged higher than others may develop hate towards the seller. On the other hand, it is considered acceptable if personalized incentives or discount coupons are offered to some individuals, while others pay a full price. Compared to personalized pricing, personalized discounting (or personalized incentives) is more commonly applied, especially in online retailing.

We propose personalized discounting policies for application on managed lane tolling, with an objective of maximizing revenue subject to practical considerations. We develop a personalized toll pricing system that relies on a DTA system to capture interactions between demand and supply and predict traffic, where the demand model is personalized. The new system is built upon the real-time non-personalized toll pricing system discussed in Section 1.1.3.

Price/discount heterogeneity among individuals results from heterogeneity in choice behaviors. Personalized discounting is not meaningful unless behavior heterogeneity among individuals is captured. We develop a personalized choice model that predicts each individual's choice between managed lanes and general-purpose lanes. We specify a logit model that captures the systematic heterogeneities in individuals' preferences to managed lanes and sensitivities to toll and travel time and estimate model parameters with personalized trip data.

1.2 Proposed Framework

We propose, develop, implement, and test a toll pricing system where proactive decisions on toll rates are made. Toll rates are generated by solving an optimization problem, where we apply a DTA system to evaluate the objective function. The performance of the toll pricing system relies on the DTA system's capability to understand and predict traffic conditions, and thus we enhance the online calibration methodology for the DTA system, featuring a heuristic for calibration of supply parameters to match speed measurements.

Offline calibration techniques are applied to calibrate a microscopic traffic simulator to real data on a real network consisting of managed lanes and general-purpose lanes. Toll optimization is tested in closed-loop for this real network, where the microscopic traffic simulator serves as the "real world" and toll optimization is evaluated in the simulator.

We extend the system to allow for personalized discounting. We design personalized discount policies so that the price ratio for each individual is generated by a personalized discount function, which is formulated based on the policy. Only some travelers are eligible to receive a discount, while others always get a price ratio of 1, representing full price. Some parameters of the personalized discount function are optimized in real time, which include the displayed toll rate and a discount control parameter that controls the overall level of discount. Parameters of the personalized pricing function also include the individual's preference to managed lanes and sensitivities to toll rate and traffic condition metrics, and some parameters are different at different times of the day. Those parameters are provided by a choice modeling system that is external to the toll pricing system. We develop a personalized choice model and estimate it with real trip data.

In the remainder of this section, we first describe a managed lane network where the proposed toll pricing system can be applied and then discuss enhancements to the infrastructure that is necessary for a personalized toll pricing system. Then we present the proposed real-time personalized toll pricing system. This system can also accommodate non-personalized toll pricing, which is simply a special case of personalized toll pricing where no one is eligible to receive a discount.

1.2.1 The Managed Lane Network

The toll pricing system is developed for a corridor of freeway that consists of toll managed lanes (ML) and general-purpose lanes (GPL) in parallel to each other. The toll rate is adjusted in real time and is displayed on a sign before travelers have to make the decision whether to take ML or not. The toll is collected through an ETC system, which relies on the gantries that record each vehicle using ML. Gantries rely on transponder readers that record vehicles equipped with transponders and use cameras to take photos for vehicles without a transponder. Sensors are mounted along both ML and GPL to detect traffic flow and speed in real-time. Figure 1.1 illustrates

an ML network in a simplified way. ML networks in the real world are usually more complex and may have multiple entry and exit ramps along the corridor on both ML and GPL. Some networks consist of multiple tolling segments. Traveling through each tolling segment requires a separate toll payment, and toll rates are adjusted separately for different tolling segments. Travelers may switch between ML and GPL at the beginning of each tolling segment.

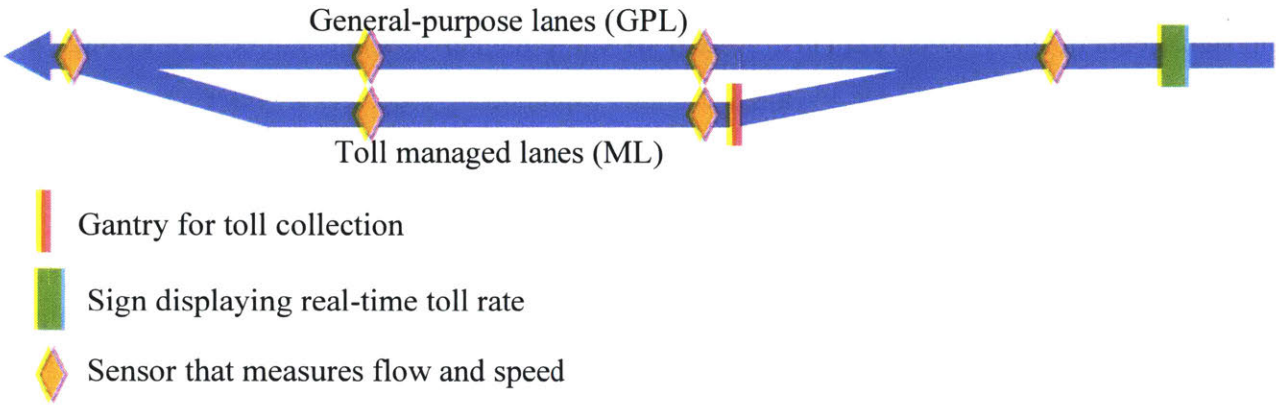

Figure 1.1: Illustration of a toll managed lane network

The above infrastructure is typically available in real-world ML networks. With flow and speed sensors providing measurements in real-time, the ML operators have been able to apply reactive toll control strategies so that toll rates are adjusted in reaction to observed traffic conditions. Relying on the same infrastructure, we develop a pricing system where toll rates are determined proactively based on predicted traffic conditions.

Besides flow and speed sensors, some ML networks are equipped with AVI sensors. Transponder users would generate a record each time when passing through an AVI sensor. The record contains a timestamp and a location. Such data, along with gantry records and flow and speed data, may be used to generate trip data for transponder users. Trip data contain travel time, duration, entry and exit points to the network, the choice between ML and GPL, traffic conditions along the trip, as well as other information. With a large number of trip data available, travelers' choice behaviors could be modeled with a personalized choice model. The personalized choice model enables us to extend the pricing system to a personalized toll pricing system, which allows discounts to be offered to some travelers. However, to implement the new system in the real world, additional infrastructure is required.

As we develop a real-time personalized toll pricing system, toll discount is optimized in real time based on predicted traffic conditions in the near future. Therefore, the discount will only be valid in the near future, because the optimal discount may be different at a later time. Therefore, a traveler would need to inform the system before arriving at the network in order to receive a discount for each trip. This communication is referred to as a signal from the traveler to the system. The signal contains traveler's estimated time of arrival to the network and can be sent through a mobile app, by text message, or by a system that automatically detects traveler's movement towards the network. Besides, a discount is only meaningful if the traveler receives it and takes it into consideration when making the choice decision. Therefore, two-way communication between the traveler and the system is essential. In addition, due to legal considerations, is may be necessary to obtain consent from the traveler before his/her personal trip data are used to generate the personalized discounts and the discount offers are sent to him/her.

Therefore, we design the infrastructure in a way that travelers need to sign up for the discount program, and they need to enable two-way communications with the system in order to receive discounts. Before arriving at the corridor, travelers send a signal to the system indicating his/her arrival time, and then they become eligible for a discount in this trip. Such travelers are referred to as eligible travelers, while the others are ineligible travelers who need to pay at the full displayed toll rate if choosing ML. Upon receiving the signal, the system informs the traveler of his/her personalized price factor (i.e., one minus discount rate). The price factor could be 1, which indicates that a discount is not available. The traveler then decides whether to take ML or GPL based on the discount.

Therefore, in order to accommodate the proposed personalized toll pricing system, additional infrastructure needs to be available, which may include an app and a background system to handle communications between travelers and the pricing system. This research does not focus on the details of the infrastructure system, and the rest of this thesis is based on the proposed infrastructure being available.

1.2.2 Proposed Real-time Personalized Toll Pricing System

Figure 1.2 illustrates the real-time personalized toll pricing system we propose, referencing the framework in Azevedo et al. (2018). The system optimizes the displayed toll rate that is effective for all travelers and generates discount offers for eligible travelers. An additional infrastructure system needs to be developed to allow for two-way communications between eligible travelers and the toll pricing system, which is not a focus of this research.

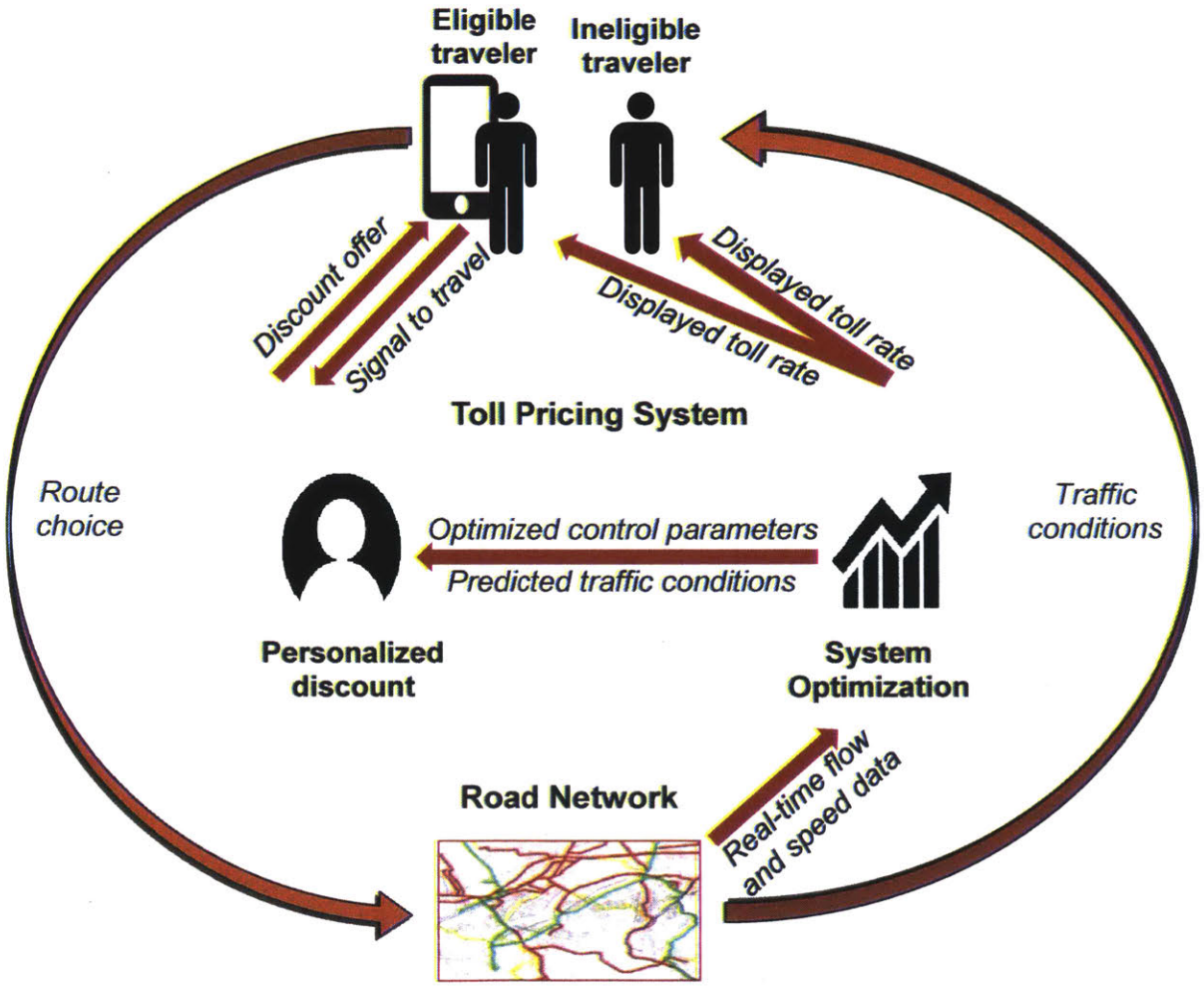

Figure 1.2: The proposed real-time personalized toll pricing system

The toll pricing system consists of a system optimization module and a personalized discount module. System optimization module solves a simulation-based optimization problem to find the toll rate and discount control parameter that maximize toll revenue. We design a personalized discount policy and formulate a personalized discount function based on the policy. The function generates a personalized discount for each eligible traveler, taking into consideration that traveler's preferences to ML and sensitivities to toll rate and travel time. While the personalized pricing function specifies how the discount is calculated, the actual amount of discount is also affected by the control parameters, which is optimized at the system level. Control parameters are arguments to the personalized discount function that are optimized in the system level and consists of the displayed toll rate and a discount control parameter.

Traffic dynamics are complex. While many researchers apply closed-form models to represent traffic dynamics in a usually simplified manner, we apply a DTA system to model the complex interactions between demand and supply. The system optimization module relies on a DTA system to predict traffic and evaluate the objective function for the simulation-based optimization problem. The DTA system is calibrated online towards real-time speed and flow data to ensure it accurately reproduces and predicts traffic dynamics on the real network.

The system optimization module is integrated with a DTA system and runs in rolling horizons. For each rolling time period, it generates the displayed toll rate and discount control parameter for the optimization horizon. The length of the rolling time period (e.g., 5 minutes) is the same as the length applied in the DTA system. The length of the optimization horizon (e.g., the next 15 minutes) is the same as the DTA system's prediction horizon.

The personalized discount module operates in real time. It is triggered each time when receiving a signal from an eligible traveler, and generate a discount offer for that traveler. As the module does not involve simulation, the eligible traveler is able to receive a discount immediately after sending a signal.

The toll pricing system relies on a choice modeling system not illustrated in Figure 1.2. We develop a logit model for travelers' choice behavior. Systematic heterogeneity among travelers is modeled

by incorporating characteristics and personal history variables. We estimate the personalized choice model with real trip data, and the model parameters may be updated online as new trip data becomes available. The personalized choice model enables the toll pricing system to predict different travelers' probabilities to choose ML given toll price and traffic conditions. The choice modeling system estimates and updates the personalized choice model using disaggregate trip data from transponder users. It maintains a database of travelers' preferences to ML and sensitivities to toll and traffic and updates these parameters with new trip data. Whenever a signal is received, the personalized discount module obtains the eligible traveler's preferences and sensitivities from the choice modeling system in order to calculate what discount to offer.

The choice modeling system also estimates the distributions of personalized preferences and sensitivities among the population and provide the distributions to the DTA system at each rolling horizon time interval (e.g., 5 minutes) so that the DTA can accurately simulate travelers' choice behaviors. The population is defined as the set of all potential travelers on the ML network, including those who choose GPL.

As disaggregate trip data are not commonly available, developing and estimating a personalized choice model is not always practical and thus personalized toll pricing is not always possible. Besides, the infrastructure for personalized toll pricing is not yet available in the real world, so we first implement the non-personalized toll pricing system and evaluate its performance with a case study using a real network and real data from that network. The non-personalized toll pricing system is a simplified version of the personalized system, as shown in Figure 1.3. As all travelers are ineligible for discounts, the personalized discount module is removed from the system. Evaluation of the system is performed in close-loop, by using a microscopic simulate the mimic the real world and evaluate toll revenue on the simulator.

After that, we extend the system to allow for personalized pricing and evaluate the added benefit based on optimization results from the system optimization module.

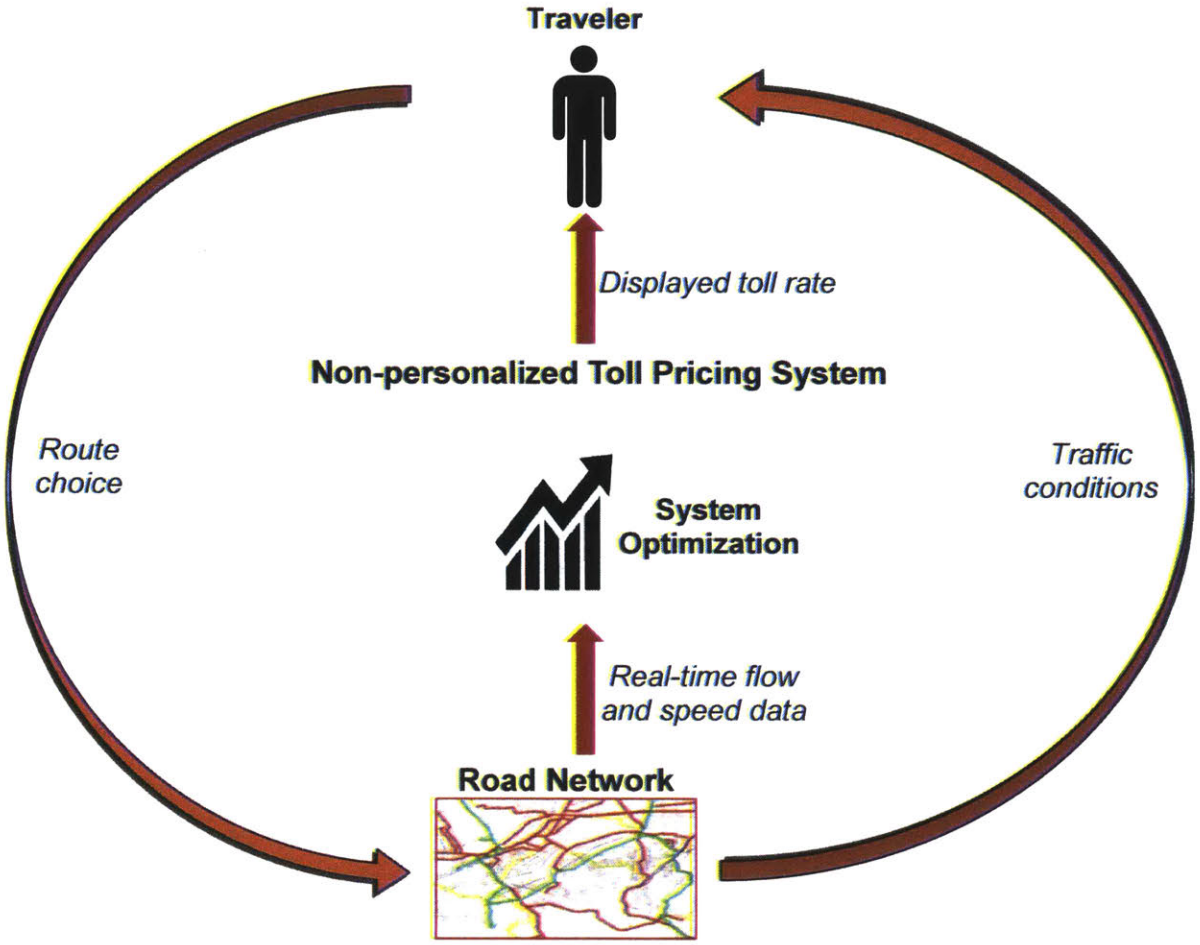

Figure 1.3: The non-personalized toll pricing system

1.3 Summary of Findings

The DTA-based toll pricing system is tested for real-time performance, and it is found to be feasible for real-time deployment. We find that the system's revenue maximization performance relies on the DTA system's prediction capability and that the proposed online calibration methodology is effective in improving the DTA system's estimation and prediction accuracy. The personalized choice model we develop is effective in capturing heterogeneities among individuals.

The personalized and non-personalized toll pricing systems we develop and implement are tested under different scenarios. They show potential in boosting revenue if applied in a real-world ML network. The personalized toll optimization problem we formulate does not significantly increase the dimension of the decision variables compared to non-personalized toll optimization. Given our objective of revenue maximization, personalized toll pricing is more effective in improving the objective.

1.4 Thesis Outline

The thesis consists of the following chapters. Chapter 1 gives an introduction, including the proposed framework and summary of findings. Chapter 2 reviews the literature on DTA systems, toll pricing, discrete choice models, and personalized pricing. Chapter 3 gives methodologies and a case study for non-personalized toll optimization and includes details of online calibration. Chapter 4 presents methodologies and a case study for personalized toll optimization and includes details of the choice model. Chapter 5 is for final conclusions.

Chapter 2 Literature Review

In this chapter, we review literature on four research topics: (1) dynamic traffic assignment (DTA) system and calibration algorithms developed to calibrate its demand and supply parameters, (2) road pricing schemes and toll managed lanes, (3) discrete choice models and modeling of heterogeneities, with a focus on applications on route choice and toll managed lanes, and (4) price discrimination and personalized pricing and incentives in transportation. Finally, we summarize the research gap.

2.1 Calibration of the Dynamic Traffic Assignment System

A DTA system models the interactions between traffic demand and supply, through which it assigns time-dependent traffic to road networks and outputs transportation performance measurements. It is applicable for real-time or near real-time traffic simulation and prediction. Typically, demand parameters include origin-destination (OD) demand, travel behavior, such as route choice, lane changing (gap acceptance, merging, yielding, and look-ahead), car following (acceleration, braking, and desired speed), and compliance (response to arterial signals, ramp meters, and toll plazas). Supply models primarily include speed-density and speed-flow relationships (i.e., the fundamental equation). DTA systems can be classified as macroscopic, mesoscopic, and microscopic depending on the fidelity of the supply and demand models. Usually, the supply model of a macroscopic DTA system uses the speed-flow relationship to replicate traffic dynamics, while that of a microscopic system simulates individual driver's driving behavior. The fineness of the mesoscopic system is in-between and may use different models depending on traffic conditions. Peeta and Ziliaskopoulos (2001) systematically summarize the foundations of DTA.

2.1.1 Offline Calibration

Calibration can be formulated as an optimization problem minimizing the deviation between the simulated and actual measurements. Earlier literature calibrates different groups of parameters in a separate manner. There is abundant research on time-dependent OD flow estimation (e.g., Ashok and Ben-Akiva, 2000; Hazelton, 2000; Cascetta and Postorino, 2001; Toledo et al., 2015), among which the Generalized Least Squares (GLS) algorithm is most often applied; Kim and Rilett (2004) developed genetic algorithms (GA) to calibrate driving behavior parameters. We review the calibration of route choice models in detail in Section 2.3. On the supply side, parameters of the fundamental equation are calibrated (e.g., Leclercq, 2005; Kundé, 2002; Park et al., 2006). More recent literature calibrates different groups of parameters jointly. The joint calibration is first accomplished iteratively (Darda, 2002; Toledo et al., 2004; Balakrishma et al., 2005).

Simultaneous calibration is proposed later to capture the interactions among different models. Balakrishma (2006) and Balakrishma et al. (2007) show that Simultaneous Perturbation Stochastic Approximation (SPSA) (Spall, 1992) can be applied for calibrating demand and supply parameters simultaneously and performs best in fitness. Later, various studies demonstrate the performance of SPSA (e.g., Ma et al., 2007; Vaze et al., 2009). However, in later studies (Paz et al., 2012; Lu, 2013), SPSA is found to be limited in terms of convergence rate and long-run accuracy, when applied to large-scale and noisy problems without analytical representations and with correlated parameters and measurements (Lu et al., 2015; Cipriani et al., 2011; Cantelmo et al., 2014). Balakrishna and Koutsopoulos (2008) extend SPSA by incorporating transition equations for the calibration of OD flows. Cipriani et al. (2011) adopt asymmetric OD matrix estimation and polynomial interpolation to the step size, the performance of which is further investigated by Cantelmo et al. (2014). To simultaneously calibrate demand and supply parameters, Lu et al. (2015) incorporate a weight matrix, which represents the correlation structure. The Weighted SPSA (W-SPSA) algorithm proposed improves convergence rate and accuracy when the dimensionality of parameters is large. Antoniou et al. (2015) propose the full framework for W-SPSA, including a series of methods to formally generate the weight matrix, as well as heuristic and combined approaches. Zhang (2017) propose further improvements in weight matrix generation and parameter update methods. Tympakianaki et al. (2015) propose c-SPSA for OD calibration, which enhances SPSA by simultaneously perturbing the gradient of a small number of constructed

homogeneous clusters each time. c-SPSA is proven to be more robust and converges faster than SPSA. Oh et al. (2019) treat the demand as a discrete variable instead of a continuous one, which is more realistic, and apply a Weighted Discrete SPSA (W-DSPSA) algorithm for GLS minimization.

Recent research further improves computational efficiency through embedding analytical problem-specific information within a simulation-based optimization (SO) algorithm (Osorio, 2019). In this line of research, Osorio (2019) formulates the offline OD calibration problem for large-scale simulation-based network models as an SO problem and propose a scalable and efficient metamodel SO algorithm. This research extends the previous work (Zhang and Osorio, 2017; Osorio, 2017) that considers only a single time interval.

2.1.2 Online Calibration

Offline calibration provides a priori values of the parameters which are then calibrated online. Online calibration is vital to ensure the accuracy in the estimation and prediction of traffic conditions, which is one of the most significant barriers for deploying large-scale DTA systems (Ben-Akiva et al., 2015). Yin and Lou (2009) propose a calibration framework that jointly and iteratively calibrates OD, behavioral, and supply parameters in a mesoscopic DTA system. The OD demand is calibrated by the GLS algorithm, while behavior and supply parameters are estimated with specific empirical methods. In a traffic management context, Hashemi and Abdelghany (2015) propose online calibration using GLS for OD and empirical methods for supply parameters. These methods are then applied to a corridor in Dallas (Hashemi and Abdelghany, 2016). Hashemi and Abdelghany (2016) use a DTA system to mimic the real-world and generate control strategies for travel time reduction with a meta-heuristic search algorithm. Their DTA system predicts significant time savings under the optimal strategies, but the actual impacts of such strategies are not tested in an environment different from the DTA system itself.

Recent advancements in online calibration methods also include simultaneous calibration of all parameters with a unified model. Antoniou et al. (2007) propose the Extended Kalman Filter (EKF) for the online calibration of DynaMIT. DynaMIT (Ben-Akiva et al., 2010) is a mesoscopic DTA

system that is capable of reading sensor data to estimate current traffic and generating control strategies (e.g., toll rates) based on predictions of future traffic. The EKF algorithm represents the evolvement of the network state by formulating it as a non-linear state-space model. It is capable of simultaneously calibrating all parameters to match all measurements, including speed, by linearizing the relations between measurements and parameters.

The online calibration of large-scale transportation networks requires specific techniques to reduce the calibration dimensionality and computational effort. Frederix et al. (2014) spatially decompose the OD of the entire network in a hierarchical approach and calibrate OD matrices of subnetworks using different methods. Djukic et al. (2012) and Djukic (2014) apply a principal component approach (PCA) in online OD calibration using a Kalman filtering approach. PCA effectively reduces the computational effort of large-scale online calibration by calibrating smaller-dimensional principal components of parameters. Prakash et al. (2017) apply a PCA in the online calibration of both supply and demand parameters using GLS. Although the estimation accuracy is slightly lower compared to regular GLS, the prediction accuracy improves because PCA captures inherent correlations of the parameters and filters the noise. Prakash et al. (2018) then apply PCA to the online calibration of both supply and demand parameters using EKF. Similar results are obtained, demonstrating the capability of PCA in reducing dimensionality for large-scale, simultaneous online calibration problems. Zhang et al. (2017) address the calibration of a large OD demand matrix by embedding an analytical model that relates calibrated parameters to the objective function. This methodology is demonstrated with the Berlin metropolitan area network. Zhang (2018) propose the augmented State Space Model (SSM) to capture delayed measurements, and a gradient estimation procedure called partitioned simultaneous perturbation (PSP) to reduce computations. Using these methods, the online calibration of the Singapore expressway network demonstrates a 90% reduction in the number of computations compared with the traditional finite difference method and the same calibration accuracy. Qurashi et al. (2019) combine SPSA with PCA to form a new algorithm named PC-SPSA, which is found to statistically reduce dimensionality and non–linearity. The performance of PC–SPSA is demonstrated in an offline network. However, the authors claim that it can be potentially applied for online calibration for networks of medium size and complexity.

2.2 Road Pricing and Toll Managed Lanes

Congestion pricing, as a tool for traffic demand management, has been a research interest since the late 1990s. It can be classified based on the zone of influence such as pricing at a cordoned area, on urban roads or specific lanes. Our particular interest is in road pricing, especially pricing on toll managed lanes, as roading pricing is predominantly applied on toll managed lanes in the US. Theories of road pricing have been developed, and real-world applications are worldwide (e.g., Meyer, 1999; Tsekeris and Voß, 2009; Li and Hensher, 2010; de Palma and Lindsey, 2011; Gu et al., 2018). Wood et al. (2016) review applications of managed lane pricing in Texas, the state where many managed lane networks have been operated. Tolls are applied using two strategies: static and dynamic. In a static setting, the toll either stays flat or varies in a predetermined way during the day, which is also referred to as time-of-day (TOD) tolling. Static pricing does not capture the interaction between real-time traffic conditions and travelers' willingness to pay. Conversely, in dynamic pricing, the toll rates are updated dynamically based on demand predictions and real-time traffic conditions. We refer to Chung and Recker (2011) for a review of state-of-the-art dynamic pricing in the US. We discuss two different dynamic pricing schemes in the remainder of this section.

2.2.1 Reactive Toll Pricing Schemes

As examples of dynamic pricing applications, we refer to the I-15 managed lane in San Diego and I-394 high-occupancy toll (HOT) managed lane in Minnesota (Yin and Lou, 2009). The base price for I-15 varies from \$0.50 to \$4.00 depending on the time of day, and the price can be adjusted in real-time in response to the traffic conditions. Performance evaluations of these examples show that dynamic pricing works well in practice (Supernak et al., 2003; Cambridge Systematics, 2006). Recently, Ferrovial/CINTRA applied dynamic pricing on a managed lanes network in Texas, where toll rates are adjusted based on a look-up table with offline calibrated parameters.

Optimization of dynamic toll rates for managed lanes is a flourishing research area. Nagae and Akamatsu (2006) present a case where a toll road manager can switch between two levels of the toll in order to maximize the revenue based on long-term dynamics of the transportation demand. Xu (2009) proposes a simulation-based dynamic pricing model that minimizes the total travel time

on the network predicted offline using DynaMIT, incorporating route choice and departure time choice models. Yin and Lou (2009) propose two dynamic pricing approaches for toll managed lanes based on real-time traffic conditions, which are referred to as feedback-control approach and reactive self-learning approach. The objective is to improve the free-flow travel service on the toll managed lanes while maximizing the throughput of the general-purpose lane. Morgul (2010) proposes a dynamic pricing model reactive to real-time traffic predictions by traffic simulation software Paramics and TransModeler, where a logit model represents travelers' response to toll rate and travel time information. Michalaka et al. (2011) develop a scenario-based toll optimization model that accounts for the stochasticity of traffic demand with a two-stage formulation. The targeted inflows are determined in the first stage, with an objective of maximizing the throughput of the general-purpose lane while achieving a desired density on the toll lane. Then the toll rate is determined in the second stage to achieve the targeted inflows. Jang et al. (2014) develop a reactive toll pricing system that determines toll rates based on predictions of traffic system performance measures, which are obtained from a closed-form model. Chen et al. (2014) propose a family of surrogate-based models to simplify the dynamic toll optimization problem. The surrogate models are implemented in a DTA system – DynusT and tested in a corridor in Maryland for optimizing objectives, including travel time, throughput, and revenue.

2.2.2 Proactive Toll Pricing Schemes

While earlier dynamic tolling strategies are often reactive and based on simplified supply-demand interactions, recent research develops proactive tolling strategies, where pricing is based on forward-looking, predicted traffic conditions. Using a DTA system named DYNASMART, Dong et al. (2011) proactively optimize toll rates to match the desired network conditions and conclude the benefits of such a proactive control strategy. Gupta et al. (2016) use a GA to optimize toll rates based on traffic predictions. They implement the algorithm in DynaMIT and optimize 13 toll rates in the Singapore expressway network. To the best of our knowledge, proactive pricing schemes are relatively less studied and have not been applied in real-world road networks.

2.3 Discrete Choice Models

Ben-Akiva and Lerman (1985) systematically review the theory of discrete choice models and application to travel demand modeling. A discrete choice model depicts an individual's (i.e., decision maker's) preference using a parameterized utility function which includes observed independent variables and unknown error terms. It assumes the decision is made based on utility maximization. There are many advancements in discrete choice models, among which are logit mixture models (Ben-Akiva et al., 2001) that incorporates random heterogeneity as well as hybrid choice models (Ben-Akiva et al., 2002) that can incorporate flexible random heterogeneity, flexible substitution patterns, and latent variables.

2.3.1 Modeling of Systematic and Random Heterogeneities

Individuals' preferences are complex, dynamic, context-dependent, heterogeneous, and even contradictory (Castells et al., 2015). There are two types of heterogeneity, i.e., systematic and random (unobserved). While the systematic heterogeneity can be captured by incorporating individuals' characteristics into the systematic component of the utility function, the random heterogeneity is captured in the error component. There are two types of random heterogeneity among the population: inter-personal and intra-personal. Inter-personal heterogeneity accounts for preference variations among individuals, while intra-personal heterogeneity accounts for preference variations of the same individual among different choice scenarios, which addresses the fact that individuals do not have permanent preferences (Ben-Akiva et al., 2019). In the presence of multiple observations per individual, logit mixture models can be developed to capture both inter- and intra-personal heterogeneity (Hess and Rose, 2009; Hess and Train, 2011; Yáñez et al., 2011). However, this type of model is computationally burdensome (Hess and Train, 2011), because it does not have a closed-form likelihood function as a flat logit model does and thus requires simulation. To address the computational issue, Becker et al. (2017) introduce a Hierarchical Bayes (HB) estimator by extending the Allenby-Train procedure (Train, 2009). Danaf et al. (2017) further incorporate Gibbs sampling to allow updating individuals' preferences in real-time.

2.3.2 Discrete Choice Models for Route Choice

The route choice model is a core behavior model in a DTA system. Ramming (2002) systematically summaries route choice models. C-logit and path-size logit models correct for the issue of overlapping paths. Cross-nested logit, probit, and hybrid choice models are also discussed and compared. With growing data availability, route choice models that capture systematic heterogeneity have been developed (Kurri et al., 2000; Danielis et al., 2005; Fowkes and Whiteing, 2006). Route choice models that capture random heterogeneity are also available. Latent class models identify the impact of truck size (Feng et al., 2013) or travel distance (Rowell et al., 2014). Using random coefficient models, Kawamura (2000) and Toledo et al. (2013) estimate the distribution of the value of time (VOT) and Toledo et al. (2018) estimate distributions of both VOT and value of reliability (VOR). Li et al. (2016) incorporate the OD pair and choice-situation-specific random error components.

2.3.3 Discrete Choice Models for Managed Lanes

In the context of toll managed lanes, Burris and Brady (2018) point out that travel time saving and toll rate are not the only factors determining the choice between managed lanes and general-purpose lanes. In this sense, a good representation of travelers' behaviors incorporating preference heterogeneity is vital for welfare analysis and real-time operations. Among such discrete choice models, inter-personal preference heterogeneity is widely considered (e.g., Small et al., 2005; Toledo and Sharif, 2019), which is crucial for welfare analysis (Small et al., 2005) and provides a basis for individual-level parameter inference (Train, 2009). Intra-personal preference heterogeneity is considered by Börjesson et al. (2013) and Hess and Rose (2009) in the context of route choice, capturing systematic and random variations, respectively. However, the two studies are both based on stated preference (SP) data reflecting preferences in hypothetical scenarios, which is likely to be biased against real choice scenarios (Ben-Akiva et al., 2019). The collection of reveal preference (RP) data traditionally relies on travel surveys (e.g., Brownstone et al., 2003; Börjesson et al., 2007), and it is usually time- and money-consuming. In recent years, with the emergence of ETC facilities, massive RP trip data become available. Exploratory studies using RP data to explain managed lane references have been conducted (e.g., Burris and Brady, 2018; Burris and Ashraf, 2019), yet econometric models have not been developed in these studies. In general,

there is limited research on modeling managed lane preferences using RP data and traveler's characteristics, due to the scarcity and proprietary nature of the required data. Bhat and Castelar (2002) propose a unified framework for combining RP and SP data in estimating logit mixture models that consider inter-personal heterogeneity. A most recent study by Xie et al. (2019) takes advantage of massive RP trip records and travelers' characteristics obtained from a managed lane operator. They develop a discrete choice model capturing both inter- and intra-personal heterogeneity in managed lane preference and use HB for efficient model estimation.

2.4 Personalization in Transportation and Price Discrimination

Discrete choice models that represent individual preferences provide grounds for personalized policy designs. In the context of tolling, the concept of personalization can be realized via price discrimination, which means offering personalized pricing, discounting, or incentives for potential toll road users based on understandings for traffic conditions and individual user's behaviors. Under this context, personalization can be applied for various purposes, e.g., toll revenue improvement, congestion reduction, and social welfare improvement.

2.4.1 Price Discrimination and Personalized Pricing

Pigou (2017) defines the three levels of price discrimination: first-degree (i.e., personalized pricing), second-degree (i.e., product versioning), and third-degree price discrimination (i.e., group pricing). Murthi and Sarkar (2003) and Fudenberg and Villas-Boas (2006) extensively review the literature on behavior-based price personalization. Specific forms of personalization have been considered. For example, Kuo et al. consider negotiation in a dynamic pricing model (2011), and Netessine et al. (2006) and Aydin and Ziya (2008) consider cross-selling where price depends on the customer's past purchases. Personalized pricing is the most aggressive strategy as it charges the equilibrium price to each customer for revenue maximization. If sufficient individual-specific data is available, a personalized pricing strategy is practical and significantly improves revenue (Shiller, 2013). Shriller (2013) demonstrates a personalized pricing strategy based on a demand model that resembles an ordered Probit. This pioneering work improves the revenue of personalized pricing by modeling customer behaviors, rather than using willingness to pay (WTP)

or customer demographics. Elmachtoub et al. (2018) provide mathematical proofs to evaluate the value of personalized pricing under different market and information scenarios.

Early literature on personalized pricing often assumes static and unlimited inventory. However, inventory (i.e., supply-side) considerations are important for perishable products, e.g., seats of an aircraft, traffic capacity at a specific time. Revenue management (RM) is the pricing and inventory control of perishable products, which can be considered as an inventory evaluation technique. In the context when inventory is limited and fixed, pricing is based on not only WTP but also expected future revenue, which is estimated by dynamic programming (DP) (Talluri and Van Ryzin, 2004). Talluri and Van Ryzin (2004) propose optimal pricing strategies under a general choice model for demand. Chapuis (2007) provides a comprehensive review of DP techniques for RM. More recent literature refines the demand model representing customer's responsiveness to price. Aviv and Pazgal (2005) investigate dynamic pricing under a highly uncertain demand that is modeled as a function of the price. Aviv and Pazgal (2008) later investigate the case when customers act strategically. The impact of price on demand in future periods are also investigated (Popescu and Wu, 2004; Ahn et al., 2007). Chen and Wu (2018) formulate a Bayesian DP that solves the dynamic pricing problem with limited inventory and an unknown (but observable) distribution of customer WTP. Chen and Chen (2015) provide a review of dynamic pricing with a brief mention of personalized pricing developments.

Aydin and Ziya (2009) first incorporate personalized and dynamic pricing of limited inventory. Customer WTP distributions are estimated using the signal (e.g., demographic information) provided by the arriving customer, and personalized price depends on the signal, inventory level, and time. Suh (2010) investigates dynamic pricing and an extension to personalized pricing in the case where a retailer sells limited inventories of substitutable products. Chen et al. (2015) develop a modeling framework for RM problems that allows customized pricing based on a logit model for purchasing probability as well as personalized assortment optimization.

2.4.2 Personalized Pricing/Discounting/Incentives in Transportation

There is limited literature on personalized pricing in the field of transportation. One particular application is airlines, which may offer personalized offers to passengers. Wittman and Belobaba (2017) propose two heuristics that provide fare offers to specific passengers based on estimated WTP. Dynamic discounting that provides targeted discounts to leisure passengers can increase the yields and load factors. Wittman and Belobaba (2018) propose a heuristic based on estimated WTP for customized dynamic pricing of airfares using observations of passenger characteristics. The results show that this pricing strategy increases revenue by 3-4% in the case study of one airline.

We find several other studies using personalized incentives instead of personalized pricing to encourage altering travel decisions. Sun et al. (2018) propose dynamic and personalized incentives that influence travelers' travel decisions. They use a logit model that represents the traveler's route choice given incentives. They use analytical models to represent traffic dynamics and propose a mathematical model to minimize the total expected travel time under the budget constraint. Social welfare (i.e., travel time savings) is demonstrated using a toy network. Azevedo et al. (2018) and Araldo et al. (2019) present Tripod and its System Optimization (SO) framework, based on the concept of a flexible mobility demand system proposed by Atasoy et al. (2015). Tripod is a smartphone-based system that offers information and personalized incentives in order to influence travelers' real-time travel decisions. Before starting trips, travelers can access the personalized menu in the Tripod smartphone application, which offers several travel options, including departure time, mode, and route. The traveler is offered tokens (i.e., incentives) associated with each travel option, the amount of which corresponds to system-wide energy savings. The SO framework is an integrated bi-level transportation management system. The SO computes the best token amounts for travel options to maximize energy savings under token budget constraints. The SO is multimodal, dynamic, predictive, and personalized, meaning that the token amounts are dependent on predicted traffic network conditions as well as individual traveler's profile. SimMobility (Adnan et al., 2016) is the simulation tool that houses the SO, and the traveler behavior modeling framework consists of a series of discrete choice models that determine the menus presented to travelers (Xie et al., 2019; Song et al., 2018). Zhu et al. (2019) present a personalized system that has similar functionalities and objectives as Tripod. The system is different from Tripod in the definition and realization of personalization. The amount of

personalized incentives is determined to ensure the probability of accepting to be greater than a predefined threshold, and individual preferences are mined by a particle filter approach. The authors mention that a future research direction can be combining the SO framework in Tripod with the proposed personalized incentivizing concept.

2.5 Summary

We review literature in four areas related to our proposed research. We summarize the research gaps below.

(1) Proactive toll pricing on ML

Dynamic pricing schemes have been widely studied and applied in the real world. The dynamic pricing concept is evolving from reactive to proactive, often based on simplified representations of demand and supply. While there is limited literature on real-time toll pricing based on DTA systems, where comprehensive demand-supply interactions are replicated in simulation, this research area is still scarce and often has a lack of proper validation.

(2) Online calibration of simulation-based DTA

The online calibration of a DTA system is vital for the accurate representation and prediction of the traffic network. Simultaneous online calibration of OD demand, behavior parameters, and supply parameters take the correlations among the parameters into consideration. Sensor counts are the most widely used measurements, yet research rarely shows how accurate the simulated speed is compared to actual data.

(3) Route choice models for managed lanes

Discrete choice models can capture both systematic and random heterogeneity, the latter including inter- and intra-individual heterogeneity. In the context of toll managed lanes, while a simple choice model that consists of time and money is most commonly used, the literature finds it insufficient to model the complicated and heterogeneous behaviors on ML route choice. A more comprehensive model specification is needed, which captures the heterogeneity among individuals.

There is limited literature on managed lane discrete choice models estimated using large-scale real trip record data, and nearly no literature incorporates travelers' history data in modeling.

(4) Personalized road pricing

There is extensive research on dynamic and personalized pricing. On the supply side, inventory limitation is studied in recent research. On the demand side, consumer behaviors are being increasingly incorporated in demand modeling, yet the interactions between supply and demand are often simplified. There are limited applications of personalized pricing or incentives in transportation. To the best of our knowledge, there is no literature in the context of road pricing despite its potential in further improving traffic management. Optimization models for personalized road pricing are not available. While the elasticity-based personalized pricing strategy and evaluation of inventory value are widely used in other fields, for road traffic, it is more difficult to derive because dynamic demand-supply interactions are complex and difficult to model analytically. Equity and other practical concerns need to be incorporated into personalized road pricing.

Chapter 3 Real-time Toll Optimization

In this chapter, we develop a toll pricing system that makes proactive decisions on toll rates by solving an optimization problem at each rolling period. We apply the system in the context of a ML network from the viewpoint of the ML operator with an objective to maximize revenue while offering a premium level of service. We formulate the toll optimization problem to maximize revenue subject to network conditions and tolling regulations. Toll pricing relies on a DTA system to predict traffic conditions, and we develop an enhanced online calibration methodology for the DTA system to improve its estimation and prediction capabilities. The toll pricing system is tested with data from a real toll managed lanes (ML) network in a closed-loop evaluation framework.

3.1 Introduction

We formulate an optimization problem to maximize revenue. Since the toll rate is not personalized, decision variables are the displayed toll rates at different times and different locations of the network. We apply a DTA system to predict traffic in order to evaluate revenue under candidate toll rates. The toll pricing system is fully integrated with the DTA system.

The DTA system runs in rolling time horizons, which means, it models the real-time progression of traffic interval by interval. Traffic flow and speed are measured for each time interval and represent the average flow/speed within each interval. Therefore, the length of an interval should be long enough so that flow and speed can be effectively measured, and random fluctuations in both the real world and the simulators can be partially averaged out. On the other hand, an interval should be short enough so that real-time traffic dynamics could be captured, and traffic conditions would not change tremendously within an interval. Previous studies related to DTA systems often set interval length to 5 minutes.

Under the rolling horizon framework, traffic conditions are considered constant within any given interval and thus optimized toll rates would be constant in a given interval. Therefore, the real-time toll pricing system is also designed to run in rolling time horizons with the same interval length as the DTA system. The system generates toll rates for each interval of the prediction horizon and maximizes expected revenue for the entire prediction horizon, with the DTA system predicting traffic conditions for the prediction horizon. Coincidentally, some ML networks in the real world adjust toll rates every 5 minutes. Therefore, in our case study, we set the interval length to be 5 minutes for the DTA system and the toll pricing system, which is consistent with previous researches on DTA as well as industry practice of toll pricing. We set the length of the prediction horizon to be 15 minutes.

Effective toll optimization relies on accurate predictions of future traffic. Before making predictions, the DTA system needs to firstly understand current traffic conditions and estimate demand and supply parameters, and this process is called state estimation. Online calibration is essential for accurate state estimation, as it calibrates demand and supply parameters to match real-time observations of traffic flow and speed. Following that, the DTA system predicts demand and supply parameters and performs traffic simulation to predicted flow, speed, travel time, and other traffic condition metrics, and calculate expected toll revenue for the prediction horizon.

We propose and apply an online calibration methodology that consists of OD calibration and supply calibration. OD stands for origin-destination demand and is a parameter measuring how many travelers are to travel between each origin-destination pair at each time interval. We apply the generalized least squares (GLS) method for calibration of OD demand to match traffic flow measurements. Supply calibration refers to the calibration of supply model parameters, where the supply model in a macroscopic DTA system means the speed-density relationship. In this study, we apply a mesoscopic DTA system where the supply model is more comprehensive and contains additional parameters than the speed-density function parameters. We develop a heuristic for calibration of supply parameters to match the speed and density measurements, where density is calculated from flow and speed. The performance of the online calibration methodology is tested by calibrating a mesoscopic DTA system online towards real data from an ML network.

Toll optimization is tested with a closed-loop evaluation framework, so the platform optimizing the toll rates is different from the platform evaluating the system. A microscopic simulator is used as the second simulation platform to evaluate the toll rates by calculating the simulated revenue, and it takes the place of the real world in this closed-loop framework. Under the closed-loop framework, the DTA system is calibrated online towards sensor measurements provided by the microscopic simulator.

The credibility of closed-loop evaluation can only be confirmed when the demand, supply, and travel behaviors are represented accurately in the microscopic simulation platform so that it mimics the real world. In this thesis, we briefly discuss the offline calibration of the microscopic simulator. It is calibrated to real data before the deployment of closed-loop evaluation.

In the case study, we test toll optimization under a based scenario and multiple experimental scenarios. We evaluate the toll rates and simulated revenue as evaluated by the microscopic simulator between the proposed toll pricing system and a base toll rate. The based toll rate is a toll profile that varies at different times of the day but is not affected by real-time fluctuation of demand or traffic conditions.

Methodology and the case study in this chapter have been presented in Zhang et al. (2019).

3.2 Optimization Formulation

For each rolling time interval, toll optimization is formulated as follows.

$$\max_{\boldsymbol{P^D}} z = \sum_{t=1}^{m} \sum_{k=1}^{g} P_{tk}^{D} Q_{tk}$$

s.t.

$$(\boldsymbol{Q}, \boldsymbol{V}) = \mathrm{dta}(\boldsymbol{P^D}, \boldsymbol{D}, \boldsymbol{H}, \boldsymbol{b})$$

$$\boldsymbol{L P_t^D} \leq \boldsymbol{p^{constraint}} \; \forall t$$

$$Q_{tk} \leq q_k^{UB} \; \forall t \; \forall k$$

$$V_{tk} \geq v_k^{LB} \; \forall t \; \forall k$$

(3.1)

where

z is the objective, i.e., the sum of expected revenue over all gantries and all intervals

t is the index for time intervals in the prediction horizon

m is the number of intervals in the prediction horizon

k is the index for gantries in the network

g is the number of gantries in the network

P_{tk}^{D} is the displayed toll rate for toll gantry k in the t-th time interval

Q_{tk} and V_{tk} are the expected flow and speed at gantry k in interval t

$\boldsymbol{D}$ and $\boldsymbol{H}$ are the predicted OD demand and supply parameters in the prediction horizon

$\boldsymbol{b}$ is the choice model parameters

$\boldsymbol{L}$ is a transformation matrix

$\boldsymbol{p^{constraint}}$ is the right-hand-side of toll constraints

q_k^{UB} and v_k^{LB} are upper/lower bound for flow/speed at gantry k

Regulations by the government or tolling policies by the operator may require ML flow to be subject to an upper bound q_k^{UB} at each gantry k, and ML speed to be subject to a lower bound v_k^{LB} at each gantry k. To accommodate these requirements while ensuring there is always a feasible solution to the optimization problem, we incorporate these considerations as penalty terms in the objective function, where W_{tk}^{Q} specifies the dollar amount of the penalty if ML flow at gantry k in interval t exceeding upper bound by one unit, and W_{tk}^{V} is the penalty for ML speed at gantry k in interval t exceeding lower bound by one unit. Penalty coefficients $\boldsymbol{W^Q}$ and $\boldsymbol{W^V}$, the flow upper bound $\boldsymbol{q^{UB}}$, and the speed lower bound $\boldsymbol{v^{LB}}$ are pre-determined parameters.

The first constraint specifies that the expected ML flow $\boldsymbol{Q}$ and expected ML speed $\boldsymbol{V}$ are obtained through the traffic prediction module of a DTA system. The next sections present details of how demand and supply parameters are predicted, and details about the choice model.

The second constraint is a generic representation of any linear constraints imposed on toll rates. $\boldsymbol{L}$ is a matrix that converts toll vector $\boldsymbol{P_t^D}$ into a vector of metrics on which constraints are imposed.

Those metrics are constrained to be no larger than $\boldsymbol{p}^{\mathbf{constraint}}$. Constraints are imposed for each time interval independently. These constraints may include but are not limited to the following.

(1) Toll rate P_{tk}^{D} does not exceed an upper bound p_k^{UB} for all gantry k.

(2) The network may consist of multiple tolling segments, and travelers traveling through the full corridor may switch between ML and GPL at the beginning of each tolling segment. If each tolling segment is of similar length, toll rates on different tolling segments should be similar, which makes sure that travelers are not surprised by a much higher or lower toll rate when finishing one tolling segment and continuing to the next segment. Therefore, the difference between toll rates for different segments may be constrained to be no more than a certain value, for example, 1 dollar.

(3) There may be multiple entry ramps to ML in the middle of a tolling segment. Travelers using these ramps do not travel through the whole segment and may be charged at a lower toll rate based on where they enter. Therefore, a gantry on a ramp in the middle of a tolling segment may charge at a fraction of the full segment toll rate, where the fraction is pre-determined based on the location of the ramp.

It should be noted that traffic simulations in the DTA system are stochastic, and thus, the flow, speed, and resulting revenue are also stochastic. In this thesis, stochasticity of the DTA system is not taken into consideration, so that we formulate the problem with respect to expected flow, speed, and revenue. In the case study, toll optimization relies on only one simulation realization to evaluate the objective, and thus, the resulting solution is generally not optimal due to simulation stochasticity.

3.3 DTA-based Solution Framework

In the case of non-personalized tolling, the pricing system consists of only one module called system optimization. System optimization relies on a DTA system to provide traffic prediction under different candidate toll rates, and it is fully integrated with the DTA system. Figure 3.1 illustrates system optimization. The state estimation module performs traffic simulation for the current time interval (denoted as estimation horizon) and estimates the current demand and supply parameters. Traffic simulation is performed through interactions of the demand and supply models. The prediction-based toll optimization module involves iterations of the traffic prediction module

and the optimization module. In each iteration, the optimization module proposes candidate toll rates. Based on that, as well as the predicted demand and supply parameters, the traffic prediction module simulates traffic and evaluates the objective function value for the prediction horizon. This process iterates until a solution is found. The toll rates are sent to the network for implementation, and also serve as input for system optimization to be performed in the next rolling period.

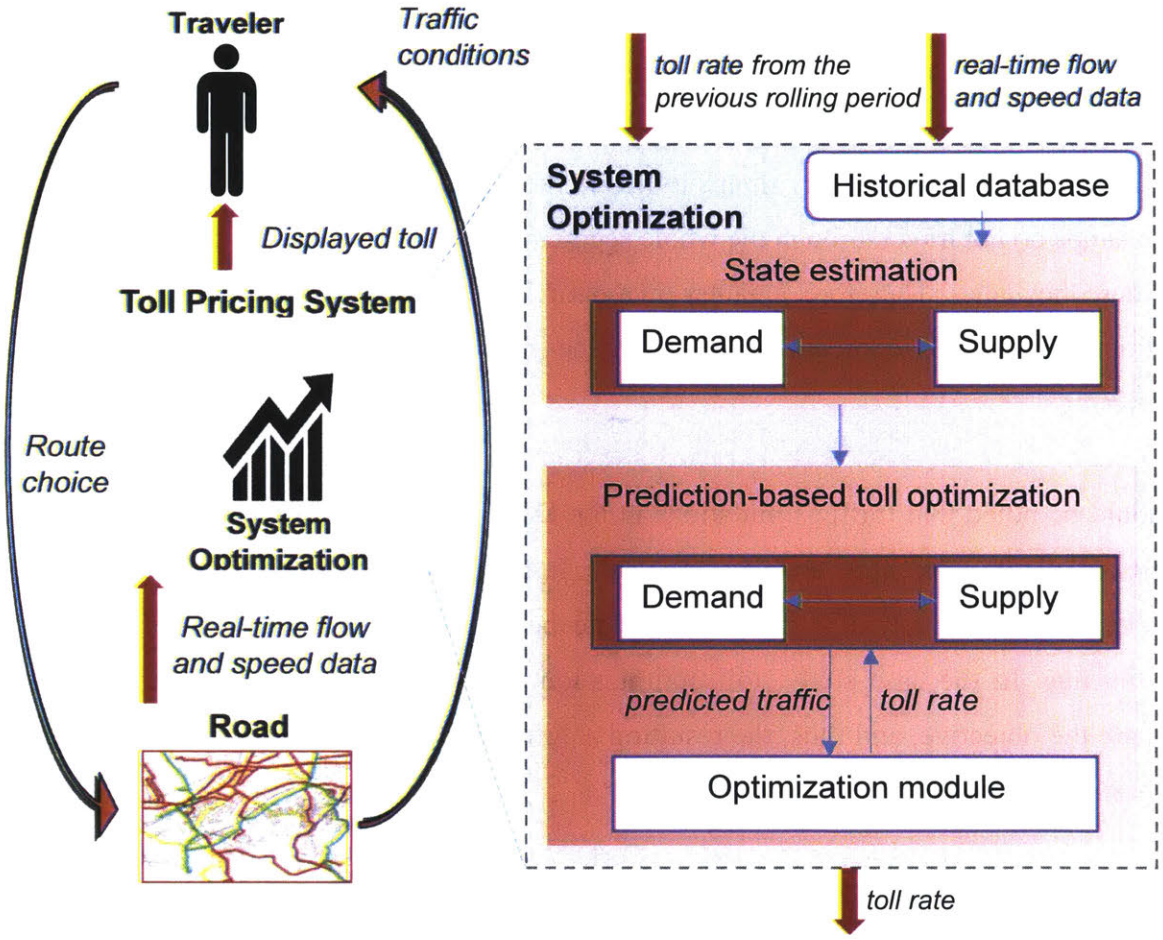

Figure 3.1: Overview of the real-time non-personalized toll pricing system

In the case study of this research, we apply a grid search algorithm that searches all nearby values of the decision variables, relative to a starting point of last time interval's toll rates. We run the system starting from a half-hour before our intended simulation starting time so that the first half-hour is considered a warm-up period and is excluded from result analysis. The initial values of the decision variables are set empirically, and their impact on simulation results is reduced as a result of the warm-up.

We apply DynaMIT for the toll pricing system. DynaMIT (DYnamic Network Assignment for the Management of Information to Travelers) (Ben-Akiva et al., 2010) is a simulation-based mesoscopic DTA system developed in the Intelligent Transportation Systems (ITS) Lab of Massachusetts Institute of Technology (MIT).

The demand model of DynaMIT predicts travelers' route choice using a multinomial path-size logit model and assigns OD demand to path flows. For each origin and the destination pair, DynaMIT identifies all possible paths, denoted as choice set CS. As the choice model is not personalized, the index for individuals is omitted. The probability of choosing path j at time interval t is given by the following equation:

$$\pi_{tj} = \frac{e^{v_{tj}+\ln PS_j}}{\sum_{r\in CS} e^{v_{tr}+\ln PS_r}} \tag{3.2}$$

where

PS_j is the path size variable for path j, which specifies the path's degree of overlapping with other paths;

v_{tj} is the systematic utility of path j at time interval t;

CS is the choice set.

While a more comprehensive personalized choice model is to be applied for personalized toll pricing in the next chapter, we specify a simple utility function as follows for non-personalized toll pricing, and calibrate the parameters offline to match real aggregate choice probabilities that are calculated from sensor flow data.

$$v_{tj} = b^{\mu}(b_{j,TOD}^{ASC} - s_{tj}^{time} - \frac{s_{tj}^{toll}}{b_{TOD}^{VOT}}) \tag{3.3}$$

where

b^{μ} is the scaling parameter, where a smaller value represents higher randomness in choice decisions;

$b_{j,TOD}^{ASC}$ is the alternative specific constant (ASC) for path j representing any other factors affecting travelers' preferences to this path. Its subscript TOD indicates the ASC could be different at different times of the day;

s_{tj}^{time} is the expected travel time on path j at time interval t;

s_{tj}^{toll} is the toll on path j at time interval t such that $s_{tj}^{toll} = \sum_{k \in Gantries_j} P_{tk}^{D}$ where $Gantries_j$ is the set of all gantries on path j;

b_{TOD}^{VOT} is travelers' value of time, representing the rate of substitution between toll and travel time. It is assumed to be a random parameter subject to a log-normal distribution so that different travelers have different values of time. Its subscript TOD indicates the mean and standard deviations of the log-normal distribution could be different at different times of the day.

As for the supply model, DynaMIT applies a modified Greenshields model to simulate traffic speed. However, for a road segment or a portion of a road segment where a queue is present, DynaMIT applies a deterministic queuing model. More details are given in Section 3.4.4.

3.4 Calibration and Prediction Models for the DTA System

Effective toll optimization relies on the DTA system's capability to predict traffic conditions under candidate toll rates. Prediction accuracy depends on state estimation performances. Offline and online calibration is essential to ensure an accurate estimation of the current network state.

3.4.1 State Estimation

A state is a vector consisting of demand and supply parameters. State estimation is the real-time process of incorporating an initial state, historical data, and real-time traffic sensor data to achieve a more reliable estimation of the current state.

Offline calibration provides a priori values of the parameters which are then calibrated online. We rely on iterative proportional fitting (IPF) algorithm to obtain a historical time-dependent OD demand table based on historical sensor flow measurements. We calibrate choice model parameters empirically so that simulated choice ratios match actual data. For supply parameters, DynaMIT applies a closed-form model, which is described in the next section, and we estimate the model parameters with actual sensor data.

To perform online calibration, the generalized least squares (GLS) method is used to estimate OD demand from real-time sensor flow measurements. For supply parameters, we propose a heuristic online calibration method to adjust supply model parameters in real-time, and we find simulation results match sensor data with satisfactory accuracy in terms of speed measurements, including when congestion is present.

3.4.2 Prediction Model

Prediction module predicts future traffic based on the current state, taking into consideration any historical information, traffic control strategies (e.g., future toll rates) to be deployed and travelers' response to traffic conditions. To predict future demand and supply parameters, we formulate the parameter prediction model as an autoregressive process (Antoniou et al., 2007):

$$x_t^{pred} - x_t^{hist} = \sum_{i=1}^{n} f_i (x_{t-i}^{est} - x_{t-i}^{hist}) \tag{3.4}$$

where

x_t^{pred} is predicted parameter value for the current interval;

x_t^{hist} is the historical parameter value for the current interval;

n is the autoregressive degree;

f_i is the autoregressive coefficient for degree i;

x_{t-i}^{est} is estimated parameter value for the i-th interval ahead;

x_{t-i}^{hist} is historical parameter value for the i-th interval ahead.

For OD demand, we estimate n and f_i using offline-calibrated time-dependent OD parameters. For supply, since we do not obtain time-dependent supply parameters offline, the above autoregressive model is simplified as follows.

$$x_t^{pred} - x^{hist} = f \cdot \left(x_{t-1}^{est} - x^{hist}\right) \tag{3.5}$$

The coefficient f is empirically determined. We then use the predicted parameters x_t^{pred} as input to simulate traffic for the prediction horizon (e.g., next 15 minutes) and obtain predicted sensor measurements.

3.4.3 Evaluation Metric for Calibration and Prediction

To evaluate the calibration and prediction accuracies, we use a metric called RMSN (Root Mean Square Normalized error) to quantify the difference between actual and simulated measurements (Antoniou et al., 2007). RMSN is specified as follows.

$$RMSN = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_i^{est} - y_i^{true})^2} \Big/ \overline{y_i^{true}} \tag{3.6}$$

where

M is the number of measurements;

y_i^{est} is the estimated value of the i-th measurement;

y_i^{true} is the true value of the i-th measurement.

3.4.4 Online Calibration of Supply Model

The optimization module of this study relies heavily on accurate prediction of travelers' choice between ML and GPL, and travel time (or travel speed) is an important factor for their decisions. Therefore, it is essential to make sure the state estimation module is able to reveal the supply parameters accurately, and thus, the simulated travel speed accurately matches actual sensor speed measurements.

As DynaMIT is a mesoscopic DTA system, in its traffic simulation model, a road segment is represented with a queuing part (downstream) and a moving part (upstream) (Ben-Akiva et al., 2010). A queue will form if flow on the segment exceeds Segment Capacity, or queue in the downstream segment spills out. Traffic speed on the queuing part is subject to a deterministic queuing model. A moving part exists only if a queuing does not occupy the full segment or does not exist. Simulated traffic speed on the moving part is determined by a traffic dynamics model specified as follows.

$$\begin{aligned} v &= max(v_{min}, v_s) \\ v_s &= v_{max} && \text{when } k \le k_{min} \\ v_s &= v_{max}\left(1 - \left(\frac{k - k_{min}}{k_{jam}}\right)^{\beta}\right)^{\alpha} && \text{when } k > k_{min} \end{aligned} \tag{3.7}$$

k is the density, v is the speed, v_s is an intermediate variable, and the other 6 parameters (v_{min}, v_{max}, k_{min}, k_{jam}, α, β), as well as the Segment Capacity, are referred to as supply parameters. The above traffic dynamics model and the queuing model together are referred to as the supply model of DynaMIT.

For the 7 supply parameters of each road segment, we estimate their a priori values from speed and flow measurement data offline. When deploying real-time toll optimization, we adjust a selection of supply parameters online in reaction to real-time sensor measurements. Figure 3.2 illustrates these operations. Step [b] ~ [e] constitute the heuristic online calibration method for supply parameters. Among them, step [c] adjusts the speed-density relationship of the supply model to match incoming data. Step [d] is triggered only when simulation does not capture a sudden emergence of congestion, while [e] is triggered when the dissipation of congestion is not captured in simulation.

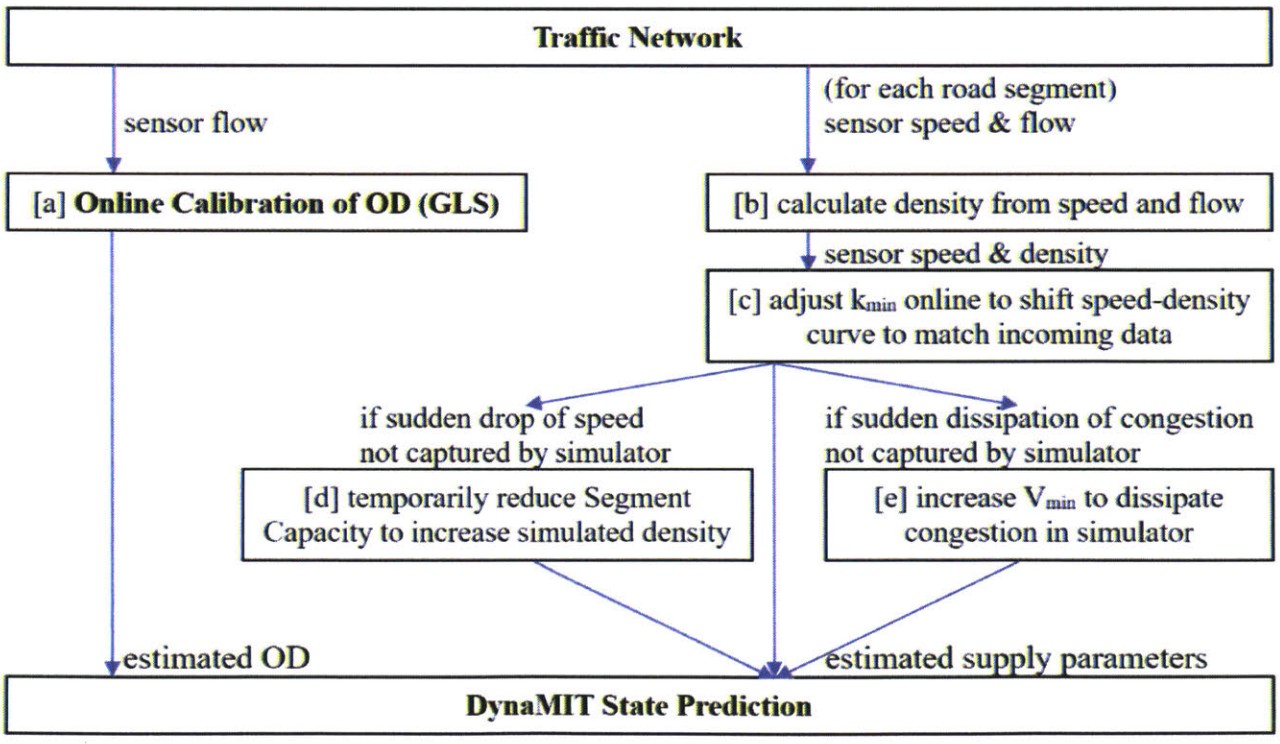

Figure 3.2: Calibration of DynaMIT

3.5 Closed-loop Evaluation Framework

Before the toll pricing system is implemented in the real world, the validity and performance of the developed models and algorithms need to be tested in a simulation environment. Therefore, a closed-loop evaluation framework is applied by using a microscopic simulator as a representation of the actual traffic network.

In this study, we use MITSIM as the testbed. MITSIM is a microscopic traffic simulator developed in the ITS Lab of MIT (Yang et al., 2000). It incorporates road topography, time-dependent OD demand, driving behavior (e.g., car following, lane changing) models and route choice models, simulates individual vehicle's movements and generates simulated sensor measurements.

Route choice is modeled as a path-size logit model, which takes into account the similarities between paths that are overlapping. Travelers make route choice decisions based on information on toll rates and travel times. To mimic the real-world, we assume that travelers have access to real-time traffic information, so they are aware of current toll rates and traffic conditions (i.e., travel time) on downstream links.

The optimized toll rates generated by the toll pricing system are implemented in MITSIM. DynaMIT is provided data from MITSIM sensors rather than a real-world traffic surveillance system. The closed-loop testing framework requires that the microscopic traffic simulator represents the real-world accurately, i.e., travelers in MITSIM behave similarly to those in the real-world, and demand-supply interactions occur in the same way. This is achieved by calibrating MITSIM towards real data.

The calibration of the microscopic traffic simulator relies on an enhanced weighted-SPSA algorithm (Zhang, 2017). Demand parameters and selected driving behavior model parameters, as well as choice model parameters, are calibrated simultaneously to minimize the discrepancies between simulated and actual sensor measurements.

3.6 Case Study

The toll pricing system is tested for a real-world ML network with real data from the network and evaluated in a closed-loop evaluation framework.

3.6.1 Network and Data

The toll pricing system is applied to the North Tarrant Express (NTE) network, a 13-mile corridor on US Interstate Highway I-820 and Texas State Highway TX-183. The corridor consists of ML, which is branded as TEXpress lanes, and GPL (Figure 3.3). The network is equipped with sensors that provide traffic flow and speed measurements, and toll gantries for non-stop tolling.

The operator of this corridor has provided us with samples of data collected on 9 Fridays in summer 2017, which includes sensor flow and speed measurements, toll rates, and some AVI data.

The tolls are applied to two tolling segments. Segment 1 is highlighted darker in the figure, and segment 2 has lighter color. Toll gantries are located at the beginning of each tolling segment, and entry ramps to ML in the middle of the tolling segments. A traveler pays a toll when entering ML. The toll rate is determined based on the entry point but not the exit point. If the traveler continues from Segment 2 to Segment 1 on westbound, he/she pays a second toll.

In this case study, we focus on the westbound (WB) of the network. For ease of analysis, the WB corridor is divided into 9 parts based on locations of entry and exit ramps on ML, as shown in the figure. Part 1~4 belong to tolling segment 2, and Part 5~9 belong to tolling segment 1.

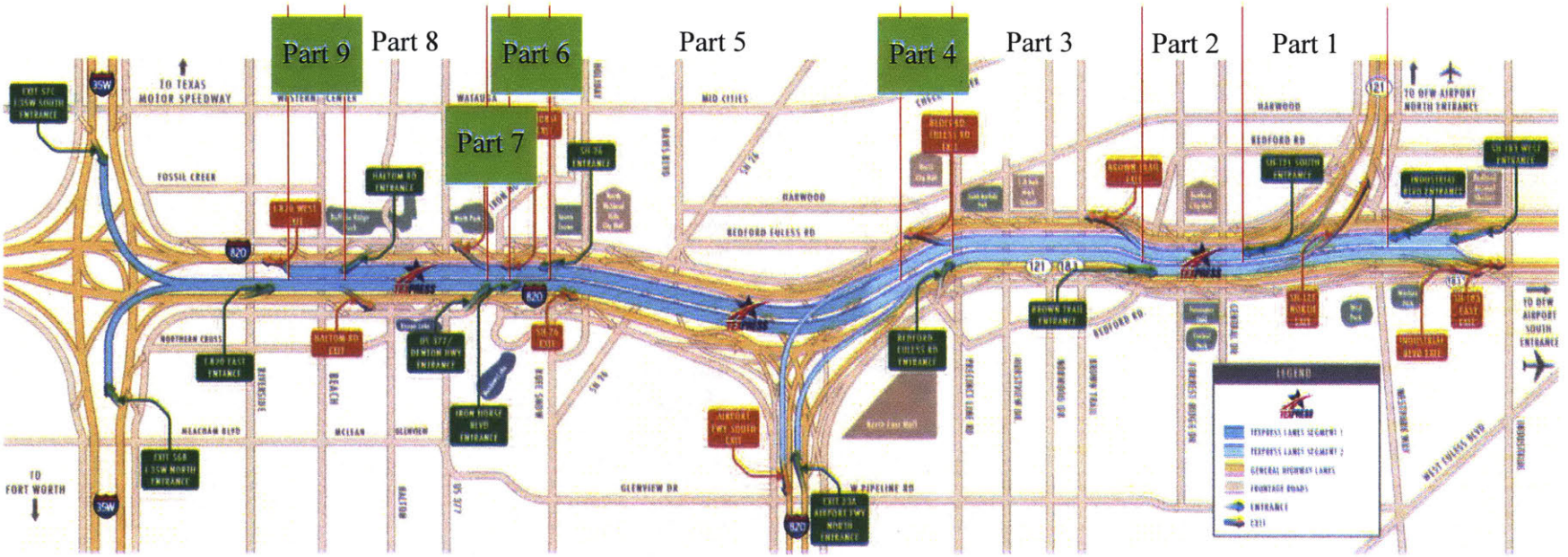

Figure 3.3: The North Tarrant Express network

3.6.2 Offline Calibration

The AVI data give an insight into the OD pattern, but they only include a fraction of vehicles. The data are used as seed OD for better offline calibration. We use the IPF algorithm to scale up the AVI-based OD, according to flow at origin and destination nodes. Flow data are available at most origin and destination nodes, either obtained from sensors on the corresponding origin and destination links or calculated from sensor flow on nearby links according to the flow conservation law. The IPF algorithm converges with no more than 0.1% error in terms of fitting origin or destination flow.

We assume that different travelers have a different value of time, which is subject to a log-normal distribution. The choice model is estimated empirically to make sure simulated choice ratios match actual data. For a successful calibration, we introduce and calibrate a constant term to capture some network-specific phenomena. We also allow the model parameters to be different in different periods, which includes morning (5:30-9:00), mid-day (9:00-14:00), afternoon (14:00-18:00) and evening (18:00-21:00). These periods are determined based on historical toll rates on the network.

We estimate supply parameters offline with Day 1 data. We firstly estimate a set of supply parameters for each type of road segments (ML, GPL, ramp). Using the results as starting values, we estimate supply parameters for each road segment. The statistics of estimated supply parameters are presented in Table 3-1.
Figure 3.4 shows the data points and the estimated supply curve for selected road segments.

Table 3-1: Statistics of offline estimated supply parameters

	Segment Type	v_{min} (mph)	v_{max} (mph)	k_{min} (veh/m)	k_{jam} (veh/m)	α	β	**Segment Capacity** (veh/s)
min	ML	8	57	0.005	0.08	2.4	1	0.56
	GPL		54	0.005				
	Ramp		40	0.002				
median	ML	8	72	0.009	0.09	3	1	1.11
	GPL		65	0.014	0.10	3		1.67
	Ramp		64	0.007	0.11	2.4		1.00
max	ML	40	76	0.012	0.16	3	1	1.70
	GPL	31	69	0.019	0.12			2.78
	Ramp	8	72	0.011	0.16			2.28

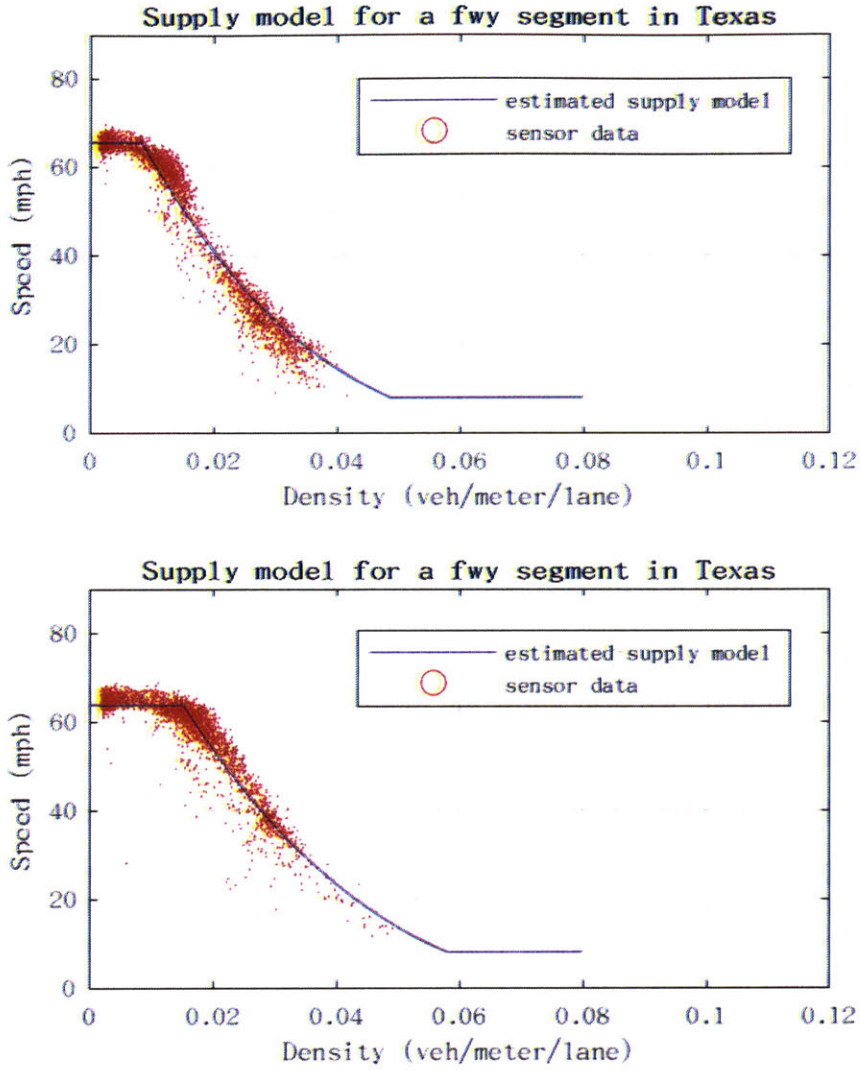

Figure 3.4: The data points and the estimated supply curve for selected road segments

After the offline calibration process, we obtain a set of parameters for Day 1, and the simulation results have an error of 19% in RMSN for flow measurements and 15% for speed measurements.

3.6.3 Online Calibration and Prediction

We calibrate DynaMIT offline to Day 1 data and obtain a set of parameters. Using Day 1 parameters as a priori values, we then calibrate DynaMIT online for the other 8 days.

For each 5-minute time interval, DynaMIT first runs traffic simulation once using predicted parameters from the last interval, obtains simulated measurements, and then applies demand and supply calibrations to obtain better estimations of demand and supply parameters. Finally, DynaMIT simulates traffic with calibrated parameters. The GLS algorithm works well for calibrating OD demand parameters, and the heuristic is effective in replicating real-world congestions in the simulator. Results of online calibration are shown in Table 3-2 as "OC demand&supply". The "No OC" case is a base case where historical OD and supply parameters are used in the simulation. The "OC demand only" case has OD calibrated by the GLS algorithm, but historical supply parameters are used in the simulation. At the bottom of the table are average RMSN among the 8 days, as well as standard deviation (SD) of the RMSN among the 8 days. The SD indicated that online calibration performs differently among the tests we performed. Such difference results from the difference in demand and supply conditions among the 8 days, as well as simulation stochasticity among the 8 tests.

We take Day 1 offline calibration results as a baseline. Simulations of other days have a much larger error for flow if online calibration is not performed because those days have different demand from Day 1. Error for speed is about the same because supply parameters are static in these cases and are similar in different days. The online calibration of demand significantly improves flow accuracy. The addition of supply calibration then improves speed accuracy due to its capability to adjust supply parameters dynamically. In all cases, prediction RMSNs are slightly larger than estimation, which is as expected and acceptable, because the prediction model incorporated additional errors while the autoregressive model is used to predict future demand and supply parameters.

Note that, the results presented above are based on a single test for each of the 3 scenarios on each of the 9 days. As traffic simulation by the DTA system involves stochasticity, the RMSN shown in the table may be different based on different realizations of the DTA simulation.

Table 3-2: Calibration and prediction accuracies

RMSN(%)		Estimation (0~5min ahead)		Prediction (0~15min later)					
				0~5min later		5~10min later		10~15min later	
		Flow	Speed	Flow	Speed	Flow	Speed	Flow	Speed
Day 1	Offline calibration results	19	15						
Day 2	No OC	22	16	22	15	22	15	22	15
	OC demand only	12	16	16	15	19	15	19	15
	OC demand&supply	12	13	17	11	19	12	22	12
Day 3	No OC	23	12	23	14	23	14	22	14
	OC demand only	12	12	16	14	18	14	19	14
	OC demand&supply	12	10	16	10	19	11	21	11
Day 4	No OC	23	13	23	15	23	15	23	15
	OC demand only	12	13	16	15	18	15	19	15
	OC demand&supply	13	11	17	11	19	12	22	12
Day 5	No OC	38	22	38	23	38	24	38	23
	OC demand only	16	23	23	24	25	24	26	24
	OC demand&supply	18	19	24	17	26	18	29	18
Day 6	No OC	33	17	33	14	33	14	33	14
	OC demand only	13	17	19	14	22	14	23	14
	OC demand&supply	15	15	21	10	23	11	25	12
Day 7	No OC	23	14	23	14	23	14	23	14
	OC demand only	12	14	16	14	18	14	19	14
	OC demand&supply	12	12	16	10	19	10	22	11
Day 8	No OC	23	12	24	13	24	13	24	13
	OC demand only	14	12	18	13	20	13	21	13
	OC demand&supply	14	10	19	10	21	10	23	10
Day 9	No OC	22	12	22	13	23	13	23	13
	OC demand only	11	12	16	13	18	13	19	13
	OC demand&supply	14	9	19	9	21	10	22	10
Average and SD (Day2~9)	No OC	25.9 ±6.1	14.8 ±3.5	26.0 ±6.0	15.1 ±3.3	26.1 ±6.0	15.3 ±3.6	26.0 ±6.0	15.1 ±3.3
	OC demand only	12.8 ±1.6	14.9 ±3.8	17.5 ±2.5	15.3 ±3.6	19.8 ±2.5	15.3 ±3.6	20.6 ±2.6	15.3 ±3.6
	OC demand&supply	13.8 ±2.1	12.4 ±3.3	18.6 ±2.8	11.0 ±2.5	20.9 ±2.5	11.8 ±2.7	23.3 ±2.6	12.0 ±2.6
Average improvement and SD (Day 2~9)	OC demand only over No OC	13.1 ±4.9	-0.1 ±0.4	8.5 ±3.7	-0.1 ±0.4	6.4 ±3.6	0.0 ±0.0	5.4 ±3.5	-0.1 ±0.4
	OC demand&supply over No OC	12.1 ±4.4	2.4 ±0.5	7.4 ±3.7	4.1 ±0.8	5.3 ±3.7	3.5 ±1.1	2.8 ±3.6	3.1 ±0.8

We present more detailed results for Day 6. Figure 3.5 shows the simulated flow and speed after the online calibration of demand and supply, compared with true measurements. Figure 3.6 shows the predicted flow and speed for the next 5 minutes based on OD and supply parameters obtained from online calibration. Each small plot shows average flow or speed on one of the nine parts of the GPL. We can see the proposed online calibration methods are successful in replicating flow and speed fluctuations in each part of the westbound GPL, despite in some cases simulated congestions are still not as severe as actual measurements. ML has overall less congestion, and their plots are omitted.

These results demonstrate that we are capable of understanding and predicting traffic conditions when congestions are present, the predictions used for toll optimization in the DTA system are accurate, and thus evaluation of the objective function is accurate. Therefore, the system is able to obtain informed decisions on toll rates.

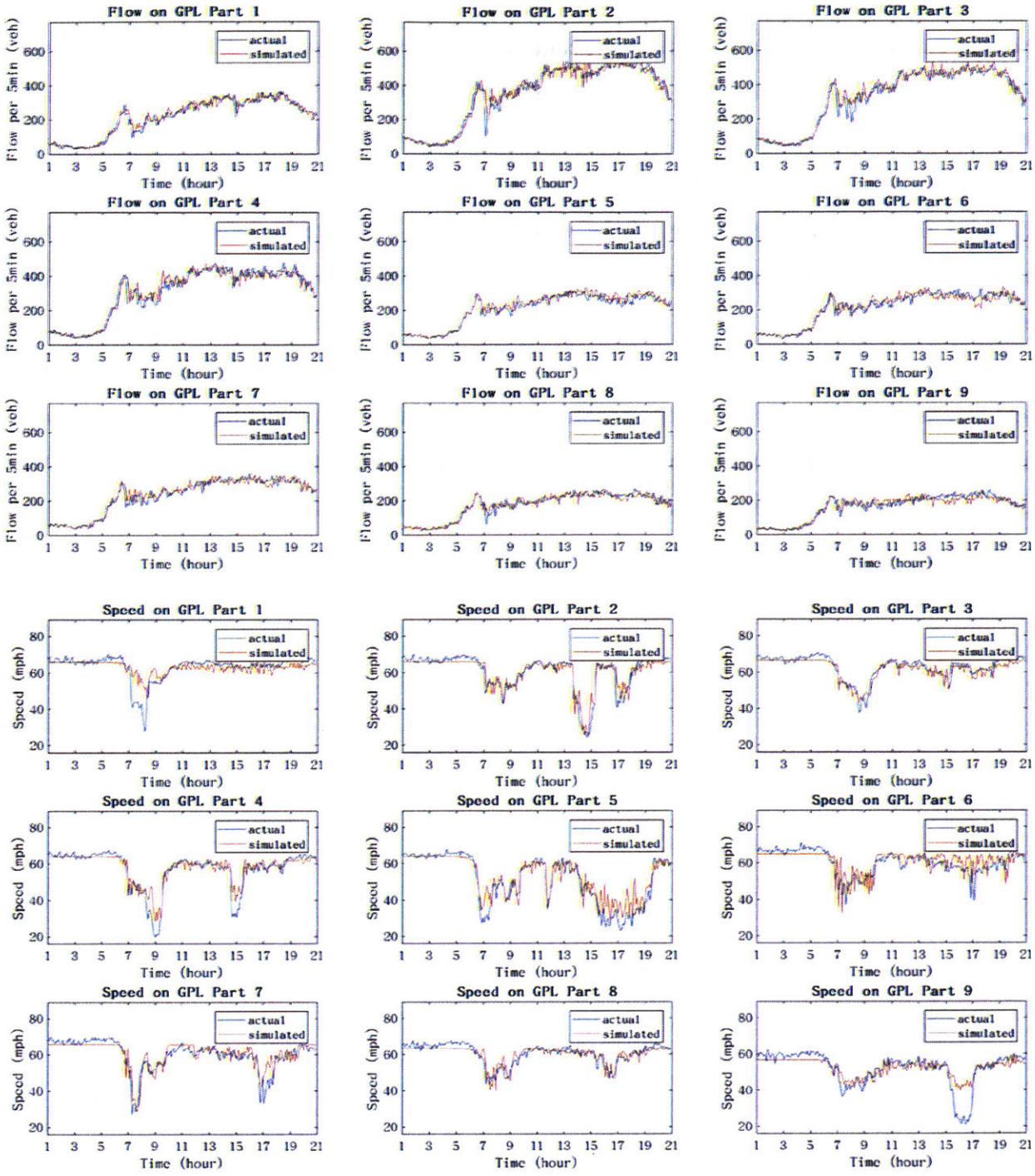

Figure 3.5: Calibration results on Day 6

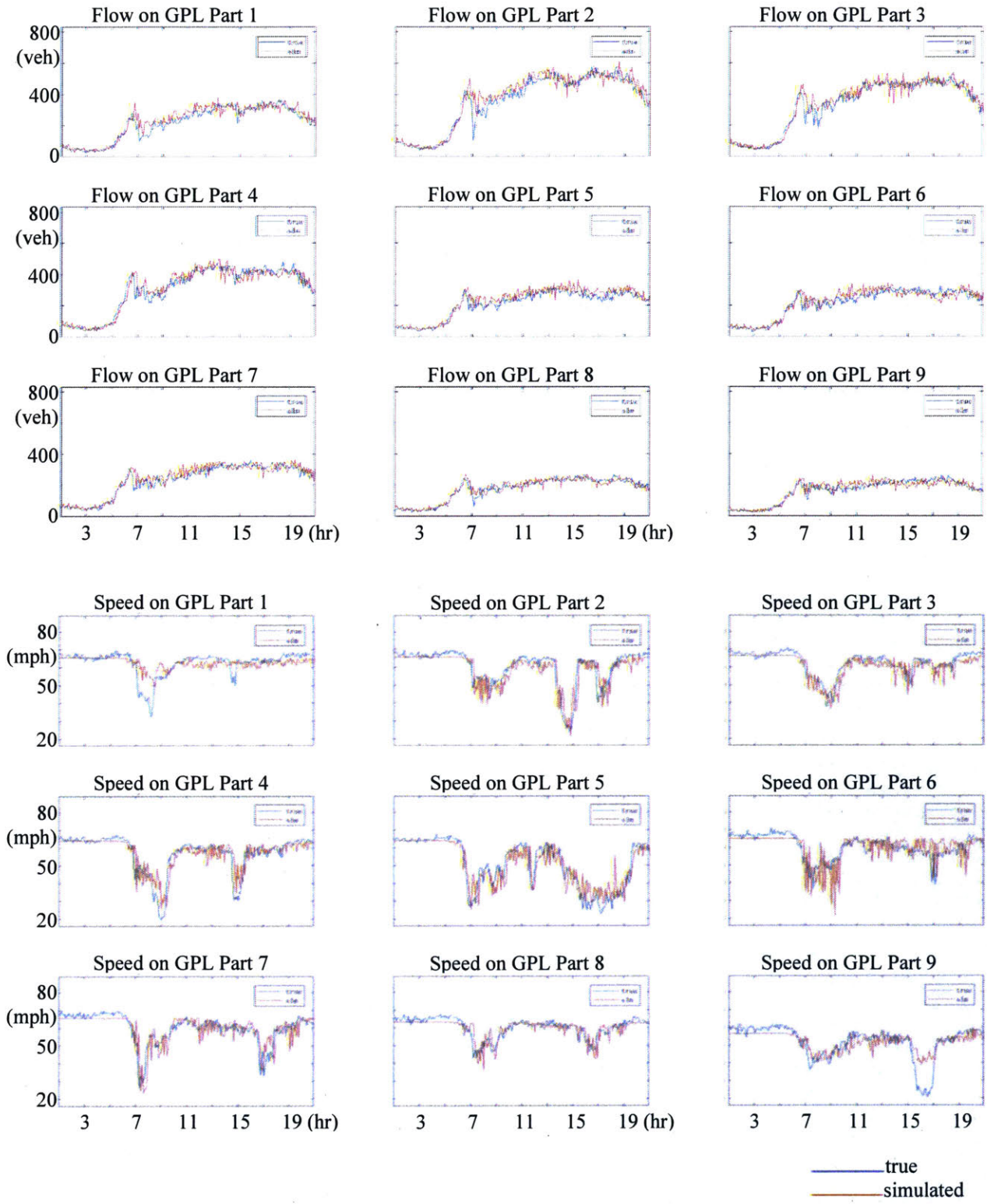

Figure 3.6: Traffic prediction results on Day 6

3.6.4 Toll Optimization

We evaluate the toll optimization framework in closed-loop for Day 6. We firstly calibrate MITSIM towards sensor measurements of Day 6. RMSN of the calibration result is 19% for flow and 17% for speed. We then apply the toll optimization framework and implement the optimized toll rates in MITSIM. We compare these toll rates with a base toll, which is obtained with the same toll pricing system, except that online calibration is not enabled. In such a situation, DynaMIT is fed with parameters that have been calibrated offline towards historical (Day 1) data, and the base toll rates represent a toll profile that is optimized for historical demand and traffic conditions but no adaptive to real-time data. Comparing the optimized toll with this base toll highlights the added benefit of online calibration in the prediction-based dynamic tolling.

There are 5 gantries on the westbound of the network. The toll optimization model generates toll rates for the 2 gantries located at the beginning of each tolling segment. The toll rate at each of the other 3 gantries is a pre-determined fraction of the gantry at the beginning of the corresponding tolling segment. Per tolling regulations, the toll rates may change dynamically every 5 minutes, and the amount of change is restricted to no more than \$0.5. Toll rates on tolling segments 1 and 2 are subject to an upper bound of \$5.3 and \$5.7, respectively. Besides, we impose a constraint that toll rates on the two tolling segments cannot be different by more than \$1, which is for practical considerations and is consistent with historical toll rate data from the network. For this study, we use a search algorithm that searches 3 toll values for each tolling segment, i.e., reduce by \$0.2, keep the same, or increase by \$0.2. The algorithm then evaluates the objective function by predicting toll revenue in the next 15 minutes.

Figure 3.7 shows the optimized toll rates for the full corridor compared to the base toll and per-5-minute revenue under these two toll profiles. Note that the revenues shown are calculated from simulation results by MITSIM, our testbed for evaluating the toll optimization framework. Figure 3.8 and Figure 3.9 show flow on ML and speed on GPL, comparing our simulation results under optimized toll rates and base toll rates.

Our optimization results suggest, in general, higher toll rates compared to the base toll except during PM peak when they both reach the upper bound, because online calibration successfully

captures most congestions, and travelers' route choice model in our system shows room for toll increase under congestion. According to our simulation of the 5:30-21:00 period in the closed-loop framework, revenue is 8.1% higher under optimized toll rates. Under the optimized toll, flow on ML is generally lower since toll rates are higher than the base toll, and thus flow on GPL gets higher and speed on GPL gets lower. However, on tolling segment 2 (Part 1~4) GPL becomes so congested after 17:00 that optimized toll rates maintain at high levels even after the PM peak period. Meanwhile, there is still a higher flow on ML at Part 1, which leads to much higher revenue during that period. Note that our framework is not addressing congestions on GPL. Based on our evaluation in the closed-loop framework, the above results demonstrate that the dynamic toll pricing framework with online calibration is promising in improving revenue while making GPL more congested.

Flow on GPL is not shown because it is complimentary to flow on ML. Speed on ML is not shown because ML is generally not congested. With optimized toll rates, speed on ML is maintained at a high level. Since we use different model parameter values in 4 periods of the day, there may be sudden changes in simulated flow between periods.

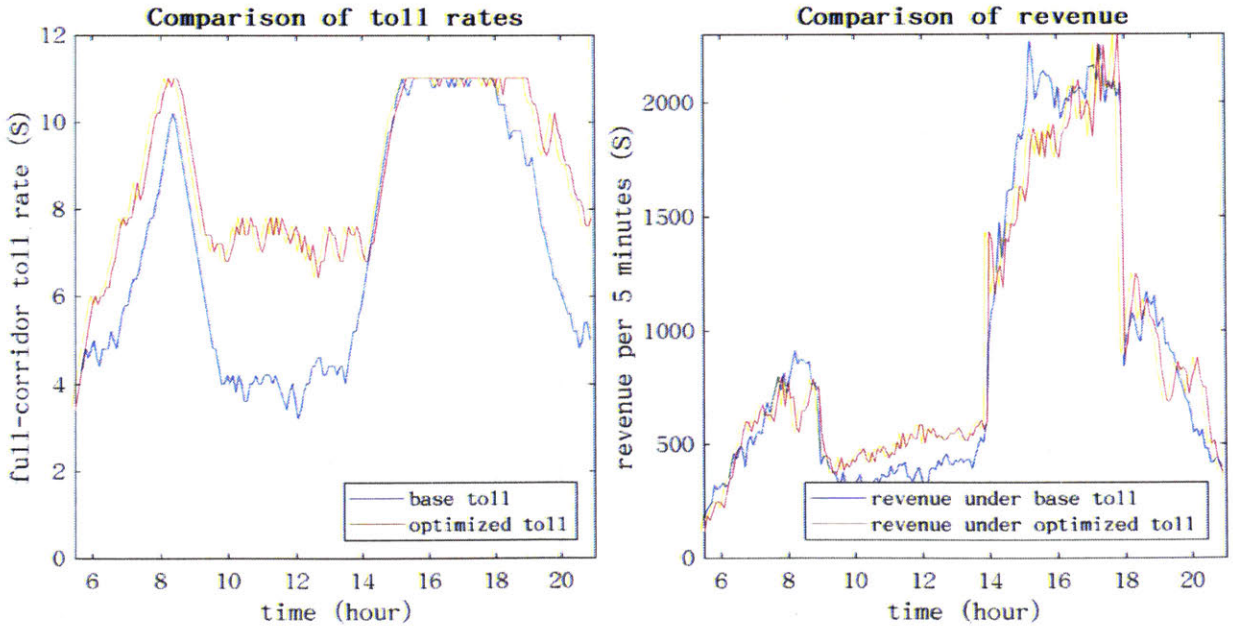

Figure 3.7: Comparison between optimized and base toll and corresponding revenue

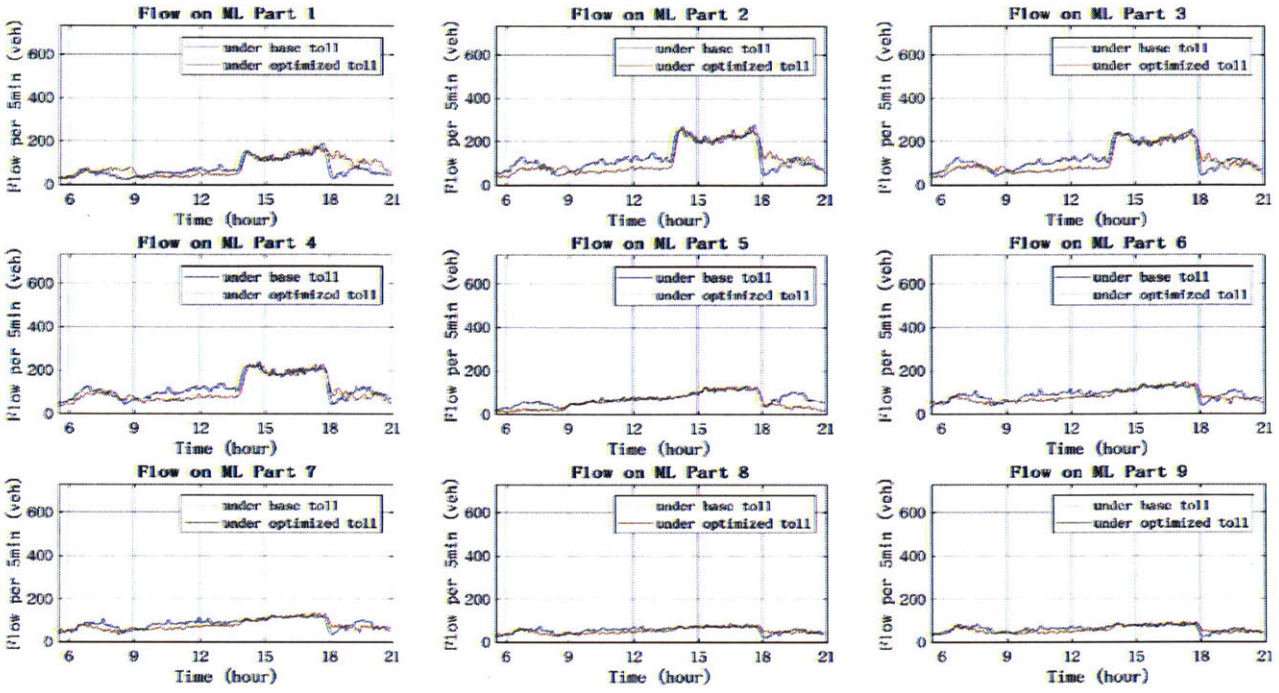

Figure 3.8: Flow on ML

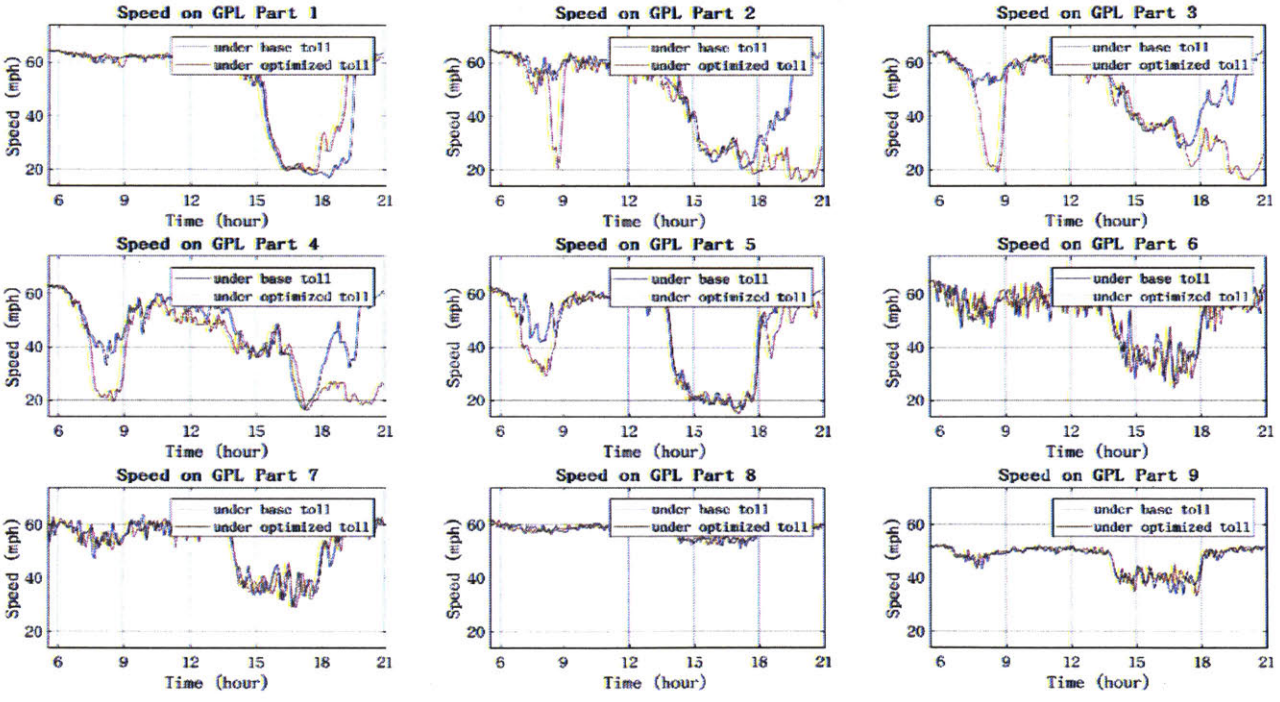

Figure 3.9: Speed on GPL

3.6.5 Toll Optimization under Multiple Scenarios

We further evaluate toll optimization under some experimental scenarios, which represent different regulatory, demand, or supply conditions (Table 3-3).

Table 3-3: Experimental scenarios

Test ID	Supply and demand condition	Tolling strategy
[1A]	[1] Base condition (Same as Day 6)	[A] Use the base toll
[1B]		[B] Optimize toll
[1C]		[C] Optimize toll without upper bound
[2A]	[2] Lower demand (Demand is 20% lower)	[A] Use the base toll
[2B]		[B] Optimize toll
[3A]	[3] Worse supply (In micro-simulator, vehicle deceleration rate is 50% lower)	[A] Use the base toll
[3B]		[B] Optimize toll

Condition [3] represents a day when road conditions are adverse so that drivers' braking behavior is more conservative. This is simulated by setting vehicle deceleration rates to 50% lower.

Optimized toll rates under the 3 supply/demand conditions are presented in Figure 3.10 in comparison to the base toll. Revenue in each test is evaluated in MITSIM and compared to the revenue under the same condition but applying different tolls. These experiments are performed for 5:30-18:00, a shorter period than the previous section. Note that the comparison between [1A] and [1B] has been presented in the previous section, for a longer simulation period. Table 3-4 shows the aggregate metrics that evaluate each test.

Table 3-4: Revenue improvement compared to the base toll under experimental scenarios

Test ID	Tolling strategy	Revenue (×$1000)	ML market Share (%)	Average toll cost Per ML trip ($)	Average travel time per trip (min)
Base condition					
[1A]	Base toll	133	14.6	4.18	7.17
[1B]	Optimized toll	138	12.8	5.04	7.42
[1C]	Optimized, no UB	146	10.5	6.77	10.54
Lower demand					
[2A]	Base toll	105	13.4	4.39	5.63
[2B]	Optimized toll	109	12.4	4.92	5.70
Worse supply					
[3A]	Base toll	131	14.5	4.35	10.65
[3B]	Optimized toll	144	13.2	5.59	13.78

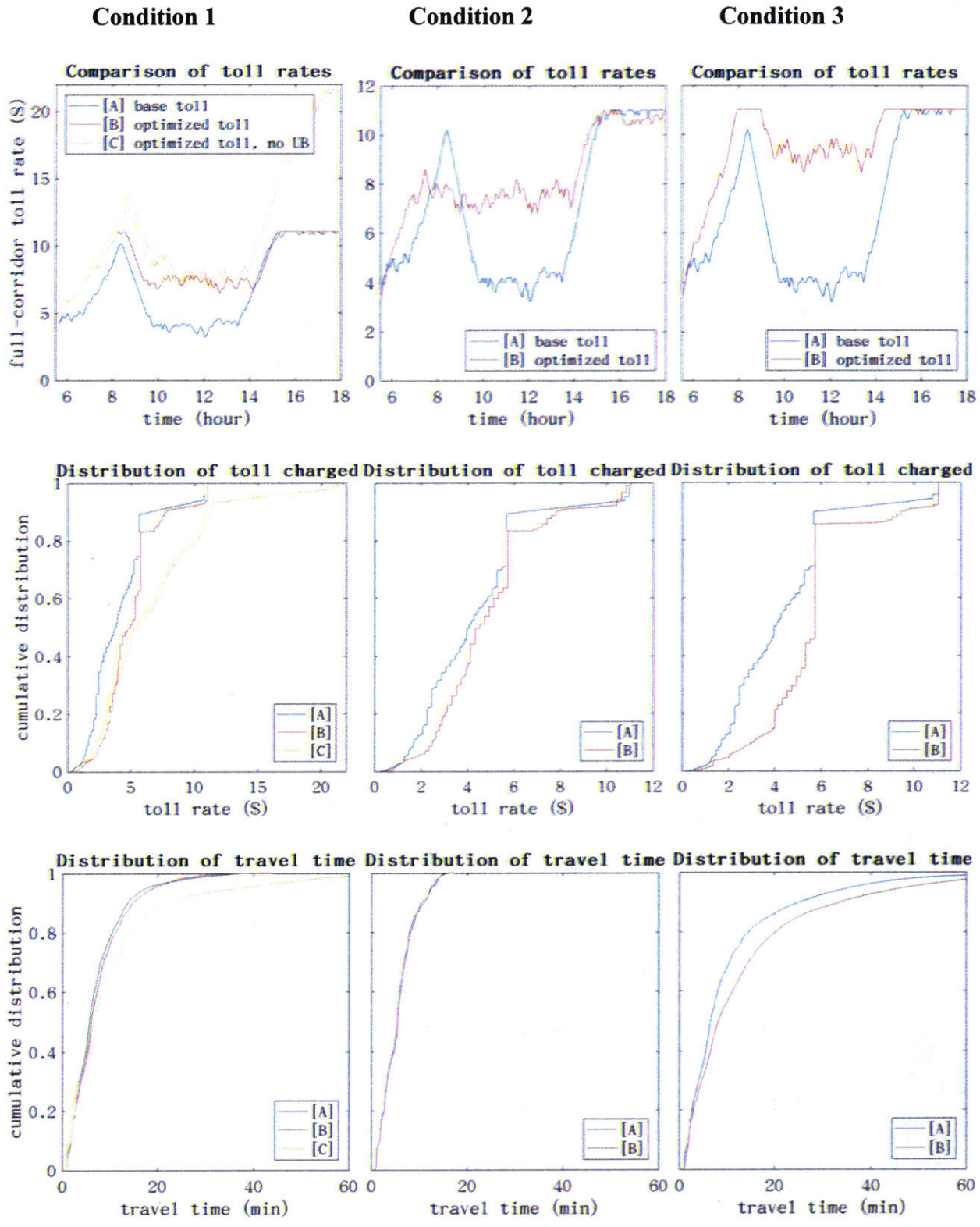

Figure 3.10: Toll rates in different scenarios

When upper bound on toll is not in effect, toll rates during AM and PM peak periods would increase to as high as twice the original upper bound, generating a revenue gain of 9.6% during the simulation period of 5:30-18:00, which is higher than 3.6%, the case where there is an upper bound. This indicates that there is still room for raising the toll rates above upper bound, based on travelers' elasticity to toll as implied by our route choice model. Nevertheless, the rate of increase reduces as the toll gets higher since the response of travelers to a higher toll is eventually effective, and the supply-demand interaction is working properly under the proposed framework. However, further increasing the toll rate makes ML more exclusive to those travelers who pay a much higher toll on average. As more travelers are pushed to GPL, average trip travel time increases.

When demand is 20% lower, optimized toll rates become lower than the base scenario due to less congestion on GPL but are still higher than the base toll during mid-day. Since mid-day is not congested, reducing demand does not affect optimized toll rates. Toll revenue would be lower than the base condition because of fewer trips, but applying toll optimization with online calibration still increases revenue by 3.5% compared to applying the base toll rates that are not adjusted dynamically. Travel times under the two scenarios are almost the same, as there is almost no congestion.

Condition [3] simulates travelers driving more conservatively, potentially because of bad weather. Due to slower deceleration rate, the headway between vehicles has to increase, thus the overall capacity of the freeway decreases. Due to more congestions, our toll optimization algorithm chooses to maintain much higher toll rates than [1B], while flow on ML is even higher than [1B] due to the congestion, thus generates a revenue gain of 9.8% comparing to base toll rates. Under more massive congestion, travelers choose ML even when toll rates are much higher, due to more substantial savings in travel time, and our proposed toll optimization framework benefits from online calibration to estimate and predict congestions.

The above tests in the simulation environment demonstrate the critical role of online calibration in the prediction-based dynamic toll pricing framework. When online calibration is enabled, and the system can replicate and predict traffic conditions with satisfactory accuracy, decisions on toll rates made by the DTA-based optimization is better than the case when no online calibration is

available, in terms of revenue maximization. The added benefit of online calibration is especially substantial when there is significant congestion on the network and is less evident when no congestion is present, which confirms that online calibration of supply parameters in an effort to match estimated and actual traffic speed is key to the success of the prediction-base toll pricing.

3.7 Conclusion

This chapter presents the calibration and optimization methodologies for a real-time toll pricing system. This system is integrated with a DTA system to optimize toll rates by evaluating toll revenues under predicted traffic conditions. Thus, online calibration is essential to ensure the DTA system accurately understands and predicts traffic conditions. We propose a heuristic online calibration algorithm to dynamically adjust supply parameters in the DTA system in response to real-time surveillance data. This algorithm is tested with real sensor data from a corridor consisting of ML and GPL, and the calibration accuracy is impressive, even when significant congestion is present. With online calibration enabled, we test the toll optimization in a closed-loop evaluation framework. A microscopic simulator is calibrated offline towards real data and integrated into the closed-loop evaluation framework as a representation of the real network. The DTA-based toll pricing system generates optimized toll rates, which are implemented in the microscopic simulator instead of in the real network.

The closed-loop evaluation is performed under 3 demand and supply conditions, including a base condition and 2 experimental conditions. Under the base condition, we also tested toll optimization without an upper bound on the toll rate. In each scenario, optimized toll rates are higher than the base toll rate, which is consistent with our prior belief, and higher revenue is obtained when optimized toll rates are implemented in the microscopic simulator, compared to the base toll rates generated in a system without online calibration. We also observe that the system is maintained to be real-time, i.e., the optimized toll is always obtained in less than 5 minutes.

We further discuss the results and implications from the regulators' perspective. Regulators usually represent the public interest and have the objective of maximizing social welfare. They negotiate with the ML operator to come up with mutually agreed tolling regulations before the operator

decides to invest in the ML project. Assume the ML operator applies the toll pricing system we develop, in each of the above scenarios, the optimized toll is generally higher than the base toll, which benefits the ML operator with higher revenue, while potentially benefit some ML users as the traffic on ML becomes lighter. However, some travelers would be pushed to the GPL due to the higher toll, while GPL users suffer from more traffic on the GPL. As the ML is generally not congested while the GPL is congested at some locations and times, reducing ML market shares makes GPL more congested while the benefit to ML users is limited. This makes the average travel time longer for all travelers, and therefore, the toll rate optimized for revenue maximization does not represent a socially optimal solution based on our simulation. If the toll rate is lower than the optimized toll, average travel time reduces. Therefore, the regulator has the potential to improve travel time (or other social welfare measurements) by altering the ML operator's tolling strategies, for example, providing a subsidy for ML usage or imposing a penalty for GPL congestion, in order to incentivize the operator to charge a lower toll than the optimized toll that maximizes revenue. Future research may look into the optimal regulations that maximize social welfare, given that the ML operator maximizes toll revenue under regulations. Note that, the optimized toll presented in this thesis does not reflect what the operator charges in the real world. Our recommendations to regulators are based on our simulation where the operator charges a revenue-maximizing toll while a lower toll would reduce overall travel time. Optimal strategies for the regulator may be different in other scenarios.

This research is conducted in a simulation environment where stochasticity is present. The results we present are based on a single test of the system for each scenario, and thus may be different in different replications due to the stochasticity in the microscopic simulator we apply to evaluate toll optimization. The revenue improvement we present may be different based on different realizations of the microscopic traffic simulation.

Due to stochasticity in traffic predictions by the DTA system, optimized toll rates are in general not the actual optimal values and may be different in different realizations. In addition, The current algorithm to solve the optimization problem is a grid search algorithm that in general does not find the actual optimal solution. A wider search range or a finer grid may improve the optimized revenue. Future research may improve the optimization solution algorithm without sacrificing

computational efficiency so that the revenue under optimized toll rates could be even higher. Robust toll optimization algorithms could be another future direction to account for the situation that the DTA system may not have perfect knowledge of travelers' choice behaviors and future network conditions.

The system relies on a discrete choice model to predict travelers' route choice under different traffic conditions and toll rates, and parameters in that model are known to the DTA system optimizing the toll. The results, in general, show that the optimized toll rate would be higher when overall demand is higher traffic congestions are more severe. We also expect the optimized toll to be highly correlated with the choice model parameters. Specifically, optimized toll rates would be higher if travelers are insensitive to toll. The results presented in the chapter are based on the choice model parameters we calibrated from flow and speed data for the travelers who utilize the corridor. Recent research by Burris and Brady (2018) suggests travelers' route choice behaviors may be more complicated than a discrete choice model that only depends on travel time and monetary cost. The next chapter develops a more comprehensive and personalized model for travelers' decisions to use managed lanes and apply the personalized choice model in toll optimization.

Contributions in this chapter include design, implementation, and testing of a heuristic online calibration method for supply parameters with the complexity of $O(n)$, where n represents the number of road segments in the network. Note that the algorithm is parallelizable, and it works simultaneously with the existing GLS OD calibration algorithm. Secondly, we implement the enhanced dynamic toll pricing framework and test it under multiple scenarios in a closed-loop testing framework. These tests show added benefit to toll optimization due to online calibration.

Chapter 4 Real-time Personalized Toll Optimization

This chapter presents the mathematical formulation of the real-time personalized toll optimization problem, and the DTA-based solution framework applied to solve the toll optimization problem. We then discuss the personalized choice model that is necessary for personalized toll optimization. Lastly, we present a case study with the optimization results.

4.1 Introduction

The previous chapter has discussed a toll pricing system where non-personalized toll rates are optimized on rolling horizons, and travelers passing through the same gantry in the same time interval are charged the same. The case study has demonstrated the importance of adjusting toll rates based on real-time traffic conditions. As discussed in Section 1.1.4, in addition to adaptive pricing, personalized pricing has the potential to boost revenue further if travelers are heterogeneous. However, completely personalized toll pricing may be unrealistic due to equity and public acceptance concerns. In addition, as two-way communication is expected to be available for only a fraction of travelers, the displayed toll rates are always necessary to accommodate travelers who do not have two-way communication with the system. Therefore, personalized discounting is more realistic than personalized pricing for practical considerations. While displayed toll rates are controlled dynamically, at the same time discounts are offered to eligible travelers.

With the availability of personalized trip data collected from AVI sensors, it is possible to develop a personalized choice model that captures the heterogeneity among travelers. Based on individuals' preferences to ML and sensitivities to toll and traffic, it is then possible to apply real-time personalized toll discounts with the infrastructure that enables two-way communication between eligible travelers and the system.

In Section 4.2, we firstly set aside any practical considerations discussed above and assume that personalized pricing is possible. In this way, each traveler may be priced differently, so that displayed toll rates are not needed, all travelers are considered an eligible traveler who is eligible for a personalized toll rate, and every traveler needs to send a signal before arrival to inquire his/her personalized toll rate. We derive optimization formulations under this hypothetical scheme. Then we bring back practical considerations and reformulate the optimization model to reflect the existence of displayed toll rates and derive the optimization formulation to accommodate personalized discounting for eligible travelers.

Then based on the toll pricing system developed for non-personalized tolling, in Section 4.3, we enhance the estimation and prediction modules of the DTA system by adding a personalized discount model and applying a personalized choice model. With these extensions to the system optimization module and a new personalized discount module that calculated individual traveler's discount when receiving a signal, the new toll pricing system can generate personalized discounts in addition to displayed toll rates.

In Section 4.4, we present details of the personalized choice model, which is developed based on personalized trip data. In the case study, the estimation results of the personalized choice model are presented, and the personalized toll pricing system is evaluated.

4.2 Optimization Formulation

In this section, we start with a naïve optimization formulation based on a simple ML network that consists of only one tolling segment and does not have intermediate ramps, as shown in Figure 1.1. Then we incorporate practical considerations by adding restrictions to the optimization problem in multiple steps and finally obtain a personalized toll optimization formulation that is ready to be solved by the real-time personalized toll pricing system which will then be presented in the next section.

With modest extensions, this optimization formulation is adapted to an ML network with multiple tolling segments and intermediate ramps.

4.2.1 Naïve Formulation of Personalized Pricing

To maximize revenue through personalized pricing, it is straightforward to produce the following optimization formulation.

$$\max_{\boldsymbol{p}} z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$\boldsymbol{\pi} = \mathrm{dta}(\boldsymbol{p}, \boldsymbol{D}, \boldsymbol{H}, \boldsymbol{B}) \tag{4.1}$$

where

p_i is the personalized toll price for individual i;

π_i is individual i's probability for choosing ML;

n is the number of individuals expected to make a choice in the prediction horizon;

$\boldsymbol{D}, \boldsymbol{H}, \boldsymbol{B}$ are OD, supply, and behavior parameters applied by the DTA system to perform traffic predictions.

The formulation is a direct extension to the non-personalized toll optimization formulation in the previous chapter, without incorporating any practical constraints. Instead of optimizing displayed toll rates in different time intervals, we optimize a vector of personalized toll rates $\boldsymbol{p}$, where p_i stands for individual i's personalized toll rate. Objective z is given by summing up expected revenue from all travelers making choices in the prediction horizon, where n is the number of travelers. π_i is individual i's probability of choosing ML, which is given by the prediction module of the DTA system denoted as $\mathrm{dta}(\cdot)$. $\boldsymbol{D_t}$ and $\boldsymbol{H_t}$ represent predicted OD demand and supply model parameters during the t-th time interval of the prediction horizon. $\boldsymbol{B_i}$ represents individual i's overall preference to ML and sensitivities to toll and traffic condition metrics, which are given by the personalized choice model. The above three parameters, in addition to the personalized toll rates $\boldsymbol{p}$, serves as input to the traffic simulation model applied in the prediction phase.

The traffic simulation model can be further decomposed as a demand model (choice model) generating route choice probabilities, and a supply model generating traffic condition metrics. The two models interact with each other. Therefore, the naïve optimization simulation can be re-written as follows.

$$\max_{\boldsymbol{p}} z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$\pi_i = \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})$$

$$\boldsymbol{S} = \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H})$$

(4.2)

where

$\boldsymbol{S_{t(i)}}$ represents the traffic condition metrics in interval t during which individual i makes the choice, which are given by the supply model as a function of travelers' choice probabilities $\boldsymbol{\pi}$ and the OD and supply parameters $\boldsymbol{D}$ and $\boldsymbol{H}$.

Individual i's probability of choosing ML (π_i) is given by a choice model as a function of that individual's personalized toll rate p_i, the traffic condition metrics at the time interval when individual i makes the choice $\boldsymbol{S_{t(i)}}$, and individual i's overall preference to ML and sensitivities to traffic $\boldsymbol{B_i}$. Details of the choice model are given in Section 4.4.

4.2.2 Two-level Formulation of Personalized Pricing

While the above naïve formulation is simple and easy to interpret, it is unrealistic for real-world applications. Firstly, the decision variable is n-dimensional, where n is the number of travelers during the prediction horizon. As we are using a simulation-based DTA system to predict traffic conditions, traffic simulation needs to be performed whenever the objective function needs to be evaluated. As the dimension of the decision variable is large, the number of times that the objective needs to be evaluated would be huge. Considering the computational time required to perform traffic simulation, a simulation-based optimization problem with an n-dimensional decision variable is not practical to be solved in real time. Secondly, as discussed in Chapter 1, the optimization problem is to be solved with a proactive toll pricing system that runs in rolling horizons. As the system operates at each rolling time period and takes time to solve simulation-based optimization, a traveler is unable to receive his/her personalized price immediately after sending a signal. After sending a signal, the traveler needs to wait for the system's next operation and wait until the system finishes optimization before obtaining his/her personalized price. Therefore, with the naïve formulation, the system is unable to present a personalized toll rate to a

traveler immediately after receiving his/her signal, and in case the signal is not received well in advance before a traveler arrives, a toll rate would not be available when the traveler needs to make a choice. These two issues make the naïve optimization formulation impossible to be implemented based on the infrastructure system we propose. To resolve these issues, we apply a two-level framework that consists of a personalized pricing module and a system optimization module, introduced in Section 1.2.2.

The personalized pricing module runs in real time and generates a personalized toll rate immediately after receiving a signal. This module does not involve simulation and the personalized toll rate is obtained analytically, which ensures its real-time performance so that an eligible traveler can receive a personalized price immediately after sending a signal. Personalized pricing depends on not only traffic conditions and individual preferences, but also a control parameter that is optimized at the system level. The system-level optimization is solved based on simulations.

Personalized pricing module:

$$p_i = \text{price}(x^*_{t(i)}, \boldsymbol{S}^*_{\boldsymbol{t(i)}}, \boldsymbol{B_i}) \tag{4.3}$$

System optimization module:

$$\max_{\boldsymbol{x}} z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$\begin{aligned} p_i &= \text{price}(x_{t(i)}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \\ \pi_i &= \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \\ \boldsymbol{S} &= \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H}) \end{aligned} \tag{4.4}$$

where

$x_{t(i)}$ is referred to as the control parameter for time interval $t(i)$, which are arguments to the personalized pricing function and controls how personalized prices are generated;

$x^*_{t(i)}$ is the control parameter's optimized value for time interval $t(i)$ found by system optimization;

$t(i)$ represents the time interval when individual i arrives at the network and makes a choice between ML and GPL.

The personalized toll rate p_i is generated with a personalized pricing function, conditional on the control parameter for time interval $t(i)$ that is optimized in the system level, denoted as $x^*_{t(i)}$, as well as the predicted traffic conditions $\boldsymbol{S}^*_{\boldsymbol{t(i)}}$ and personalized preferences and sensitivities $\boldsymbol{B_i}$. The personalized pricing function is formulated based on a personalized pricing policy we propose, and the next section is devoted to the policy.

The control parameter controls the overall price level based on overall traffic conditions and is not personalized. The system optimization module finds the control parameters for each time interval that maximize overall expected revenue, where revenue is evaluated by the traffic prediction module of the DTA system. As the DTA system runs in rolling horizons, the system optimization module also runs in rolling horizons. System optimization is formulated as Equation 4.4. In the DTA demand model denoted as $\text{choice}(\cdot)$, the personalized price p_i is given by the same personalized pricing function as in the personalized pricing module. However, as we allow eligible travelers to send the signal right before arrival, the pricing system does not have full knowledge on which travelers are traveling during the prediction horizon. Therefore, system optimization is based on predictions of travelers who are arriving and making choices during the prediction horizon. Therefore, in system optimization, the personalized pricing function is applied to predicted travelers, whereas the personalized pricing module applies personalized pricing function to actual travelers upon receiving a signal.

Compared to the naïve formulation, the two-level formulation significantly reduces the dimension of the decision variable from n to m, where n is the number of travelers while m is the number of time intervals in the prediction horizon. It also resolves the second issue of the naïve formulation, as the personalized price can be calculated immediately by the personalized pricing module whenever a signal is received, whereas with the naïve formulation personalized prices are optimized in each rolling period, and an eligible traveler is unable to receive a personalized price immediately after sending a signal. However, the two-level formulation imposes restrictions compared to the naïve formulation, as each individual is prices based on the same personalized pricing function with the same control parameter for each time interval, and thus the two-level formulation theoretically may not find the optimal vector of personalized toll rates while the naïve formulation can.

4.2.3 Personalized Pricing Policy

In this section, we derive a personalized pricing policy that is consistent with the system-level objective of maximizing expected revenue in the near future (i.e., the prediction horizon). We apply a first-degree price discrimination approach, as this is most effective in maximizing revenue. Firstly, we derive p_i^O that maximizes expected revenue from individual i given traffic conditions, without taking into consideration the impact of individual i's choice to traffic conditions. We refer to p_i^O as the personal optimal price. As we apply a random utility choice model, the personal optimal price is given by the following equation.

$$p_i^O = \text{argmax}_p \; p \cdot \text{choice}\left(p, \boldsymbol{S}_{t(i)}, \boldsymbol{B}_i\right) \tag{4.5}$$

Equivalently,

$$p_i^O + \frac{\text{choice}\left(p_i^O, \boldsymbol{S}_{t(i)}, \boldsymbol{B}_i\right)}{\left.\frac{\partial \text{choice}\left(p, \boldsymbol{S}_{t(i)}, \boldsymbol{B}_i\right)}{\partial p}\right|_{p=p_i^O}} = 0 \tag{4.6}$$

where

p_i^O is the personalized price that maximizes expected revenue from individual i.

Note that, individual i's price elasticity of choice probability equals -1 when the price is at p_i^O. However, while charging individual i at p_i^O maximizes revenue from this particular individual, if we charge every individual at their personal optimal price, overall revenue is not maximized, as demand-supply interactions are not taken into consideration.

Individual i's decision to choose ML rather than GPL affects expected traffic conditions on both ML and GPL for later users. GPL traffic gets slightly better while ML traffic gets slightly worse if individual i switches from GPL to ML. Such an impact on traffic reduces other travelers' WTP for ML, and thus expected revenue from those travelers is reduced. Note that, WTP represents the portion of ML utility except for price, or the net benefit of ML over GPL without consideration of price.

As individual i's switching from GPL to ML reduces expected revenue from later travelers, we denote the amount of such revenue loss as $X_{i,t(i)}^C$. From the ML operator's point of view, $X_{i,t(i)}^C$ is an opportunity cost if individual i takes ML. It also represents the value of a marginal unit of

capacity on ML. Maximizing revenue independently for each traveler by charging p_i^O would be short-sighted as there is an opportunity cost for providing ML to a traveler.

By taking into consideration the opportunity cost, the optimal price for individual i is then given by the following equation.

$$p_i^C = \text{argmax}_p \, (p - X_{i,t(i)}^C) \cdot \text{choice}(p, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \tag{4.7}$$

Equivalently,

$$p_i^C + \frac{\text{choice}(p_i^C, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\left.\frac{\partial \text{choice}(p, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\partial p}\right|_{p=p_i^C}} = X_{i,t(i)}^C \tag{4.8}$$

where

p_i^C is the personalized price that maximizes expected revenue minus cost from individual i;

$X_{i,t(i)}^C$ is the opportunity cost of providing ML to individual i during time interval $t(i)$.

The opportunity cost for each individual is unknown, and due to the complexity of traffic dynamics, it is difficult to be obtained analytically. However, we are aware that overall revenue is maximized, only if we price each traveler based on a correct estimate of the opportunity cost. Therefore, we formulate an optimization model to find the best estimate of opportunity cost by maximizing revenue. We solve the optimization with a DTA-based solution framework where we apply the DTA system to evaluate the objective under candidate decision variable values.

$$\max_{\boldsymbol{X^C}} z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$p_i = p_i^C \text{ such that } p_i^C + \frac{\text{choice}(p_i^C, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\left.\frac{\text{choice}(p, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\partial p}\right|_{p=p_i^C}} = X_{i,t(i)}^C$$

$$\pi_i = \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})$$

$$\boldsymbol{S} = \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H}) \tag{4.9}$$

However, as discussed previously, the simulation-based optimization problem with an n-dimensional decision variable $\boldsymbol{X^C}$ is not computationally practical for real-time applications. We attempt to reduce the dimension by assuming that the opportunity cost is homogeneous among individuals but different at different time intervals.

Opportunity cost $X^C_{i,t(i)}$ depends on the product of two factors: (A) how individual i's presence on ML and absence from GPL affects traffic conditions on ML and GPL, and (B) how ML and GPL traffic conditions affect expected revenue from later travelers. Factor B is independent of any characteristics of individual i, thus homogeneous among travelers. However, it depends on the OD demand and traffic conditions at each interval, and thus, it is different at different time intervals.

Factor A mainly depends on the vehicle size and driving behavior of individual i. A larger vehicle has a more significant impact on traffic. Drivers who follow the traffic flow, in general, have less impact on the traffic than those who switch lanes and change speed frequently. As this study only involves personalized pricing for passenger cars, the heterogeneity in vehicle size is minimal. Besides, the mesoscopic DTA system applied in this study to solve the optimization problem relies on a traffic dynamics model and a queuing model for traffic simulation, and thus any heterogeneity in driving behavior is not captured in the DTA-based solution framework, as well as any heterogeneity in vehicle size as long as all vehicles are modeled as passenger cars. Therefore, we assume that each traveler has the same impact on traffic, and thus, Factor A is homogeneous among individuals. While this assumption may seem strong in reality, it is an inherent assumption that is already in effect in the DTA system. Therefore, imposing this assumption in the optimization formulation theoretically does not affect the solution. For other research that involves different types of vehicles, type-specific opportunity cost may be specified.

The opportunity cost depends on Factor A and B, and we have assumed Factor A to be homogeneous among individuals. Since Factor B is also homogeneous but time-dependent, the opportunity cost is then assumed homogeneous among individuals but different for each time interval, so that it is x^C_t

$$x^C_t = X^C_{i,t(i)} \; \forall \; i \text{ such that } t(i) = i \tag{4.10}$$

With the above assumption, the following optimization formulation is derived, which is consistent with the two-level model structure proposed in the previous section.

Personalized pricing module:

$$p_i = \text{price}^{\text{C}}(x_{t(i)}^{C*}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \tag{4.11}$$

where

i is the index for the actual travelers sending a signal to the system;

$t(i)$ is the time interval when individual i makes the choice;

p_i is the personalized price for individual i;

$x_{t(i)}^{C*}$ is the optimized estimate of the opportunity cost in interval $t(i)$, which serves as the control parameter;

$\boldsymbol{S_{t(i)}}$ is the traffic condition metrics in interval $t(i)$;

$\boldsymbol{B_i}$ is individual i's preference to ML and sensitivities to toll and traffic;

$\text{price}^{\text{C}}(\cdot)$ is the personalized pricing function formulated with the opportunity-cost-based Personalized pricing policy:

$$\text{price}^{\text{C}}\left(x_{t(i)}^{C}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}\right) = p_i^C \text{ such that}$$

$$p_i^C + \frac{\text{choice}\left(p_i^C, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}\right)}{\left.\frac{\partial \text{choice}\left(p, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}\right)}{\partial p}\right|_{p=p_i^C}} = x_{t(i)}^{C} \tag{4.12}$$

System optimization module:

$$\max_{\boldsymbol{x^C}} z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$\begin{gathered} p_i = \text{price}^{\text{C}}(x_{t(i)}^{C}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \\ \pi_i = \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}) \\ \boldsymbol{S} = \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H}) \end{gathered} \tag{4.13}$$

where

i is the index for the travelers that the system predicts will use the network during the prediction horizon;

n is the number of travelers predicted to use the network during the prediction horizon;

$\boldsymbol{x^C}$ is the vector of opportunity cost estimates for each interval within the prediction horizon, which serves as the control parameters for personalized toll pricing;

z is the objective, i.e., the expected revenue during the prediction horizon;

p_i is individual i's personalized toll rate;

π_i is individual i's probability of choosing ML.

The real-time personalized toll optimization problem we propose is formulated above. This formulation is valid for a simple ML network and when practical considerations are not incorporated. In this formulation, we apply a two-level decision process, where the personalized toll pricing module provides a personalized toll rate based on toll pricing policy, while the system optimization module optimizes a parameter for personalized toll pricing.

In this section, we propose an opportunity-cost-based pricing policy, where each individual is priced by maximizing expected revenue from that individual while taking opportunity cost into consideration. The estimate of opportunity cost is the control parameter that is optimized by system optimization. If the assumption that each traveler has the same impact on traffic is true, then the above optimization formulation is most effective in maximizing revenue, and theoretically generates the highest possible expected revenue. If the assumption does not hold, then the restriction that each individual is priced based on the same estimate of opportunity cost would reduce expected revenue compared to another formulation and solution framework that can capture the heterogeneity in an individual's impact on traffic. The next section introduces restrictions to the optimization problem in order to account for practical concerns so that the methodology is realistic to be applied in a real-world ML network with an infrastructure system available as we have proposed.

4.2.4 Formulation of Personalized Discounting

As discussed previously, personalized pricing where each individual gets a different price may be considered immoral and lack of public acceptance, and thus not realistic for real-world applications. In addition, the previous formulation requires all travelers to be able to make two-way communications with the system, which is also impossible. Therefore, instead of charging each traveler at a different price, we optimize a displayed toll rate and offer personalized discounts to eligible travelers. To do so, we impose restrictions on the optimization problem such that for each time interval, (1) price for ineligible travelers are the same, and (2) personalized prices for eligible travelers do not exceed the price for ineligible travelers. We also rephrase personalized pricing as personalized discounting. Besides, penalty terms for ML flow or speed exceeding regulatory requirements are added to system optimization, as well as an upper bound on displayed toll rate. Then the new formulation is obtained as below.

Personalized discount module:

$$r_i = \text{discount}^{\text{C}}(x_{t(i)}^{C*}, P_{t(i)}^{D*}, \boldsymbol{S}_{\boldsymbol{t(i)}}, \boldsymbol{B_i}, a_i) \tag{4.14}$$

where

r_i is the discount ratio for individual i upon receiving the signal from individual i;

$P_{t(i)}^{D*}$ is the optimized displayed toll rate at interval $t(i)$ generated by system optimization;

a_i is a dummy variable indicating if individual i is eligible for a discount, which always equals 1 in the personalized discount module, but may be 0 in system optimization;

$\text{discount}^{\text{C}}(\cdot)$ is the personalized discount function formulated with the opportunity-cost-based personalized discount policy:

$$\text{discount}^{\text{C}}\left(x_{t(i)}^{C}, P_{t(i)}^{D}, \boldsymbol{S}_{t(i)}, \boldsymbol{B_i}, a_i\right) = r_i \text{ such that}$$

$$\text{if } a_i = 0, r_i = 1 \tag{4.15}$$

$$\text{if } a_i = 1, r_i = \min\left(1, \frac{p_i^C}{P_{t(i)}^D}\right) \text{ such that} \tag{4.16}$$

$$p_i^C + \frac{\text{choice}\left(p_i^C, \boldsymbol{S}_{\boldsymbol{t(i)}}, \boldsymbol{B_i}\right)}{\left.\frac{\text{choice}\left(p, \boldsymbol{S}_{\boldsymbol{t(i)}}, \boldsymbol{B_i}\right)}{\partial p}\right|_{p=p_i^C}} = x_{t(i)}^C \tag{4.17}$$

System optimization module:

$$\max_{\boldsymbol{x}^C, \boldsymbol{P}^D} Z = \sum_{i=1}^{n} p_i \pi_i$$

s.t.

$$p_i = P_{t(i)}^D \times \text{discount}^C\left(x_{t(i)}^C, P_{t(i)}^D, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}, a_i\right)$$

$$\pi_i = \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})$$

$$(\boldsymbol{S}, \boldsymbol{Q}, \boldsymbol{V}) = \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H})$$

$$LP_t^D \leq p^{constraint} \; \forall t$$

$$Q_{tk} \leq q_k^{UB} \; \forall t \; \forall k$$

$$V_{tk} \geq v_k^{LB} \; \forall t \; \forall k$$

(4.18)

The above optimization problem is to be solved in the DTA-based solution framework.

4.2.5 Alternative Personalized Discount Policies

The two-level decision process is flexible to accommodate alternative personalized discount policies, where the opportunity-cost-based personalized discount policy is one example. A discount policy may incorporate different considerations, for example, equity concerns. It may also be designed to accommodate third-degree price discrimination, where the discount rate is based on a group that the traveler belongs to. The interpretability of the discount policy is also important.

We propose an alternative policy called a reduced discount policy. With this policy, the personalized discount module first obtains the optimal personal price p_i^O, calculate the amount of toll discount based on this price, and reduce the discount by a discount reduction factor $x_{t(i)}^R$. For example, if $p_i^O = \$6$ and displayed price $P_{t(i)}^D = \$10$, the full discount is \$4 or 40%. If discount reduction factor $x_{t(i)}^R = 25\%$, discount will be reduced to \$3 or 30%, and thus the discount ratio is $1 - 30\% = 70\%$. Toll optimization with the reduced discount policy is formulated below.

Personalized discount module:

$$r_i = \text{discount}^{\text{R}}(x_{t(i)}^{R*}, P_{t(i)}^{D*}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}, a_i) \tag{4.19}$$

where

$x_{t(i)}^{R*}$ is the discount reduction factor that serves as a control parameter;

$\text{discount}^{\text{R}}(\cdot)$ is the personalized discount function formulated with the reduced discount policy:

$$\text{discount}^{\text{R}}\left(x_{t(i)}^{R}, P_{t(i)}^{D}, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}, a_i\right) = r_i \text{ such that}$$

$$\text{if } a_i = 0 \text{ or } p_i^O > P_{t(i)}^D, r_i = 1 \tag{4.20}$$

$$\text{if } a_i = 1 \text{ and } p_i^O < P_{t(i)}^D, r_i = 1 - (1 - x_{t(i)}^R)(1 - \frac{p_i^O}{P_{t(i)}^D}) \tag{4.21}$$

where

$$p_i^O + \frac{\text{choice}(p_i^O, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\left.\frac{\text{choice}(p, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})}{\partial p}\right|_{p=p_i^O}} = 0 \tag{4.22}$$

$\frac{p_i^O}{P_{t(i)}^D}$ is the price factor for individual i, $(1 - \frac{p_i^O}{P_{t(i)}^D})$ is the discount rate. As the discount is reduced by $x_{t(i)}^R$, $(1 - x_{t(i)}^R)(1 - \frac{p_i^O}{P_{t(i)}^D})$ represents the reduced discount rate, and thus $1 - (1 - x_{t(i)}^R)(1 - \frac{p_i^O}{P_{t(i)}^D})$ is the price factor after discount reduction.

System optimization module:

$$\max_{\boldsymbol{x^R}, \boldsymbol{P^D}} Z = \sum_{i=1}^{n} p_i \pi_i + \sum_{t=1}^{m} W_t^Q \min(0, Q_t - q^{UB}) + W_t^V \min(0, v^{LB} - V_t)$$

s.t.

$$p_i = P_{t(i)}^D \times \text{discount}^{\text{R}}\left(x_{t(i)}^R, P_{t(i)}^D, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}, a_i\right)$$

$$\pi_i = \text{choice}(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i})$$

$$(\boldsymbol{S}, \boldsymbol{Q}, \boldsymbol{V}) = \text{supply}(\boldsymbol{\pi}, \boldsymbol{D}, \boldsymbol{H})$$

$$P_t^D \leq p^{constraint} \; \forall t$$

$$0 \leq \boldsymbol{x^R} \leq 1 \tag{4.23}$$

4.2.6 Generalization to a Full Managed Lane Network

In previous sections, we have formulated the personalized toll optimization problem for a simple ML network consisting of only one tolling segment without intermediate ramps. However, a real ML network usually consists of multiple tolling segments where different toll rates are applied. There may also be intermediate entry ramps, and travelers entering the ML through those ramps are charged a pre-determined fraction of the full-segment toll rate. As travelers are charged when entering the network, their toll rates are not affected by where they exit.

The two-level personalized toll optimization model we develop can be extended to accommodate a full ML network. In the system optimization module, dimensions of decision variables will increase as the network gets more complicated. As there are g gantries on the network, the dimension of the displayed toll rate $\boldsymbol{P^D}$ would be $g \times m$ where m is the number of time intervals in the prediction horizon. As discussed in Chapter 3, linear constraints may be added on $\boldsymbol{P_t^D}$. In particular, the displayed toll rate on a gantry at intermediate entry ramps is a pre-determined fraction of the full-segment toll rate. These equality constraints effectively reduce the dimension of $\boldsymbol{P_t^D}$ from $\boldsymbol{g}$ to the number of tolling segments in the network.

As for the discount control parameter, ideally, it needs to be optimized for each origin-destination (OD) pair on the network. In case the opportunity-cost-based discount policy is applied, and we optimize an estimate of opportunity cost $\boldsymbol{x^C}$, it is reasonable to assume the opportunity cost is proportional to the displayed toll price for each OD pair. Therefore, instead of optimizing the opportunity cost itself, we optimize the ratio of opportunity cost over the displayed toll, which is assumed to be the same across the network at the same time. In case the reduced discount policy is applied so that we optimize the discount reduction factor $\boldsymbol{x^R}$, we restrict $\boldsymbol{x^R}$ to be the same across the network at the same time. In both cases, the dimension of the discount control parameter does not scale as the network scales.

In the personalized discount module, the discount is calculated based on expected traffic conditions for the specific OD that the traveler is on. In case there are more than 2 possible paths, the discount is generated with the assumption that the traveler chooses either full ML or full GPL. Based on the

trip data used for the estimation of the choice model, we find that only a small percentage of travelers decides to switch between ML and GPL in the middle of their trips.

4.3 DTA-based Solution Framework

Figure 4.1 demonstrates the real-time personalized toll pricing system we develop to solve the personalized toll optimization problem. The personalized discount module generates a discount offer r_i upon receiving a signal from individual i. The system optimization module is illustrated in Figure 1.2. It is the same as that for non-personalized toll pricing as presented in Chapter 3, except for the following enhancements:

(1) The demand model in the DTA system consists of a personalized choice model where each individual has different preferences to ML and sensitivities to toll and traffic.

(2) The demand model includes a new personalized discount model that applies the personalized discount function to determine personalized toll rates for simulated travelers in the DTA system, based on the control parameters.

(3) The control strategy refers to controlling not only the displayed toll rate but also the discount control parameter.

(4) Revenue is calculated within the demand model as a sum of each traveler's expected revenue based on the choice model, whereas in the non-personalized toll pricing system revenue is the sum over all gantries of the product of gantry toll and predicted gantry flow.

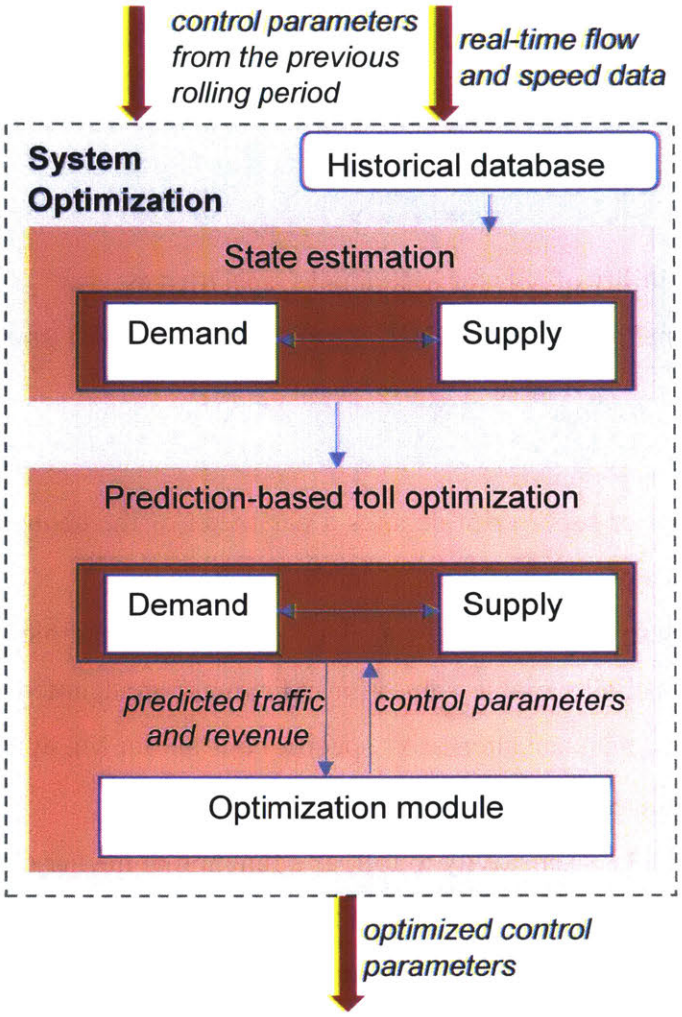

Figure 4.1: The system optimization module

4.4 Personalized Choice Model

Choice model parameters are important arguments to the personalized discount function. For a given choice set at a given time, the personalized discount module generates different discounts for different eligible travelers only because they are heterogeneous in their choice behavior. Therefore, personalized discounting is only meaningful when heterogeneities among travelers' choice behavior are modeled. We develop a personalized choice model that captures systematic heterogeneity in individuals' overall preferences to ML and sensitivities to toll rate and traffic condition metrics.

4.4.1 Model Specification

The choice model is a binary logit model so that individual i's probability of choosing ML π_i is given as:

$$\pi_i = \frac{e^{v_i}}{e^{v_i}+1} = \frac{1}{1+e^{-v_i}} \tag{4.24}$$

where v_i is the systematic utility of ML for individual i, specified as:

$$v_i = ASC_{t(i)} + senToll_i \times scaledToll_i + senTime_i \times scaledTimeSaving_{t(i)} + b^{DPspeed} \times (s_{t(i)}^{DPspeed} - s_i^{histDPspeed}) + ISC_i \tag{4.25}$$

where

b with any subscript or superscript denotes a parameter of the choice model that is to be estimated from data. The set of all parameters is denoted as $\boldsymbol{b}$.

s_i with any superscript denotes individual i's value of a variable generated from the disaggregated data. The set of all variable values for individual i are denoted as $\boldsymbol{s_i}$.

$ASC_{t(i)}$ is the TOD-dependent alternative specific constant for ML at time interval t during which individual i makes the choice

$senToll_i$ is individual i's sensitivity to toll, as a function of his/her characteristics

$scaledToll_i$ is a box-cox transformation of the toll rate for individual i

$senTime_i$ is individual i's sensitivity to travel time, as a function of his/her characteristics

$scaledTimeSaving_{t(i)}$ is the difference between GPL and ML travel times at time interval t during which individual i makes the choice, linearly scaled based on which TOD.

$b^{DPspeed}$ is travelers' sensitivity to decision-point traffic speed

$s_{t(i)}^{DPspeed}$ is the decision-point traffic speed at time interval t during which individual i makes the choice

$s_i^{histDPspeed}$ is individual i's average experienced decision-point speed in the past.

ISC_i is the individual specific constant that measures individual i's general preferences to ML, as a function of his/her characteristics

The above terms are further explained below.

(1) $ASC_{t(i)} = b^{ASC} + \sum_{k=1}^{d} b_k^{sin} \sin(2k\pi s_{t(i)}^{TOD}) + b_k^{cos} \cos(2k\pi s_{t(i)}^{TOD})$ (4.26)

$ASC_{t(i)}$ is the alternative specific constant for ML, measuring travelers' overall preference to ML and any factors that affect preference but are not captured by other terms of the utility function. It consists of a constant term b^{ASC} and a Fourier series with d pairs of $\sin(\cdot)$ and $\cos(\cdot)$ functions that capture the time-of-day variation in preferences. $x_{t(i)}^{TOD}$ is the TOD value during the interval $t(i)$ when individual i makes the choice, and is a number from 0 (representing 0:00) to 1 (representing 24:00) but does not reach 1.

(2) $senToll_i = -\exp(b^{toll} + b^{tollFreq} \times s_i^{freq} + b^{tollLoyal} \times s_i^{loyal} + b^{tollNoTrip} \times s_i^{noTrip})$ (4.27)

$senToll_i$ measures individual i's sensitivity to the (scaled) toll rates, which is a negative number specified as a function of the individual's travel frequency in the past 90 days s_i^{freq} measured as the average number of trips per day, loyalty to ML in the past 90 days s_i^{loyal}, and the no-trip indicator s_i^{noTrip} where $s_i^{noTrip} = 1$ indicates the individual has not made any trip during the past 90 days.

s_i^{loyal} measures individual i's loyalty to ML, which is specified as a function of the number of trips he/she made on ML and that on GPL during the past 90 days: $s_i^{loyal} = \log(\frac{s_i^{MLtrips}+1}{s_i^{GPLtrips}+1})$. An individual who has made more trips on ML and fewer trips on GPL is more loyal to ML. $s_i^{loyal} = 0$ if the person does not have any trip or have equal numbers of ML and GPL trips in the past 90 days. The dummy variable s_i^{noTrip} is included in the specification to distinguish these two cases.

(3) $scaledToll_i = \frac{\left(s_i^{toll}+1\right)^{b^{tollExponent}} - 1}{b^{tollExponent}}$ (4.28)

To capture the fact that travelers' sensitivity to the same increment of toll may be different at different toll levels, we apply a box-cox transformation on toll rates. s_i^{toll} is individual i's toll rate. $b^{tollExponent} > 1$ represents that travelers are more sensitive to an increment of toll when toll is higher, while $b^{tollExponent} < 1$ represents travelers are less sensitive to an increment of toll when toll is higher. If $b^{tollExponent} = 1$, the toll is not scaled.

(4) $senTime_i = \exp(b^{time} + b^{timeFreq} \times s_i^{freq} + b^{timeLoyal} \times s_i^{loyal} + b^{timeNoTrip} \times s_i^{noTrip})$ (4.29)

$senToll_i$ measures individual i's sensitivity to the (scaled) time saving on ML compared to GPL, which is specified as a function of s_i^{freq}, s_i^{loyal}, and s_i^{noTrip}.

(5) $scaledTimeSaving_{t(i)} = s_{t(i)}^{timeSaving} \times exp\Big(b^{timeNight} \times s_{t(i)}^{isNight} + b^{timeAM} \times s_{t(i)}^{isAM} + b^{timePM} \times s_{t(i)}^{isPM}\Big)$ (4.30)

As sensitivity to time savings may be different in different periods of the day, we scaled the time saving to capture this effect. We define mid-day as the base period for which time-saving is not scaled. We use dummy variables $s_{t(i)}^{isNight}$, $s_{t(i)}^{isAM}$, and $s_{t(i)}^{isPM}$ to denote if the time interval $\mathrm{t}(i)$ is during the night, AM peak, or PM peak. A scaling parameter is estimated if the choice is made during one of the above three periods.

(6) $b^{DPspeed}$, $s_{t(i)}^{DPspeed}$, and $s_i^{histDPspeed}$

$s_{t(i)}^{DPspeed}$ is the traffic speed in time interval $\mathrm{t}(i)$ at the decision point (DP) where travelers need to make the choice between ML and GPL. $s_i^{histDPspeed}$ is the average DP speed that traveler i experienced among his/her trips during the past 90 days. The different $(s_{t(i)}^{DPspeed} - s_i^{histDPspeed})$ measures his/her impression of the travel conditions when making the choice, and we expect an individual is more likely to choose ML is DP speed is lower than his/her experienced DP speed.

(7) $ISC_i = b^{pastMLtimeDiff} \times s_i^{pastMLtimeDiff} + b^{pastGPLtimeDiff} \times s_i^{pastGPLtimeDiff}$
$+b^{loyal} \times s_i^{loyal} + b^{hadMLtrip} \times s_i^{hadMLtrip} + b^{hadGPLtrip} \times s_i^{hadGPLtrip}$
$+b^{loyal30} \times s_i^{loyal30} + b^{hadMLtrip30} \times s_i^{hadMLtrip30} + b^{hadGPLtrip30} \times s_i^{hadGPLtrip30}$
$+(b^{lastTripWasML} + b^{lastTripWasMLmemory} \times memoryFactor) \times s_i^{lastTripWasML}$
$+b^{lastTripWasGPLmemory} \times memoryFactor \times s_i^{lastTripWasGPL}$
$+b^{lastTripTimeDiffML} \times s_i^{lastTripTimeDiff} \times memoryFactor \times s_i^{lastTripWasML}$
$+b^{lastTripTimeDiffGPL} \times s_i^{lastTripTimeDiff} \times memoryFactor \times s_i^{lastTripWasGP}$

where $memoryFactor = \exp\left(-b^{memoryDecayRate} \times s_i^{daysSinceLastTrip}\right)$ (4.31)

$s_i^{pastMLtimeDiff}$ is individual i's experience on his/her ML trips in the past 90 days, averaged among trips. Experience in a trip is measured by the difference between actual travel time for that trip and the expected travel time before making the choice of taking ML for that trip. $s_i^{pastGPLtimeDiff}$ is individual i's experience on his/her GPL trips in the past 90 days, averaged among trips. These variables are 0 if he/she does not have a past ML/GPL trip.

$s_i^{hadMLtrip}$, $s_i^{hadGPLtrip}$, $s_i^{hadMLtrip30}$, and $s_i^{hadGPLtrip30}$ are dummy variables indicating if individual i have had any ML/GPL trips in the past 90 or 30 days. s_i^{loyal} is individual i's loyalty to ML based on the number of trips he/she has made on ML and GPL in the past 90 days. $s_i^{loyal30}$ is the same measurement of loyalty with the past 30 days' data.

$s_i^{lastTripWasML}$ is a dummy variable indicating if individual i had trips in the past 90 days and the most recent one was on ML. There might be memory effects so that the individual is more likely to choose ML for the current trip if he/she chose ML for the most recent trip. We use a fixed component and a variable component to capture this effect, where the variable component diminishes as the most recent trip gets further away from the current trip. The diminishing effect is represented with the $memoryFactor$ which is between 0 and 1 and diminished based on an exponential decay function as the number of days since last trip increases. Similarly, if individual i had trips in the past 90 days and the most recent one was on GPL, there is also a memory effect. As the fixed component has perfect collinearity with other variables, only the variable component of this memory effect is estimated.

$s_i^{lastTripTimeDiff}$ is individual i's experience on his/her most recent trip, which would be 0 if he/she does not have a trip in the past 90 days. The term captures that current choice may be affected by the experience on the most recent trip in additional average experience on all past trips, and the impact may be different depending on if the most recent trip was on ML or GPL. All the above terms together are referred to as the individual specific constant (ISC), which measured individuals' overall preferences to ML.

4.4.2 Model Estimation and Application in the Toll Pricing System

We apply maximum likelihood estimation to estimate the model parameters $\boldsymbol{b}$ using disaggregate trip data $\boldsymbol{s}$. Note that, if an individual has multiple trips in the data set, his/her characteristics may be different as those trips are made at different times when his/her experience during the 90 days before each trip is different. So, we use a different index i to denote the same individual on different trips. Based on the estimated model parameters as well as variables values at each trip $\boldsymbol{s_i}$, we calculate each individual's parameters for each trip, and take the average among all trips by the same individual to obtain a database of travelers and their individual-specific parameters including $senToll_i$, $senTime_i$, $c_i^{histDPspeed}$, and ISC_i. As the individuals are samples of the population, where the population is defined as the set of all potential travelers on the ML network, the distributions of the sensitivities and the ISC are a fair representation of the population distribution. To perform traffic simulation in the system optimization module, the DTA demand model resamples from the samples by randomly drawing travelers from the database. Each individual's probability of being drawn depends on how many trips he/she has made during the time period (night/AM peak/mid-day/PM peak) that is being simulated.

The demand model in the DTA system predicts individuals' choice probabilities as follows:

$$\pi_i = \text{choice}\left(p_i, \boldsymbol{S_{t(i)}}, \boldsymbol{B_i}\right) = \frac{1}{1+e^{-v_i}} \tag{4.32}$$

where

$$v_i = ASC_{t(i)} + senToll_i \times scaledToll_i + senTime_i \times scaledTimeSaving_{t(i)} + b^{DPspeed} \times (s_{t(i)}^{DPspeed} - s_i^{histDPspeed}) + ISC_i \tag{4.25}$$

$$scaledToll_i = \frac{(p_i+1)^{b^{tollExponent}}-1}{b^{tollExponent}} \tag{4.33}$$

$\left\{ASC_{t(i)}, scaledTimeSaving_{t(i)}, s_{t(i)}^{DPspeed}\right\} = \boldsymbol{S_{t(i)}}$ is given by the DTA's supply model (4.34)

$\left\{senToll_i, senTime_i, s_i^{histDPspeed}, ISC_i\right\} = \boldsymbol{B_i}$ is drawn from the database (4.35)

$b^{tollExponent}$ and $b^{DPspeed}$ are constants with known values.

As the trip data do not include any traveler without a transponder, the choice model we estimate only represents travelers equipped with transponders. While non-transponder users may have different choice model parameters, given the limitation that their trips are not recorded, we apply the estimated model to the full population after calibrating the model to full-population data including aggregated flow and speed on the network. We calibrate selected parameters of the choice model offline to match the simulated ML market share to actual data.

4.5 Case Study

The real-time personalized toll pricing system is tested for the same real-world ML network that has been used in the previous chapter to test non-personalized toll pricing, which is referred to as the NTE network. However, the personalized choice model developed in this chapter requires disaggregate trip data so that model parameters can be estimated, and we do not have such data from the NTE network. We obtain trip data from the LBJ TEXpress Network to estimate the choice model. As the two networks are both physically close to each other and similar to each other, after estimating the choice model with LBJ network trip data, we calibrate model parameters with the flow, speed, and toll rate data from the NTE network and apply the calibrated model for testing of toll pricing on the NTE network.

4.5.1 Network and Data

Personalized toll pricing is tested for the NTE network, the same as the one used in the previous chapter where non-personalized toll pricing has been tested. The same flow, speed, and toll rate data are used to calibrate the DTA system and evaluate revenues.

We apply trip data from the LBJ network to estimate the personalized choice model. The LBJ network is a 13-mile corridor on I-635 and I-35E in Dallas, Texas. The corridor consists of ML and GPL that run in parallel with each other. We use the data of westbound trips on I-635 recorded by AVI sensors and toll gantries during a 5-month period from September 2017. These trips start before I-635's intersection with US-75 and continue beyond the intersection with I-35E, for a distance of around 10 miles.

The LBJ and the NTE networks are operated by the same operator and are in the same twin-cities, i.e., the Dallas-Fort Worth area. The two networks are apart from each other by only 15 miles. The NTE network consists of two tolling segments, while the LBJ westbound through-trips also traverse two tolling segments.

4.5.2 Personalized Choice Model

The ML operator provides us with trip record data that includes anonymized unique transponder IDs which we regard as traveler IDs, choice between ML and GPL, trip start time, actual travel time on the chosen paths, pre-choice expectations of travel times on both ML and GPL, the displayed toll rate, and the traffic condition (flow and speed) at the decision point. The ML operator estimated pre-choice expectations of travel time (i.e., expected travel time) based on traffic speed sensors along the corridor and then calibrated them to match the Google Maps predictions of travel time. Thus, we consider these variables as reasonable approximations of what travelers would expect the travel time to be before making the choice. While some travelers do not use Google Maps or other apps to obtain a real-time estimation of travel time and make the choice based on their estimations of the travel time, we assume the expected travel time in the data are fair approximations of their estimations. As toll rate changes every 5 minutes, and we measure traffic conditions for each 5-minute time interval, these variables are constant during the same 5-minute time intervals.

As the trips traverse through 2 tolling segments, the travelers do not have to stay on ML or GPL for the full trip. In fact, around 7% of the trips use partial ML. Those travelers either decided to take partial ML upfront or changed their minds and switched during the trip. At this stage we aim to develop a binary choice model, and thus we exclude any trip that does not stay fully on ML or GPL.

Besides, we identify time intervals when some of the AVI sensors are offline. During these time intervals, trips on some paths are not recorded, and thus trips in the data may be biased. We exclude all trips during those time intervals.

For each trip, we generate personalized variables using the same individuals' trip records in the past 90 days before that trip. Personalized variables refer to the set of are individual-specific variables, e.g., s^{freq}, s^{loyal}, s^{noTrip}. Trip data during the first 90 days are not directly used in model estimation, as a 90-day history is not available to generate personalized variables for these trips. We use data from Day 91 to Day 120 as training data to estimate model parameters, and data from Day 121 to Day 150 as testing data to test the prediction performance of the model. After generating personalized variables, we exclude weekend and holiday trips from training and testing datasets, as we aim to estimate a choice model for weekdays.

Estimation results of the choice model parameters are presented in the following tables. The star (*) besides some parameters indicate the parameter estimate is not significantly different from 0 at a significance level of 0.05. All times are measured in minutes (min) and tolls are measured in dollars ($). Speed is measured in mile-per-hour (mph).

Based on Equation 4.26 and the parameter estimates presented in Table 4-1, we plot the relationship between $\boldsymbol{ASC}$ and $24s^{TOD}$, the latter representing hours of the day.

Figure 4.2 shows that travelers tend to choose ML more during AM and PM peak periods compared to other periods when other components of the utility are equal. Two factors may explain this. Firstly, travelers during the peak period are more likely to be commuting travelers while during other periods, there are more leisure travelers. Commuting travelers may be more likely to choose ML compared to leisure travelers. As our model is purely based on revealed preferences data that do not include travel purpose variables, such effect is captured by the TOD-specific ASC. Secondly, during the peak periods, traffic flow and density along the corridor may be higher than in other periods, which is not captured in our model due to lack of data. Travelers choose ML not only to save time but also to enjoy a more comfortable driving experience due to lower density and more smooth movements on ML, and such effect is capture by the TOD-specific ASC.

Also, the ASC is lower than 0 in general, meaning that travelers prefer the GPL when other components of the utility are 0. This is understandable as the action to take ML may incur some disutility; for example, the payment itself is a disutility no matter what the toll rate is.

Table 4-1: Estimate of parameters related to ASC

Parameter	Estimate	Robust std error	Robust t-stat	
b^{ASC}	-1.13	0.0347	-32.57	
b_1^{sin}	-0.177	0.0273	-6.47	
b_2^{sin}	-0.257	0.0336	-7.63	
b_3^{sin}	-0.0658	0.0302	-2.18	
b_4^{sin}	0.0426	0.0272	1.57	*
b_5^{sin}	-0.0134	0.0233	-0.58	*
b_6^{sin}	0.0625	0.0176	3.56	
b_1^{cos}	-0.782	0.0382	-20.47	
b_2^{cos}	-0.533	0.0316	-16.85	
b_3^{cos}	0.374	0.0337	11.11	
b_4^{cos}	0.109	0.0259	4.19	
b_5^{cos}	-0.00745	0.0235	-0.32	*
b_6^{cos}	-0.0481	0.0183	-2.63	

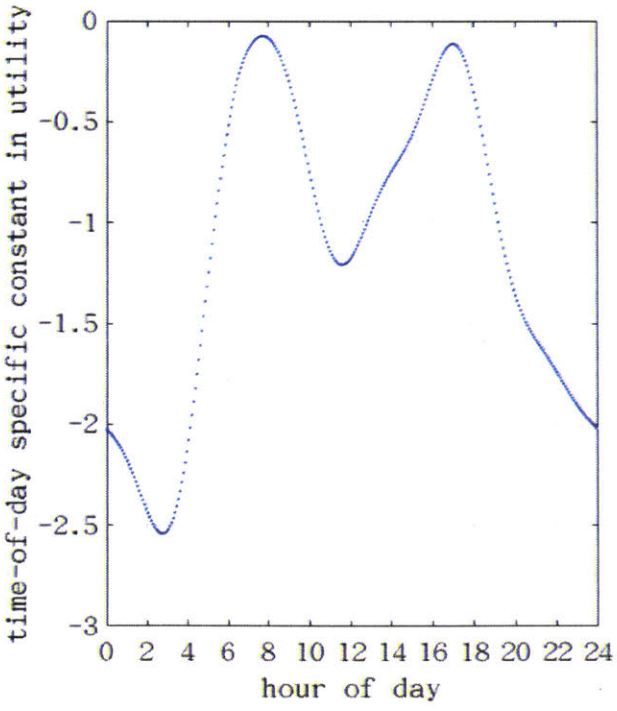

Figure 4.2: Time-of-day-specific ASC

Based on Equation 4.28, the relationship between $\boldsymbol{scaledToll}$ and $\boldsymbol{s^{toll}}$ is plotted in Figure 4.3. Table 4-2 presents the parameter estimates related to sensitivities to toll, travel time saving, and decision point (DP) traffic speed. $b^{tollExponent} = 3.68$ indicates toll has a highly non-linear effect in utility, and any increment of toll has a higher disutility when the toll is high compared to when it is low.

$b^{tollFreq} > 0$ and $b^{timeFreq} > 0$ indicate that frequent travelers are more sensitive to both toll and time-saving compared to infrequent travelers. This is reasonable as infrequent travelers may have a more imperfect understanding of the traffic conditions and make less rational decisions. $b^{tollLoyal} < 0$ and $b^{timeLoyal} < 0$ indicate that loyal travelers are less sensitive to both toll and time-saving, which means their decisions are less affected by real-time toll rates and travel time savings, but their past experiences have a relatively higher impact.

Sensitivities to time-savings are different at different periods of the day. During the morning peak, sensitivity is 1.25 times compared to mid-day. It is 1.17 times during PM peak and 1.64 times at night. Commuting travelers are more sensitive to time than leisure travelers, and thus, time sensitivity is higher during the commuting periods. Traveling during the night may incur more disutility per unit of time and travelers may be rushing to get home as early as possible, and thus the estimated sensitivity to time-saving is the highest among all periods of the day.

$b^{DPspeed} < 0$ indicates that travelers are less likely to choose ML when the traffic speed is higher at the decision point.

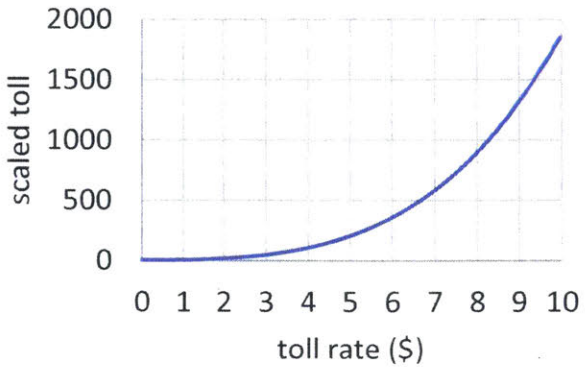

Figure 4.3: Relationship between toll rate and scaled toll

Table 4-2: Estimate of parameters related to sensitivities to toll and traffic

Parameter	Estimate	Robust std error	Robust t-stat	
$b^{tollExponent}$	3.68	0.504	7.31	
b^{toll}	-7.55	1.02	-7.43	
$b^{tollFreq}$	1.39	0.459	3.03	
$b^{tollLoyal}$	-0.412	0.095	-4.34	
$b^{tollNoTrip}$	-0.487	0.39	-1.25	*
b^{time}	-2.21	0.0934	-23.66	
$b^{timeFreq}$	0.958	0.21	4.56	
$b^{timeLoyal}$	-0.198	0.0403	-4.9	
$b^{timeNoTrip}$	-0.377	0.135	-2.8	
$b^{timeNight}$	0.496	0.141	3.52	
b^{timeAM}	0.225	0.108	2.09	
b^{timePM}	0.161	0.121	1.33	*
$b^{DPspeed}$	-0.0349	0.00247	-14.16	

Table 4-3 shows the estimates of parameters related to ISC. If a traveler has on average bad experience on ML in the past 90 days, he/she is less likely to choose ML. However, past experience on GPL does not have a significant impact. Preferences to ML are also correlated to the traveler's choices in the past 90 days, measured by loyalty. Besides, loyalty measured by the past 30 days' choices has an additional impact on choice.

Traveler's choice on the last (most recent) trip affects the current choice. The fixed component of this effect is not significant, and the variable component indicates that the traveler is more likely to choose ML if the last trip is on ML compared to not having a trip in the past 90 days. The traveler is less likely to choose ML if the last trip is on GPL compared to the same base. The effect diminishes as the last trip gets further away from the current trip, based on an exponential memory decay function, where the memory reduces by 12% per day or 58% per week. Experience on the last trip does not have a significant impact in addition to average experience on all past trips.

Table 4-3: Estimate of parameters related to ISC

Parameter	Estimate	Robust std error	Robust t-stat	
$b^{pastMLtimeDiff}$	-0.116	0.0286	-4.05	
$b^{pastGPLtimeDiff}$	-0.00396	0.00712	-0.56	*
b^{loyal}	0.782	0.0276	28.3	
$b^{loyal30}$	0.471	0.0297	15.88	
$b^{hadMLtrip}$	0.646	0.0952	6.78	
$b^{hadMLtrip30}$	-0.124	0.0445	-2.79	
$b^{hadGPLtrip}$	-0.574	0.0457	-12.57	
$b^{hadGPLtrip30}$	0.00431	0.0437	0.1	*
$b^{lastTripWasML}$	0.0409	0.0549	0.75	*
$b^{lastTripWasMLmemory}$	0.446	0.176	2.54	
$b^{lastTripWasGPLmemory}$	-0.392	0.0638	-6.14	
$b^{lastTripTimeDiffML}$	0.0125	0.0547	0.23	*
$b^{lastTripTimeDiffGPL}$	0.00584	0.0101	0.58	*
$b^{memoryDecayRate}$	0.124	0.0239	5.2	

The systematic heterogeneity captured by the model is presented in Table 4-4. Based on the estimated parameter values and the variables from the data, we calculate the mean and standard deviation (SD) of ***senToll***, ***senTime***, and ***ISC*** among all trips in the data. The coefficient of variation (CoV) is then calculated accordingly.

Table 4-4: Statistics measuring heterogeneity in travelers' sensitivities and preferences

	Mean	SD	CoV
senToll (per unit of ***scaledToll***)	-0.000736	0.000812	-1.10
senTime (per minute)	0.117	0.0583	0.499
ISC	-0.325	1.92	-5.92

Summary statistics of the model's performance on training data and testing data are presented in Table 4-5.

Table 4-5: Statistics of the model's performance on training data and testing data

	Training data	Testing data
# trips	120193	77872
# parameters	40	
log-likelihood	-39391	-25040
geomean of probabilities of chosen alternatives	72.06%	72.50%
accuracy of deterministic prediction	84.69%	85.77%

The log-likelihood of the estimated model is a measurement of how the model fits the data. The geometric mean probability of the chosen alternatives is simply a function of log-likelihood and the number of trips in the dataset. Accuracy of deterministic predictions measures the prediction accuracy when we predict each traveler chooses the alternative with higher systematic utility. Performance on the testing data seems better than training data, which may be a result of that testing data is of better quality than training data so that travelers' decisions are more correlated to the traffic conditions and their personal history variables.

4.5.3 Offline Calibration

As the choice model is estimated with transponder users' trip data from a particular OD of the LBJ network, and we are applying the model to testing personalized toll optimization on the full NTE network, we calibrate a subset of model parameters. The parameters we calibrate include the 13 parameters related to $\boldsymbol{ASC}$, which are b_k^{sin} and b_k^{cos} for all $k = 1, \ldots, 6$ and b^{ASC}, the exponent on toll $b^{tollExponent}$, and the overall toll sensitivity b^{toll}. Calibration is performed empirically to match the overall market share of ML at different times of the day and different toll levels. Overall market share is a weighted average of the market share at each gantry where the weight depends on toll rates. For gantries on the main corridor, market share is the flow on ML (i.e., gantry flow) divided by the total flow on ML and GPL. For gantries on entry ramps, market share is the gantry flow divided by total flow from the corresponding origin.

Figure 4.4 compares the simulated ML market share before and after offline calibration of the choice model parameters. Without calibration, the choice model overestimates market share when

the toll is low and during the night and underestimates market share when the toll is high and during the PM peak. After calibration, the market share profile matches better with actual data.

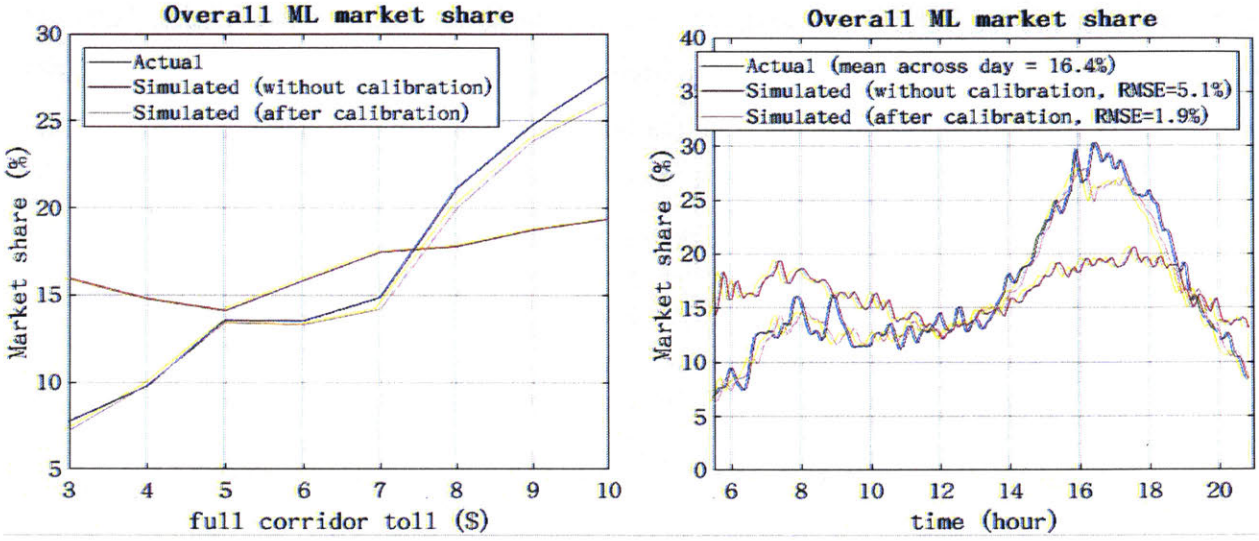

Figure 4.4: ML market share before and after choice model calibration

4.5.4 Toll Optimization

We evaluate the real-time personalized toll pricing system for Day 6 of the 9 days for which we have flow and speed data. The test is not conducted in the closed-loop evaluation framework, because the microscopic simulator MITSIM is not capable of simulating heterogeneous travelers with a personalized choice model or handling personalized toll rates for different travelers. While we are not using separate platforms to optimize and evaluate the toll as done in the closed-loop framework, we rely on separate modules of the system to optimize and evaluate the toll: the displayed toll rates and discount control parameters are optimized based on traffic predictions by the DTA system in the prediction-based optimization module, and we obtain personalized toll rates, revenue, and other metrics from simulations by the DTA system's state estimation module for evaluation purposes.

We conduct 4 tests, which represent 4 different ways to generate toll rates under the same demand and supply conditions (Table 4-6). Test A mimics a scenario when a non-personalized choice model is applied by the optimization module while actual travelers make choices based on a

personalized choice model. Non-personalized toll rates are optimized based on a non-personalized choice model but evaluated in traffic simulations where a personalized choice model is used. The percentage of eligible travelers to 0, meaning that no one is eligible to receive a discount. Test B is a test of non-personalized toll optimization based on the personalized choice model. Test C is for personalized toll optimization using the opportunity-cost-based discount policy, assuming 20% of the travelers are eligible for discounts. Test D is similar to Test C but with the reduced discount policy.

Table 4-6: Toll rate generation method in each test

Test ID	How toll is generated
A	Non-personalized toll optimization based on non-personalized choice model
B	Non-personalized toll optimization based on the personalized choice model
C	Personalized toll optimization (20% eligible to discount), using the opportunity-cost-based discount policy
D	Personalized toll optimization (20% eligible to discount), using the reduced discount policy

In Figure 4.5, we compare the optimized non-personalized toll rates between Test A and B. While different toll rates are optimized for the two tolling segments of the network, for easier comparison, the plot shows the full-corridor toll that is the sum of the two segments' toll rates. The upper bound of the full-corridor toll rate is $11. The figure shows that the optimized toll in Test B is much higher than Test A. Capturing the heterogeneities among travelers allows the system to predict expected revenue from each traveler separately. In Test A, the toll rate is optimized for an average traveler. In Test B, optimal toll rates for different travelers are different, and we define high-value customers as those whose optimal toll rates are higher, and low-value customers as those whose optimal toll rates are lower. The system makes a trade-off between a higher toll that generates higher revenue from high-value customers but lower revenue from low-value customers as the toll rates are much higher than optimal, or a lower toll that is far from the optimal value for high-value customers but generates more revenue from low-value customers. The results indicate that the system gives more weight to high-value customers and thus raises toll rates compared to when a non-personalized choice model is applied, possibly because the expected revenue from high-value customers is higher.

Note that, the optimized non-personalized toll rate is different from that in Chapter 3 because estimation and calibration methodology and data for the choice models are different. As revenues

are computed from traffic simulations using the corresponding choice models, it is not meaningful to compare the toll rates or revenues between Test 1 in this chapter and the result in Chapter 3.

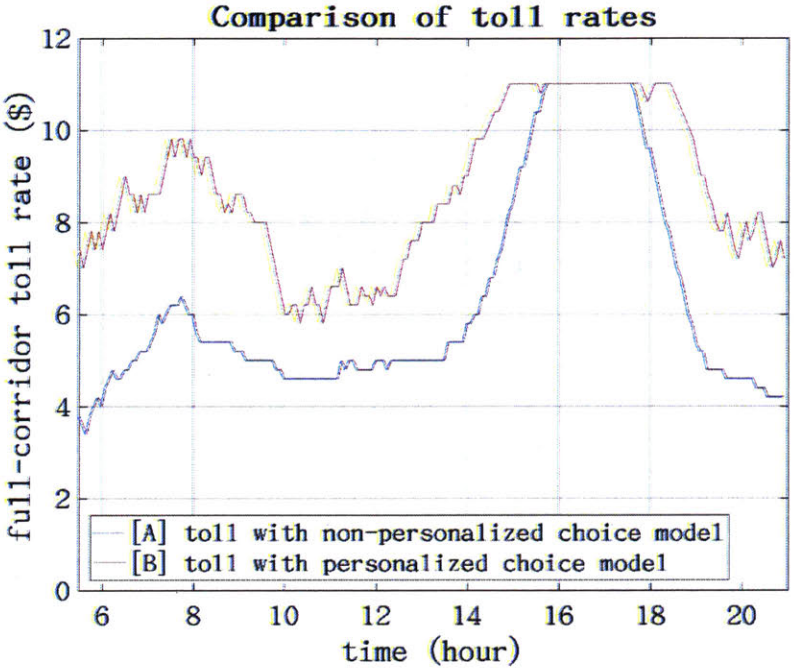

Figure 4.5: Comparison of toll rates between Test A and Test B

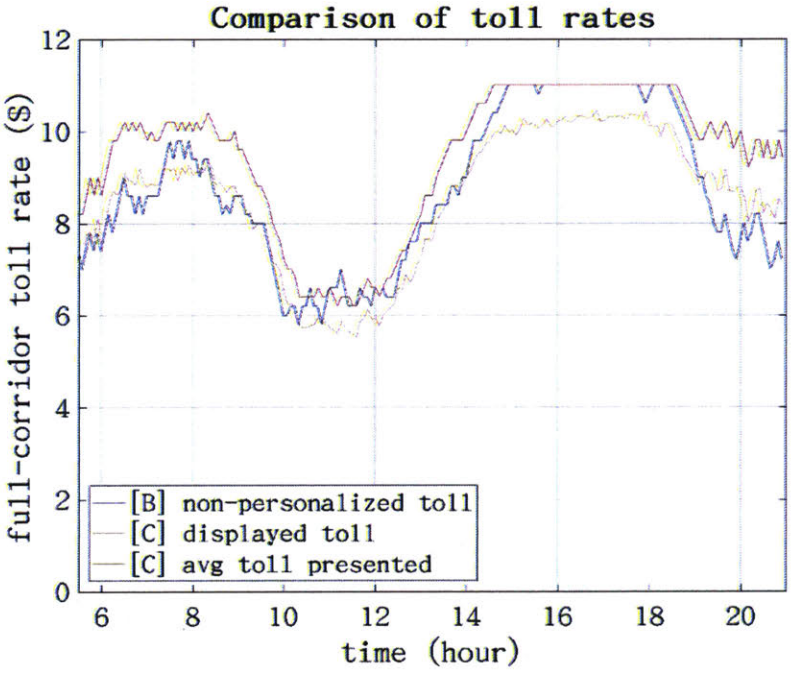

Figure 4.6: Comparison of toll rates between Test B and Test C

Figure 4.6 compares 3 curves: (1) the optimized non-personalized toll rate from Test B, (2) the optimized displayed toll rate from Test C, and (3) the average personalized toll rate among travelers from Test C.

As 20% of the travelers become eligible for discounts, a better strategy is to raise the displayed toll rate while offering discounts to some travelers. Therefore, curve (2) is, in general, higher than (1). The average toll rate (3) is generally close to the optimized toll when the toll is not personalized (1). However, in Test C, during the PM peak, the displayed toll (2) is constrained by the upper bound, but there are always some travelers for whom a discount would increase overall revenue. Therefore, the average toll (3) is lower than the upper bound while displayed toll (2) reaches the upper bound.

Figure 4.7 compares the optimized toll and the average toll under the two discount policies. The toll profile in Test D is very close to Test C. Under both discount policies, the personalized discount is generated by firstly finding the personal optimal price where elasticity is -1, and then apply some penalty to increase the price. The difference is at how the penalty is applied. Therefore, it is reasonable to see the two policies generating similar outcomes in terms of displayed toll and revenue, while the opportunity-cost-based policy is theoretically more efficient in maximizing revenue and thus leads to slightly higher revenue. We also note that in Test [D], more discounts are offered compared to Test [C].

Figure 4.8 compares the revenue in each test. We see that the improvement of expected revenue is consistent at each time interval, while in the case study of Chapter 3, it is not the case. In Chapter 3, we are evaluating the realized revenue in the microscopic simulator, which is subject to much higher stochasticity, mainly from supply-side simulations. In this Chapter, we are evaluating the expected revenue by the DTA system that is directly given by the choice model, and thus, the stochasticity is much lower. Besides, in Chapter 3 revenue improvement results from the system's better understanding and predictions of the traffic conditions, which may not be consistent in all time intervals, as online calibration is still not perfect. In this Chapter, revenue improvement is a result of more accurate modeling of the choice behaviors of the flexibility of charging different prices, and such benefits are consistent in all time intervals.

Figure 4.7: Comparison of toll rates between Test C and Test D

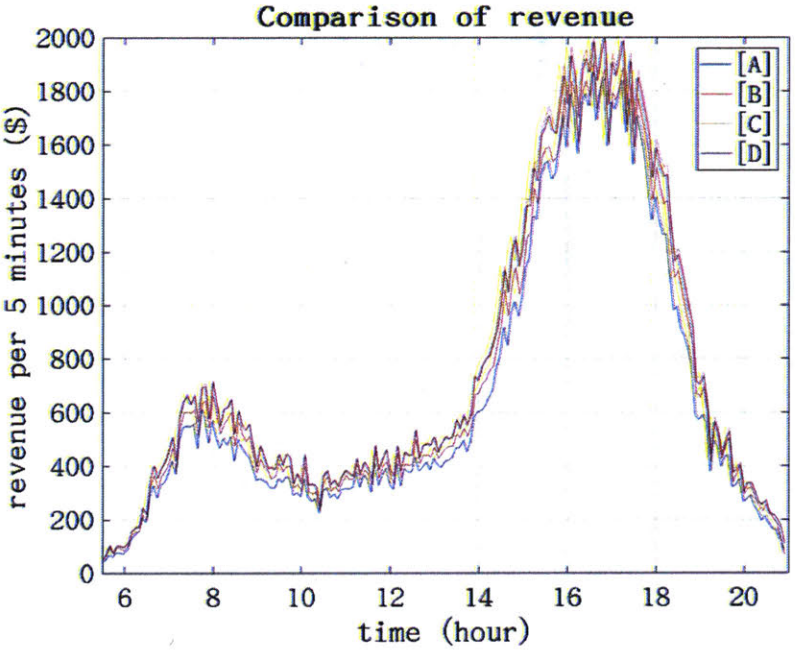

Figure 4.8: Comparison of revenue in each test

In Figure 4.9, we present the distribution of the toll rates presented to travelers in each test. All travelers are included to generate this distribution, whether or not he/she decides to pay the toll and take ML.

Figure 4.10 presents the distribution of the toll charged to travelers who decide to take ML. We see that in Test [A], the toll rate is generally lower, as discussed previously. Toll rates in Test [C] and [D] are lower than Test [B] due to discounting.

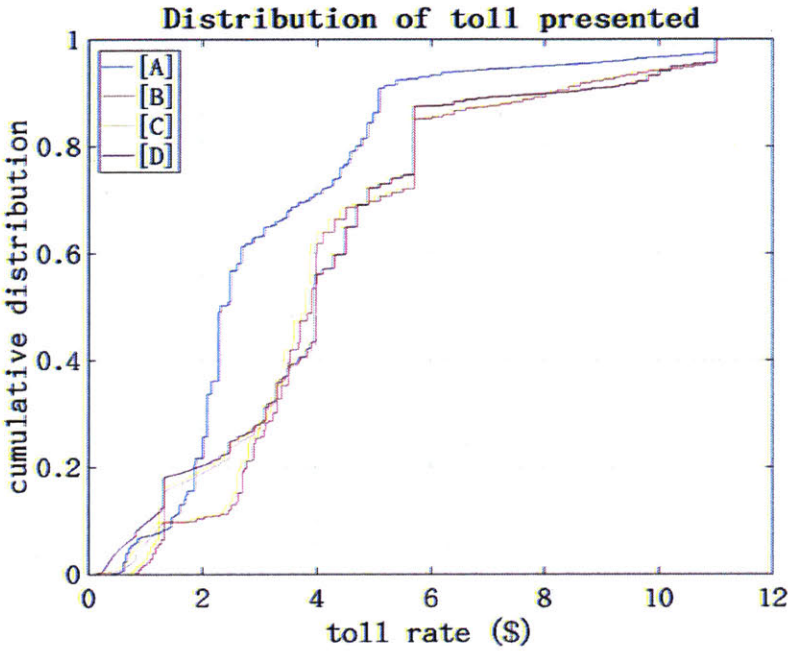

Figure 4.9: Distribution of the toll rates presented to travelers in each test

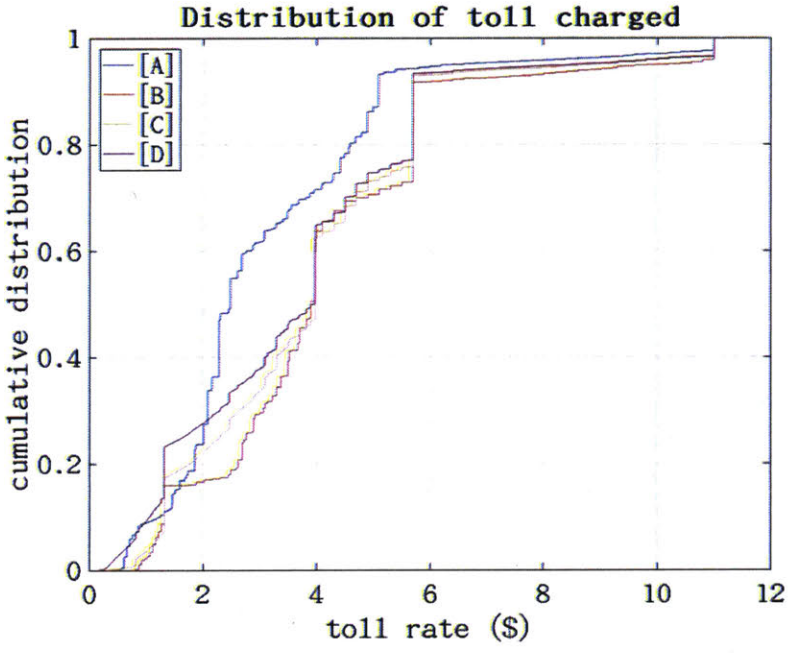

Figure 4.10: Distribution of the toll charged to ML users

Figure 4.11 shows the distribution of travel time. We see that shorter trips are less affected by the pricing strategies, while the travel time for longer trips differs more when the toll is different. As the corridor is only 13 miles long, in case there is not congestion, even full-corridor trips should take less than 13 minutes. A higher toll rate increases overall travel time mainly when there is congestion.

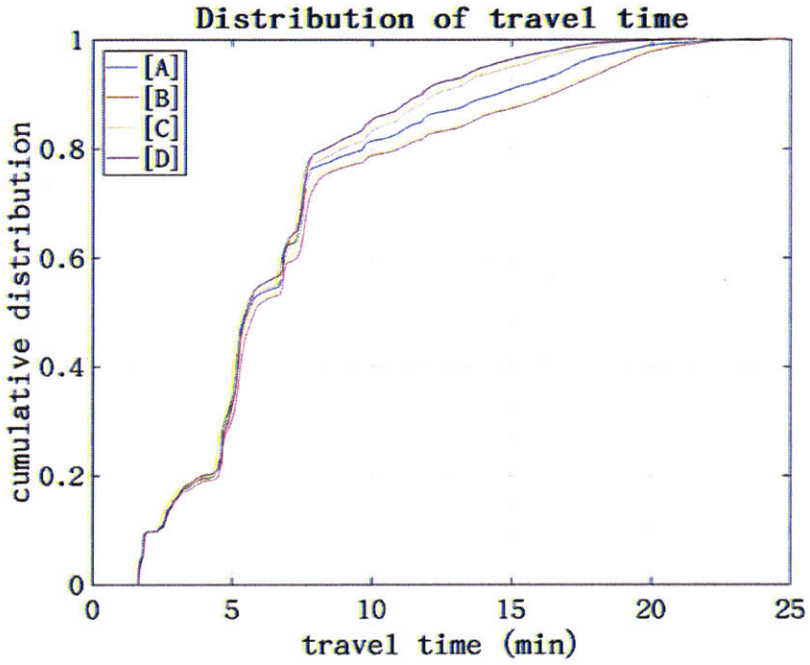

Figure 4.11: Distribution of travel time

Table 4-7 summarizes the results and discussions above. Test [B] improves revenue over Test [A] because a more accurate choice model is used in the optimization module. Test [C] and [D] further improves revenue as a result of personalized discounting, while the opportunity-cost-based discount policy in Test [C] generates a slightly higher revenue than the policy in Test [D].

We take Test [B] as the base for comparison. In test [A] more travelers are attracted to ML due to a much lower toll. In Test [C] and [D] travelers are attracted to ML due to the flexible pricing scheme, although the toll rate is only slightly lower than [B]. We see that the toll charged is slightly lower than the toll presented, as the travelers are less likely to choose ML when the toll is higher.

Table 4-7: Summary statistics of the four tests

Test ID	Total revenue (×$1000)	ML market share (%)	Average toll presented per trip ($)	Average toll cost per ML trip ($)	Average travel time per trip (min)
A	123	19.3	3.24	3.16	7.09
B	132	15.9	4.43	4.14	7.58
C	143	18.0	4.30	3.94	6.76
D	141	18.8	4.23	3.73	6.55

The results presented are based on a single test of the system. Due to stochasticity in traffic predictions by the DTA system, optimized toll rates and discount offers are not the actual optimal values and may be different in different realizations. In addition, the traffic simulation in which toll optimization is evaluated is also stochastic, and thus, the revenue improvement and other metrics may be different based on different realizations of the simulation.

4.6 Conclusion

This chapter presents a real-time personalized toll pricing system, which is an extension to the non-personalized toll pricing system, as presented in the previous chapter. We formulate personalized toll optimization as a two-level decision process, where the system optimization module optimizes system-level control parameters given the personalized pricing policy, and the personalized pricing module generates individual traveler's toll rate given optimized control parameters. While our formulation is capable of accommodating completely personalized toll pricing, we re-formulate it as personalized discounting for practical reasons, so system-level control parameters consist of the displayed toll rate and a discount control parameter. We derive an opportunity-cost-based discount policy so that the objective of personalized discounting is fully consistent with the system-level objective. It may be substituted by other policies with alternative objectives and/or constraints.

We extend the DTA-based solution framework proposed in the previous chapter to solve the personalized toll optimization problem in real time. An additional decision variable, i.e., the discount control parameter, is optimized in the prediction-based toll optimization module. The personalized discount module is implemented in the system.

Personalized pricing or discounting depends on the specific traveler's preference to ML and sensitivities to toll and time-saving. We develop a personalized choice model to predict each individual's probability of choosing ML and capture the systematic heterogeneity among individuals. The choice model specifies each individual's sensitivity to toll and time-saving as a function of their characteristics. As we rely on revealed preference data generated from AVI records to estimate the model, we do not have traveler characteristics directly available. We generate characteristics based on each traveler's personal history in the past 90 days. The personal history variables we generate include loyalty to ML, travel frequency, past experience, variables related to the most recent trip, and a number of dummy variables. While 3 variables are used to capture the heterogeneity of sensitivities, we use more variables to capture heterogeneity in travelers' overall preferences to ML. We estimate model parameters with trip data generated from AVI record on a real ML network, and the results indicate there is significant heterogeneity in both sensitivities and over preferences. The estimated model is applied in the toll pricing system after calibration.

We test the personalized toll pricing system for the same ML network used in the previous chapter. Based on one replication of the test for each scenario, we find that providing personalized discounts to some travelers increase overall revenue compared to the non-personalized tolling scheme. While displayed toll rate becomes higher, the average toll rate presented to all travelers (including those who did not choose ML) stays roughly the same as the non-personalized tolling scheme.

From the analysis of individuals' toll rates and travel times, we find that the operator's interest is not always consistent with the travelers. If the operator applies the revenue-maximizing toll rate, then reducing the toll rate may promote social welfare as overall travel time reduces, although it is not to the best interest of the operator. Note that, this is only true when the ML is not congested, which is the case in our tests. This finding is consistent with that in Chapter 3, and thus, the same recommendations could be made to the regulators.

On the other hand, we find that personalized discounting is beneficial for both the operator and the travelers when ML is not congested. By personalized discounting, the ML operator can attract more travelers to the ML. While displayed toll rates increase, average toll rates slightly decrease,

and thus travelers benefit from paying an overall lower toll. As the overall toll is lower, personalized discounting also reduces traffic on the GPL and reduces travelers' average travel time.

It should be noted that the results are based on the choice model we develop, estimate, and calibrate. We find that travelers, in general, prefer ML more during the AM and PM peak periods, and thus toll rate and simulated revenues during those periods are both higher. In addition, congestions on GPL during the PM peak also contributes to higher toll and revenue during that time. Personalized toll prices for those who are insensitive to toll and have a large ISC are higher than other travelers, and in the meantime, expected revenue from travelers with larger ISC is higher.

Different outcomes are expected if personalized toll pricing is applied to a different network with a population that is subject to a different choice model. For example, the distribution of demand and ASC at different times of the day affect the toll profile, i.e., the periods when the toll is higher/lower. A lower toll is justified for a population more sensitive to toll. In case the population is relatively homogeneous in choice behavior, the impact of personalized pricing/discounting would be minimal.

Chapter 5 Conclusions

In this chapter, we first identify the research contributions of this study and then summarize the main findings. Finally, we discuss the limitations of this research and potential future research directions.

5.1 Contributions

This research contributes to the literature on road pricing by developing a DTA-based proactive toll pricing system that is feasible to be applied in real time. It is generally agreed that a proactive pricing scheme is better than a reactive scheme in terms of optimizing a specific objective that is measured in the future, as long as predictions of the future are accurate, but an integrated system that addresses real-time performance and prediction accuracy is no available in the literature.

We contribute to the online calibration of simulation-based DTA systems by developing a heuristic that calibrates DTA supply model parameters to match simulated speed measurements with real-time sensor data and thus predictions of traffic speed. Current methodologies for online calibration mainly focus on matching traffic flow measurements while not emphasizing the importance of traffic speed. While traffic flow usually matches true measurements with satisfactory accuracy, the literature rarely shows whether congestions can be captured by the DTA system through online calibration. Accurate predictions of traffic speed and thus travel time is crucial for the success of a proactive pricing scheme, as its performance relies on accurate evaluation of the objective, which is based on accurate predictions of future traffic conditions. Travel time is an important measurement of traffic conditions.

We contribute to the literature in personalized pricing by designing a dynamic personalized pricing policy for application in the field of road pricing. We combine the elasticity-based personalized pricing strategy so that each individual is priced differently to maximize expected revenue, and the

opportunity-cost-based dynamic pricing strategy so that the price of a product changes dynamically as the value of the inventory (or opportunity cost of selling the product) changes dynamically. While such combined strategies exist in the literature on dynamic personalized pricing, the value of inventory (opportunity cost) is derived analytically or empirically for applications in different fields. In the field of road pricing, the value of road capacity (opportunity cost) is not readily available in an analytical form, and we innovatively apply a simulation-based optimization framework to evaluate the opportunity cost.

We then develop a real-time personalized toll pricing framework, which is new in the field of road pricing. We design a two-level framework so that a personalized pricing policy is applied at the lower-level, while its control parameters are optimized at the upper-level.

Elasticity-based pricing relies on a personalized behavior model that captures the heterogeneity in different travelers' choice behaviors. We develop a personalized choice model that relies on travelers' characteristics to capture travelers' heterogeneous preferences to managed lanes as well as heterogeneous sensitivities to toll and travel time. It is known that a choice model estimated with revealed preference (RP) data is more accurate in representing behaviors in the real world compared to stated preference (SP) data. While SP data usually contains travelers' characteristics, RP data generated from AVI sensors on the real network generally do not capture traveler's socioeconomic characteristics. We innovatively propose a set of personal history variables that captures travelers' characteristics based on their travel history in the past.

While our methodologies are applied to an ML network, it can be applied to a general road network with modest extensions. While we chose revenue as the objective of the optimization formulation, our methodologies are capable of accommodating different objectives, for example, social welfare. In such cases, the opportunity cost would be defined differently with regard to the objective.

5.2 Findings

We perform a single test of online calibration for each of the 8 days and find the heuristic online calibration methodology that calibrates speed-density relationship is effective in improving the accuracy of simulated speed. Proactive toll pricing based on traffic predictions by simulation-based DTA is feasible to be applied in the real world in real time. While effective optimization relies on accurate online calibration, the online calibration methodology we develop is capable of supporting the system's good performance. The proposed toll pricing system performs better than a time-of-day toll profile which is not adaptive to real-time traffic conditions, based on our closed-loop evaluation of the toll pricing system. Toll optimization that relied on online calibration performs even better when congestions are significant, and vice versa.

We estimate the choice model based on trip data generated from AVI sensors on a real network and find that the personal history variables we generate are capable of capturing the systematic heterogeneity among travelers, and therefore a personalized route choice model can be estimated with purely AVI trip data. We then test the personalized toll pricing system based on the personalized choice model we have estimated and find that given the heterogeneity we capture, offering discounts to some travelers while slightly raising displayed toll rates increase overall revenue, while only slightly increasing the dimension of the optimization problem. We also find personalized pricing reduces the average toll paid by travelers and average travel time, but given the choice model parameters we applied, travelers would benefit from lower cost and lower travel time if the ML operator charges a toll lower than the rate where revenue is maximized. Therefore, regulatory agencies have the potential to improve social welfare through subsidizing the use of ML thus altering ML operators' tolling objective, or penalizing GPL congestions, in addition to simply imposing constraints on toll rates and traffic conditions.

In this thesis, all results regarding toll optimization are based on our simulations that depend on the models and parameters we apply and are not compared to any actual data from the real network. All tests we perform include only one replication for each scenario, and thus, the results are subject to stochasticity.

5.3 Future Research Directions

We suggest the following future research directions, either to improve certain aspects of this study or to extend the toll pricing framework developed in this study.

(1) In the field of online calibration of simulation-based DTA systems, the current research focus is on the simultaneous calibration of OD, choice model, and supply model parameters to match the full set of measurements including flow and speed. Future achievements in this field may improve the calibration and prediction accuracy, thus improve toll optimization performances.

(2) We use a grid search algorithm in the system optimization module to find the toll rates and discount control parameters that maximize the objective. This algorithm generally does not find the actual optima. A better solution algorithm may be applied to improve the optimization solution.

(3) While the DTA system calibrates its parameters to fit current traffic conditions, its traffic prediction always includes errors as the predictions of demand and supply parameters are never accurate. Robust optimization formulations may be applied to account for errors in traffic predictions. This is especially important when the toll optimization problem contains constraints on future traffic conditions which may not be violated.

(4) More comprehensive choice models may be developed. Firstly, a mixed logit model may be applied to capture random heterogeneities among individuals. Besides, a multinomial logit model may be used so that the behavior of using partial ML and partial GPL in a trip may be modeled. Furthermore, choice model parameters may be calibrated online based on real-time flow and speed data to improve the accuracy of traffic predictions.

(5) More dimensions of travelers' behavior in addition to route choice may be modeled. For example, the displayed toll rate or a discount offer may change a traveler's decision whether to travel or when to travel. The toll optimization then needs to be extended to capture such behaviors.

(6) For personalized toll pricing, alternative discount policies may be tested. The discount policy may have a different objective from the system-level objective, or contain additional constraints.

Specifically, a policy not based on travelers' price elasticity of choice probability may attract better public acceptance. Group pricing may also be tested, where discount policy is purely based on characteristics. An example is to offer discounts to those who are loyal to ML in the past 90 days while not loyal in the past 30 days.

(7) More extensive testing and analysis of the proposed toll pricing system may be necessary. In this thesis, each scenario is evaluated with a single test. Due to the stochasticity in the traffic predictions by the DTA system, optimized decision variables are in general not optimal and may be different in different realizations. Due to stochasticity in the simulations where toll optimization is evaluated, the metrics that demonstrate the results are also subject to stochasticity. Additional testing may be necessary to evaluate the statistical significance of the impacts of the proposed toll pricing system. Base on the tests, additional analysis may be performed to evaluate the societal impact of different tolling strategies. While this thesis evaluates the travel time and toll cost to travelers, additional metrics may be included in the analysis.

Bibliography

Adnan, M., Pereira, F.C., Azevedo, C.M.L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C. and Ben-Akiva, M., 2016, January. SimMobility: A multi-scale integrated agent-based simulation platform. In *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record.*

Ahn, H.S., Gümüş, M. and Kaminsky, P., 2007. Pricing and manufacturing decisions when demand is a function of prices in multiple periods. *Operations Research*, *55*(6), pp.1039-1057.

Antoniou, C., Azevedo, C.L., Lu, L., Pereira, F. and Ben-Akiva, M., 2015. W–SPSA in practice: Approximation of weight matrices and calibration of traffic simulation models. *Transportation Research Procedia*, *7*, pp.233-253.

Antoniou, C., Ben-Akiva, M. and Koutsopoulos, H.N., 2007. Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. *IEEE Transactions on Intelligent Transportation Systems*, *8*(4), pp.661-670.

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*, p.0361198119864906.

Ashok, K. and Ben-Akiva, M.E., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. *Transportation Science*, *34*(1), pp.21-36.

Atasoy, B., Ikeda, T., Song, X. and Ben-Akiva, M.E., 2015. The concept and impact analysis of a flexible mobility on demand system. *Transportation Research Part C: Emerging Technologies*, *56*, pp.373-392.

Aviv, Y. and Pazgal, A., 2008. Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management*, *10*(3), pp.339-359.

Aydin, G. and Ziya, S., 2008. Pricing promotional products under upselling. *Manufacturing & Service Operations Management*, *10*(3), pp.360-376.

Aydin, G. and Ziya, S., 2009. Personalized dynamic pricing of limited inventories. *Operations Research*, *57*(6), pp.1523-1531.

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 *the 97th Annual Meeting of Transportation Research Board.*

Balakrishna, R., 2006. *Off-line calibration of dynamic traffic assignment models* (Dissertation, Massachusetts Institute of Technology).

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N. and Wen, Y., 2007. Calibration of microscopic traffic simulation models: Methods and application. *Transportation Research Record, 1999*(1), pp.198-207.

Balakrishna, R. and Koutsopoulos, H.N., 2008. Incorporating within-day transitions in simultaneous offline estimation of dynamic origin-destination flows without assignment matrices. *Transportation Research Record, 2085*(1), pp.31-38.

Balakrishna, R., Koutsopoulos, H.N. and Ben-Akiva, M., 2005. Calibration and validation of dynamic traffic assignment systems. In *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic TheoryUniversity of Maryland, College Park.*

Becker, F., Danaf, M., Song, X., Atasoy, B., and Ben-Akiva, M., 2018. Bayesian estimator for Logit Mixtures with inter-and intra-consumer heterogeneity. *Transportation Research Part B: Methodological, 117*, pp.1-17.

Ben-Akiva, M., Bolduc, D. and Walker, J., 2001. Specification, identification and estimation of the logit kernel (or continuous mixed logit) model.

Ben-Akiva, M., Gao, S., Lu, L. and Wen, Y., 2015. DTA2012 Symposium: combining disaggregate route choice estimation with aggregate calibration of a dynamic traffic assignment model. *Networks and Spatial Economics, 15*(3), pp.559-581.

Ben-Akiva, M., Koutsopoulos, H.N., Antoniou, C. and Balakrishna, R., 2010. Traffic simulation with dynamit. In *Fundamentals of traffic simulation* (pp. 363-398). Springer, New York, NY.

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

Ben-Akiva, M., McFadden, D., and Train, K., 2019. Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. *Foundations and Trends® in Econometrics, 10*(1-2), pp.1-144.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D.S. and Daly, A., 2002. Hybrid choice models: progress and challenges. *Marketing Letters, 13*(3), pp.163-175.

Bhat, C. R., and Castelar, S. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B: Methodological*, 2002. 36(7), 593-616.

Börjesson, M., Eliasson, J. and Levander, A., 2007. The value of time of car drivers choosing route: evidence from the Stockholm congestion charging trial. In *2007 European Transport Conference, Leiden, Netherlands* (p. 14).

Brownstone, D., Ghosh, A., Golob, T.F., Kazimi, C. and Van Amelsfort, D., 2003. Drivers' willingness-to-pay to reduce travel time: evidence from the San Diego I-15 congestion pricing project. *Transportation Research Part A: Policy and Practice, 37*(4), pp.373-387.

Burris, M.W. and Ashraf, S., 2019. Tracking the Impact of a Toll Increase on Managed Lane Travel Behavior. *Transportation Research Record*, *2673*(2), pp.779-788.

Burris, M.W. and Brady, J.F., 2018. Unrevealed Preferences: Unexpected Traveler Response to Pricing on Managed Lanes. *Transportation Research Record*, *2672*(5), pp.23-32.

Cantelmo, G., Cipriani, E., Gemma, A. and Nigro, M., 2014. An adaptive bi-level gradient procedure for the estimation of dynamic traffic demand. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), pp.1348-1361.

Cascetta, E. and Postorino, M.N., 2001. Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation science*, *35*(2), pp.134-147.

Castells, P., Hurley, N.J. and Vargas, S., 2015. Novelty and diversity in recommender systems. In *Recommender Systems Handbook* (pp. 881-918). Springer, Boston, MA.

Chapuis, J.M., 2007. Basics of dynamic programming for revenue management. *Available at SSRN 1123768*.

Chen, L. and Wu, C., 2018. Bayesian dynamic pricing with unknown customer willingness-to-pay and limited inventory. *Available at SSRN 2689924*.

Chen, M. and Chen, Z.L., 2015. Recent developments in dynamic pricing research: multiple products, competition, and limited demand information. *Production and Operations Management*, *24*(5), pp.704-731.

Chen, X., Owen, Z., Pixton, C. and Simchi-Levi, D., 2015. A statistical learning approach to personalization in revenue management. *Available at SSRN 2579462*.

Chen, X., Zhang, L., He, X., Xiong, C. and Li, Z., 2014. Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network. *Computer-Aided Civil and Infrastructure Engineering*, *29*(5), pp.359-381.

Chung, C.L. and Recker, W., 2011. *State-of-the-art assessment of toll rates for high-occupancy and toll lanes* (No. 11-1331).

Cipriani, E., Florian, M., Mahut, M. and Nigro, M., 2011. A gradient approximation approach for adjusting temporal origin–destination matrices. *Transportation Research Part C: Emerging Technologies*, *19*(2), pp.270-282.

Danaf, M., 2017, March. Personalized Recommendations using Discrete Choice Models with Inter-and Intra-Consumer Heterogeneity. In *International Choice Modelling Conference 2017*.

Danielis, R., Marcucci, E. and Rotaris, L., 2005. Logistics managers' stated preferences for freight service attributes. *Transportation Research Part E: Logistics and Transportation Review*, *41*(3), pp.201-215.

Darda, D., 2002. *Joint calibration of a microscope traffic simulator and estimation of origin-destination flows* (Dissertation, Massachusetts Institute of Technology).

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

Djukic, T., 2014. Dynamic OD demand estimation and prediction for dynamic traffic management.

Djukic, T., Van Lint, J.W.C. and Hoogendoorn, S.P., 2012. Application of Principal Component Analysis to Predict Dynamic Origin–Destination Matrices. *Transportation Research Record*, *2283*(1), pp.81-89.

Dong, J., Mahmassani, H.S., Erdoğan, S., and Lu, C.C., 2011. State-dependent pricing for real-time freeway management: Anticipatory versus reactive strategies. *Transportation Research Part C: Emerging Technologies*, *19*(4), pp.644-657.

Elmachtoub, A.N., Gupta, V. and Hamilton, M., 2018. The value of personalized pricing. *Available at SSRN 3127719*.

Federal Highway Administration, 2017. 2017 Urban Congestion Trends: Measuring, Managing, and Improving Operations in the 21st Century. Available at: https://ops.fhwa.dot.gov/publications/fhwahop18025/index.htm. Accessed Aug. 1, 2019.

Feng, T., Arentze, T. and Timmermans, H., 2013. Capturing preference heterogeneity of truck drivers' route choice behavior with context effects using a latent class model. *European Journal of Transport and Infrastructure Research*, *13*(4).

Fowkes, T. and Whiteing, T., 2006. The value of freight travel time savings and reliability improvements–Recent evidence from Great Britain. In *Proc., European Transport Conference 2006*.

Frederix, R., Viti, F., Himpe, W.W. and Tampère, C.M., 2014. Dynamic origin–destination matrix estimation on large-scale congested networks using a hierarchical decomposition scheme. *Journal of Intelligent Transportation Systems*, *18*(1), pp.51-66.

Fudenberg, D. and Villas-Boas, J.M., 2006. Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, *1*, pp.377-436.

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), pp.94-101.

Gupta, S., Seshadri, R., Atasoy, B., Pereira, F.C., Wang, S., Vu, V.A., Tan, G., Dong, W., Lu, Y., Antoniou, C. and Ben-Akiva, M., 2016. *Real time optimization of network control strategies in dynamit2. 0* (No. 16-5560).

Hashemi, H., and Abdelghany, K., 2015. *Integrated method for online calibration of real-time traffic network management systems* (No. 15-3240).

Hashemi, H., and Abdelghany, K., 2016. Real-time traffic network state estimation and prediction with decision support capabilities: Application to integrated corridor management. *Transportation Research Part C: Emerging Technologies*, *73*, pp.128-146.

Hazelton, M.L., 2000. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, *34*(7), pp.549-566.

Hess, S., and Rose, J.M., 2009. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. *Transportation Research Part B: Methodological*, *43*(6), pp.708-719.

Hess, S. and Train, K.E., 2011. Recovery of inter-and intra-personal heterogeneity using mixed logit models. *Transportation Research Part B: Methodological*, *45*(7), pp.973-990.

INRIX, 2019. INRIX 2018 Global Traffic Scorecard. Available at: http://inrix.com/scorecard/. Accessed Aug. 1, 2019.

Jang, K., Chung, K., and Yeo, H., 2014. A dynamic pricing strategy for high occupancy toll lanes. *Transportation Research Part A: Policy and Practice*, *67*, pp.69-80.

Kawamura, K., 2000. Perceived value of time for truck operators. *Transportation Research Record*, *1725*(1), pp.31-36.

Kim, K.O. and Rilett, L.R., 2004, January. A genetic algorithm based approach to traffic micro-simulation calibration using ITS data. In *83rd Annual Meeting of the Transportation Research Board, Washington, DC*.

Kundé, K.K., 2002. *Calibration of mesoscopic traffic simulation models for dynamic traffic assignment* (Dissertation, Massachusetts Institute of Technology).

Kuo, C.W., Ahn, H.S. and Aydin, G., 2011. Dynamic pricing of limited inventories when customers negotiate. *Operations research*, *59*(4), pp.882-897.

Kurri J., Sirkia A. and Mikola J. (2000), Value of Time in Freight Transport in Finland, Transportation Research Record 1725, pp.26-30.

Leclercq, L., 2005. Calibration of flow–density relationships on urban streets. *Transportation Research Record*, *1934*(1), pp.226-234.

Li, D., Miwa, T., Morikawa, T., Liu, P. Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets. *Transportation Research Part C: Emerging Technologies*, 2016. 67, 31-46.

Li, Z. and Hensher, D.A., 2010. Toll roads in Australia: an overview of characteristics and accuracy of demand forecasts. *Transport Reviews*, *30*(5), pp.541-569.

Lu, L., 2013. *W-SPSA: an Efficient Stochastic Approximation Algorithm for the off-line calibration of Dynamic Traffic Assignment models* (Master Thesis, Massachusetts Institute of Technology).

Lu, L., Xu, Y., Antoniou, C., and Ben-Akiva, M., 2015. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transportation Research Part C: Emerging Technologies*, *51*, pp.149-166.

Ma, J., Dong, H. and Zhang, H.M., 2007. Calibration of microsimulation with heuristic optimization methods. *Transportation Research Record*, *1999*(1), pp.208-217.

Marzano, V., Papola, A., Simonelli, F. and Papageorgiou, M., 2018. A Kalman filter for quasi-dynamic od flow estimation/updating. *IEEE Transactions on Intelligent Transportation Systems*, *19*(11), pp.3604-3612.

McFadden, D., 1975. The revealed preferences of a government bureaucracy: Theory. Bell Journal of Economics, 6(2), pp.401-416.

Meyer, M.D., 1999. Demand management as an element of transportation policy: using carrots and sticks to influence travel behavior. *Transportation Research Part A: Policy and Practice*, *33*(7-8), pp.575-599.

Michalaka, D., Lou, Y. and Yin, Y., 2011. *Proactive and robust dynamic pricing strategies for high-occupancy-toll (HOT) lanes* (No. 11-2617).

Morgul, E.F., 2010. *Simulation based evaluation of dynamic congestion pricing algorithms and strategies* (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).

Murthi, B.P.S. and Sarkar, S., 2003. The role of the management sciences in research on personalization. *Management Science*, *49*(10), pp.1344-1362.

Nagae, T. and Akamatsu, T., 2006. Dynamic revenue management of a toll road project under transportation demand uncertainty. *Networks and Spatial Economics*, *6*(3-4), pp.345-357.

Netessine, S., Savin, S. and Xiao, W., 2006. Revenue management through dynamic cross selling in e-commerce retailing. *Operations Research*, *54*(5), pp.893-913.

Oh, S., Seshadri, R., Azevedo, C.L. and Ben-Akiva, M.E., 2019. Demand calibration of multimodal microscopic traffic simulation using weighted discrete SPSA. *Transportation Research Record*, p.0361198119842107.

Osorio, C., 2017. High-dimensional offline od calibration for stochastic traffic simulators of large-scale urban networks. In *Technical Report*. Massachusetts Institute of Technology.

Osorio, C., 2019. Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C: Emerging Technologies*, *98*, pp.186-206.

Park, B.B., Pampati, D.M. and Balakrishna, R., 2006. Architecture for on-line deployment of DynaMIT in hampton roads, VA. In *Applications of Advanced Technology in Transportation* (pp. 605-610).

Paz, A., Molano, V. and Gaviria, C., 2012, September. Calibration of corsim models considering all model parameters simultaneously. In *2012 15th International IEEE Conference on Intelligent Transportation Systems* (pp. 1417-1422). IEEE.

Peeta, S. and Ziliaskopoulos, A.K., 2001. Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and spatial economics*, *1*(3-4), pp.233-265.

Pigou, A., 2017. *The economics of welfare*. Routledge.

Popescu, I. and Wu, Y., 2007. Dynamic pricing strategies with reference effects. *Operations research*, *55*(3), pp.413-429.

Prakash, A.A., Seshadri, R., Antoniou, C., Pereira, F.C. and Ben-Akiva, M.E., 2017. Reducing the dimension of online calibration in dynamic traffic assignment systems. *Transportation Research Record*, *2667*(1), pp.96-107.

Prakash, A.A., Seshadri, R., Antoniou, C., Pereira, F.C., and Ben-Akiva, M., 2018. Improving scalability of generic online calibration for real-time dynamic traffic assignment systems. *Transportation Research Record*, p.0361198118791360.

Qurashi, M., Ma, T., Chaniotakis, E. and Antoniou, C., 2019. PC-SPSA: Employing Dimensionality Reduction to Limit SPSA Search Noise in DTA Model Calibration. *IEEE Transactions on Intelligent Transportation Systems*.

Ramming, M.S., 2001. Network knowledge and route choice. *Unpublished Ph.D. Thesis, Massachusetts Institute of Technology*.

Rowell, M., Gagliano, A. and Goodchild, A., 2014. Identifying truck route choice priorities: the implications for travel models. *Transportation Letters*, *6*(2), pp.98-106.

Saleh, W., and Sammer, G., 2016. Travel demand management and road user pricing: success, failure, and feasibility. In *Travel Demand Management and Road User Pricing* (pp. 21-30). Routledge.

Schrank, D., Eisele, B., Lomax, T. and Bak, J., 2015. 2015 urban mobility scorecard.

Shiller, B.R., 2013. *First degree price discrimination using big data* (p. 32). Brandeis Univ., Department of Economics.

Small, K.A., Winston, C. and Yan, J., 2005. Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, *73*(4), pp.1367-1382.

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), pp.599-607.

Spall, J.C., 1992. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE transactions on automatic control*, *37*(3), pp.332-341.

Suh, M., 2010. *Retail Pricing of Substitutable Products under Logit Demand* (Doctoral dissertation).

Sun, Y., Xiong C., Zhang L., 2018. Designing Dynamic and Personalized Travel Incentives for Transportation Network Efficiency. Submitted for presentation only at COTA International Symposium Emerging Trends in Transportation

Supernak, J., Steffey, D. and Kaschade, C., 2003. Dynamic value pricing as instrument for better utilization of high-occupancy toll lanes: San Diego I-15 case. *Transportation Research Record*, *1839*(1), pp.55-64.

Systematics, C., 2006. Inc. I-394 MnPASS technical evaluation: Final report. *Minnesota Department of Transportation, St. Paul.*

Talluri, K. and Van Ryzin, G., 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, *50*(1), pp.15-33.

Taylor, R., 2010. *2009 Urban Congestion Trends: How Operations Is Solving Congestion Problems* (No. FHWA-HOP-10-032). United States. Federal Highway Administration.

Toledo, T., Ben-Akiva, M.E., Darda, D., Jha, M. and Koutsopoulos, H.N., 2004. Calibration of microscopic traffic simulation models with aggregate data. *Transportation Research Record*, *1876*(1), pp.10-19.

Toledo, T., Jing, P., Atasoy, B., Ding-Mastera, J., Santos, J.O. and Ben-Akiva, M., 2018. *Intercity Truck Driver Route Choice Incorporating Drivers' Heterogeneity in Toll Road Usage: Data Collection, Model Estimation, and Model Application* (No. 18-02626).

Toledo, T., Kolechkina, T., Wagner, P., Ciuffo, B., Azevedo, C., Marzano, V. and Flötteröd, G., 2015. Network model calibration studies. *Daamen W., Buisson C. and S. Hoogendoorn (eds). Traffic simulation and data: validation methods and applications, CRC Press, Taylor and Francis, London*, pp.141-162.

Toledo, T. and Sharif, S., 2019. The effect of information on drivers' toll lane choices and travel times expectations. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, pp.149-159.

Toledo, T., Sun, Y., Rosa, K., Ben-Akiva, M., Flanagan, K., Sanchez, R. and Spissu, E., 2013. Decision-making process and factors affecting truck routing. In *Freight Transport Modelling* (pp. 233-249). Emerald Group Publishing Limited.

Train, K.E., 2009. *Discrete choice methods with simulation*. Cambridge university press.

Tsekeris, T. and Voß, S., 2009. Design and evaluation of road pricing: state-of-the-art and methodological advances. *NETNOMICS: Economic Research and Electronic Networking*, *10*(1), pp.5-52.

Tympakianaki, A., Koutsopoulos, H.N. and Jenelius, E., 2015. c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin–destination matrix estimation. *Transportation Research Part C: Emerging Technologies*, *55*, pp.231-245.

Vaze, V., Antoniou, C., Wen, Y. and Ben-Akiva, M., 2009. Calibration of dynamic traffic assignment models with point-to-point traffic surveillance. *Transportation Research Record*, *2090*(1), pp.1-9.

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), pp.376-396.

Wittman, M.D. and Belobaba, P.P., 2018. Customized dynamic pricing of airline fare products. *Journal of Revenue and Pricing Management*, *17*(2), pp.78-90.

Wood, N., Baker, T., Moran, M., Pritchard, G., Lomax, T.J., Steadman, M., Glover, B. and Dell, B., 2016. *Managed lanes in Texas: a review of the application of congestion pricing* (No. PRC 15-47-F). Texas A&M Transportation Institute.

Xie, Y., Danaf, M., Azevedo, C.L., Akkinepally, A.P., Atasoy, B., Jeong, K., Seshadri, R. and Ben-Akiva, M., 2019. Behavioral modeling of on-demand mobility services: general framework and application to sustainable travel incentives. *Transportation*, pp.1-23.

Xie, Y., Zhang, Y., Akkinepally, A., Ben-Akiva, M., 2019. A Personalized Choice Model for Managed Lane Travel Behavior. Submitted to *the 99th Annual Meeting of Transportation Research Board.*

Xu, S., 2009. *Development and test of dynamic congestion pricing model* (Dissertation, Massachusetts Institute of Technology).

Yáñez, M.F., Cherchi, E., Heydecker, B.G., and de Dios Ortúzar, J., 2011. On the treatment of repeated observations in panel data: efficiency of mixed logit parameter estimates. *Networks and Spatial Economics*, *11*(3), pp.393-418.

Yang, Q., Koutsopoulos, H.N., and Ben-Akiva, M.E., 2000. Simulation laboratory for evaluating dynamic traffic management systems. *Transportation Research Record*, *1710*(1), pp.122-130.

Yin, Y. and Lou, Y., 2009. Dynamic tolling strategies for managed lanes. *Journal of Transportation Engineering*, *135*(2), pp.45-52.

Zhang, C., Osorio, C., and Flötteröd, G., 2017. Efficient calibration techniques for large-scale traffic simulators. *Transportation Research Part B: Methodological*, *97*, pp.214-239.

Zhang, C. and Osorio, C., 2017. Efficient offline calibration of origin-destination (demand) for large-scale stochastic traffic models. In *Technical Report*. Massachusetts Institute of Technology.

Zhang, H., 2018. *Online calibration for simulation-based dynamic traffic assignment: towards large-scale and real-time performance* (Doctoral dissertation, Massachusetts Institute of Technology).

Zhang, Y., 2017. *Exploration of algorithms for calibration and optimization of transportation networks* (Dissertation, Massachusetts Institute of Technology).

Zhang, Y., Atasoy, B., Akkinepally, A. and Ben-Akiva, M., 2019. Dynamic Toll Pricing using Dynamic Traffic Assignment System with Online Calibration. *Transportation Research Record*, 0361198119850135.

Zhu, X., Wang, F., Chen, C. and Reed, D.D., 2019. Personalized incentives for promoting sustainable travel behaviors. *Transportation Research Part C: Emerging Technologies*.