Title
Toward Sustainable Cities: Modeling and Data Analytics for Urban Mobility

Author
Zhang, Jingwei

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Toward Sustainable Cities:
Modeling and Data Analytics for Urban Mobility

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Jingwei Zhang

2023

ABSTRACT OF THE DISSERTATION

Toward Sustainable Cities:
Modeling and Data Analytics for Urban Mobility

by

Jingwei Zhang
Doctor of Philosophy in Management
University of California, Los Angeles, 2023
Professor Christopher S. Tang, Co-Chair
Professor Auyon A. Siddiq, Co-Chair

Urbanization accelerates greenhouse gas emissions and traffic congestion, endangering the environmental sustainability of cities and causing significant economic losses. In this thesis, we aim to tackle practical problems faced by public sector by leveraging tools including economic modeling, optimization, and data analytics to achieve sustainable urban mobility.

In Chapter 2, we focus on an incentive mechanism design problem to improve **public transportation adoption**. Due to a prolonged decline in public transit ridership over the last decade, transit agencies across the United States are in financial crisis. To entice commuters to travel by public transit instead of driving personal vehicles, municipal governments must address the "last mile" problem by providing convenient and affordable transportation between a commuter's home and a transit station. This challenge raises an important question: *Is there a cost-effective subsidy program that can improve public transit adoption by solving the last-mile problem?* To address the question, we present and analyze two incentive mechanisms applied in practice for increasing commuter adoption of public transit. In

a direct mechanism, the government provides a subsidy to commuters who adopt a "mixed mode", which involves taking public transit and hailing rides to/from a transit station. The government funds the subsidy by imposing congestion fees on personal vehicles entering the city center. In an indirect mechanism, instead of levying congestion fees, the government secures funding for the subsidy from the private sector. These two mechanisms are especially relevant because several jurisdictions in the U.S. have begun piloting incentive programs in which commuters receive subsidies for ride-hailing trips that begin or end at a transit station. We then examine the implications of both mechanisms on five self-interested stakeholders (commuters, public transit agency, ride-hailing platform, municipal government, and local private enterprises), and provide conditions where either mechanism is superior with respect to total cost and commuter welfare. Our findings offer cost-effective prescriptions to municipal governments seeking to improve urban mobility and public transit ridership.

Chapter 3 of the thesis delves into improving **cycling ridership** through bike lane network planning. Sustainable urban infrastructure plays a pivotal role in the creation of livable cities. To meet the growing demand for cycling and reduce emissions, municipal governments worldwide have made significant investments in the expansion of bike lane networks. However, re-allocating road capacity from vehicles to cycling can often prove controversial due to the risk of exacerbating traffic congestion. This chapter presents a method for bike lane network planning that accounts for ridership and congestion effects. We first develop a game theoretic model that captures the traffic equilibrium and consumers' transportation mode choice. We then present an estimator for recovering unknown parameters of a traffic equilibrium model from features of a road network and observed vehicle flows, which we show asymptotically recovers ground-truth parameters as the network grows large. We subsequently present a prescriptive model that recommends paths in a road network for bike lane construction while endogenizing cycling demand, driver route choice, and driving travel times. Our framework allows for a thorough assessment of bike lane network expansion on both cycling ridership and traffic congestion, which offers valuable insights and solutions for

social planners seeking to promote urban mobility via urban infrastructure planning.

In Chapter 4, we build on the theoretical foundation of Chapter 3 and conduct a comprehensive empirical study on the City of Chicago. We bring together data on the road and bike lane networks, vehicle flows, travel mode choices, bike share trips, driving and cycling routes, and taxi trips to estimate the impact of expanding Chicago's bike lane network. We estimate that adding 25 miles of bike lanes as prescribed by our model can lift ridership from 3.9% to 6.9%, with at most an 8% increase in driving times. We also find that three intuitive heuristics for bike lane planning can lead to lower ridership and worse congestion outcomes, which highlights the value of a holistic and data-driven approach to urban infrastructure planning.

The dissertation of Jingwei Zhang is approved.

Felipe Caro

Charles J. Corbett

Kumar Rajaram

Auyon A. Siddiq, Committee Co-Chair

Christopher S. Tang, Committee Co-Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I express my profound gratitude to my esteemed advisors, Professors Auyon Siddiq and Christopher Tang, for their invaluable guidance, unwavering support, and encouragement throughout my doctoral studies. Auyon has been my guiding light from the very beginning of my research career, and has always been there for me ever since I knocked on the door of his office and asked if we can work together. His patience and positivity have sustained me over the past five years and through the challenging job market season. Chris's boundless energy and passion have been a constant source of inspiration, and I am grateful for his insightful suggestions that have shaped my research and encouraged me to connect it with practical applications. The hours spent in his office, standing in front of the whiteboard, deriving models, and discussing problems will shine in my memory. Through countless meetings, discussions, and prompt responses, Auyon and Chris have taught me how to identify research questions I am truly passionate about, model carefully, think independently and critically, and I aspire to become the kind of advisor and researcher they exemplify. I cannot express enough gratitude to these two remarkable individuals, as without them, I would not have become the best version of myself today.

I would also like to extend my sincere thanks to my committee members, Professors Felipe Caro, Charles Corbett, and Kumar Rajaram, for their insightful feedback and contributions that have broadened my perspectives. I am grateful to Felipe for his guidance in shaping my research ideas and improving my job talk, to Charles for his suggestions on reaching a wider audience with my research, and to Kumar for his instructions and encouraging smiley face on my final paper in his OM class. I also want to acknowledge the invaluable suggestions I have received from Professors Fernanda Bravo, Francisco Castro, Elisa Long, Velibor Mišić, and Scott Rodilitz throughout my job market season. Moreover, I am also indebted to Professor Sheng Liu for numerous in-depth discussions and his mentorship during the development of the third and fourth chapters of this thesis. His expertise and guidance have been invaluable.

I would like to express my gratitude to Neli and Bethany for their support in managing the logistics, which has greatly contributed to a smoother PhD journey for me. I am thankful to my senior peers in the Decisions, Operations, and Technology Management area, Ali, Anna, Bobby, Nur, Taylor, and Yi-Chun, for their tremendous help over the years and during my job search. Additionally, I am grateful to my cohort and junior peers, Irem, Mirel, Saeed, Jian, Jingyuan, Xinyi, Zach, Abolfazl, Martín, and Nareen, for their intellectual stimulation and emotional support.

I am deeply grateful to my roommate, Kexin Li, who has been a constant source of support during my challenging moments and has generously provided me with transportation whenever needed. I would also like to thank my friends, Yu Dai, Xinzhou Ge, and Qianhui Gao (just to name a few), who have offered me different perspectives and taught me to perceive the world through various lenses.

Finally, I express my deepest appreciation to my parents, Shixiu Jiang and Xiang Zhang, whose wholehearted dedication and efforts in raising and educating me have shaped who I am today. To my partner, Bingjia Wang, I am forever grateful for your unwavering love and care. Meeting you has been the most serendipitous encounter, and words cannot adequately convey how beautifully you have transformed my life. I would also like to thank our adorable furry friend, Lulu, whose presence brings immense joy and serves as a source of comfort.

I am truly honored and blessed to have these exceptional individuals in my life. Without their support, this dissertation would not have been possible.

VITA

EDUCATION

2018 – 2023 **University of California, Los Angeles**

Ph.D. Candidate in Management (Decisions, Operations & Technology Management)

2017 – 2018 **Columbia Univsersity**

M.Sc., Marketing Science

2013 – 2017 **Tsinghua Univsersity**

B.Eng., Industrial Engineering

HONORS AND AWARDS

2022 Second Place, INFORMS IBM Best Student Paper Award

2022 Finalist, INFORMS Workshop on Data Mining and Decision Analytics Best Paper Award (Applied Track)

2022 First Place, POMS College of Sustainable Operations Student Paper Competition

2022 UCLA Dissertation Year Fellowship

PUBLICATIONS

PUBLISHED & UNDER REVIEW PAPERS

Sheng Liu, Auyon Siddiq, **Jingwei Zhang**, "Planning bike lanes with data: Ridership, congestion, and path selection," Major Revision, *Management Science*.

Auyon Siddiq, Christopher S. Tang, **Jingwei Zhang**, "Partnerships in urban mobility: Incentive mechanisms for improving public transit adoption," *Manufacturing & Service Operations Management*, vol. 24,no. 2, pp. 956–971, 2022.

CHAPTER 1

Introduction

Accelerating global urbanization is typically accompanied by increased greenhouse gas emissions and traffic congestion, which threatens the environmental sustainability of cities and incurs enormous economic losses. Alleviating emissions and traffic congestion is likely to require a multi-faceted solution, including increased adoption of sustainable travel modes (e.g., walking, cycling, and public transit). In this thesis, motivated by practical problems faced by public sector, we utilize multiple methodologies including economic modeling, optimization, and data analytics, to support organizations in making decisions that are economically, environmentally, and socially sustainable.

The thesis is organized as follows. Chapter 2 tackles an incentive design problem aimed at increasing **public transit ridership**. This issue is motivated by innovative partnerships established between public transit agencies and private ride-hailing platforms, as well as recent advancements in transit mobile apps (Corselli 2020). In Chapter 3, we introduce an estimation and optimization framework for bike lane planning that seeks to improve **cycling ridership**. This initiative is driven by recent plans by various municipal governments to expand bike lane networks. Chapter 4 builds on the foundations of Chapter 3 by conducting an empirical study of bike lane expansion in the City of Chicago and examining its policy implications. Chapter 5 concludes the thesis and lists potential future research directions.

In the following sections, we will present a high-level overview of the main ideas explored in each chapter.

1.1 Partnerships in Urban Mobility: Incentive Mechanisms for Improving Public Transit Adoption

In recent years, the decline in public transit ridership has left transit agencies across the U.S. facing financial crises. The "last mile" problem, which occurs due to the lack of convenient and affordable transit options between an individual's home and a transit station, is a major obstacle to increasing public transit utilization. To address this issue, many municipalities, including Los Angeles and Philadelphia (Schwieterman and Livingston 2018), have formed partnerships with ride-hailing companies to offer subsidies for rides to and from transit stations.

To identify a cost-effective mechanism to solve the last mile problem and improve transit adoption, two mechanisms exist in practice. The first is a direct mechanism [D] in which the subsidy program is funded by congestion fees collected from drivers, as seen in London (Transport for London 2019). The second mechanism [I] is an indirect approach, which was first implemented in 2020 by Miami (Corselli 2020). The funding source for this mechanism is partner private sector who benefit from increased transit usage, such as shops near transit stations and advertisers in public transit mobile apps.

Chapter 2 investigates the equilibrium of both mechanisms and provides conditions under which one mechanism outperforms the other. We first present a novel game-theoretic model to capture the strategic interactions among five self-interested stakeholders, including commuters, public transit agencies, ride-hailing platforms, municipal governments, and local private enterprises. By examining the equilibrium outcomes, three key findings are obtained.

First, we characterize how the optimal interventions associated with the direct or the indirect mechanism depend on: (a) the coverage level of the public transit network; (b) the public transit adoption target; and (c) the relative strength of commuter preferences for driving over taking public transit. Second, we show that the direct mechanism cannot be budget neutral without undermining commuter welfare. However, when the public transit

adoption target is not too aggressive, we find that the indirect mechanism is budget neutral, and it increases both commuter welfare and sales to the private enterprise. Finally, we show that although the indirect mechanism restricts the scope of government intervention (by eliminating the congestion fee), it can dominate the direct mechanism by leaving all stakeholders better off, especially when the adoption target is modest. In conclusion, our findings offer cost-effective prescriptions for improving urban mobility and public transit ridership.

1.2 Planning Bike Lanes: Ridership, Congestion, and Path Selection

In Chapter 3, we focus on improving cycling ridership via bike lane network planning. This chapter is strongly motivated by practice, and our modeling approach is informed by discussions with city planners from the City of Chicago's Department of Transportation and the City of Vancouver's Transportation Planning group.

Empirical evidence suggests an abundance of bike lanes is associated with increased cycling ridership and safety. In order to accommodate the increasing demand for cycling and encourage further adoption of this mode of transportation, municipal governments have recently made significant investments in the installation of new bike lanes. However, bike lane expansion remains contentious because it necessarily reduces road capacity for automotive vehicles and amplifies congestion. The key research question of Chapter 3 is: Where should bike lanes be built to promote cycling ridership while mitigating congestion effects? Realizing the lack of analytical and data-driven approach in solving such complicated infrastructure planning problems, our work fills the gap by developing an estimation procedure for quantifying the effect of bike lanes on cycling ridership and congestion, which we extend to a prescriptive model that generates recommendations for bike lane expansion.

We first develop an estimator for learning parameters of a traffic equilibrium model from

features of the road network and vehicle traffic flows. We provide sufficient conditions under which the estimator is statistically consistent, i.e., provably recovers ground-truth parameters in the limit as the network grows large. Building upon the estimation approach, we present a prescriptive model for planning new bike lane locations, which endogenizes transportation mode demand and travel times, and leads to a nonconvex integer optimization problem. For tractability, we approximate the model as a mixed-integer linear program, and we show that the approximation error (the suboptimality with respect to the exact formulation) can be bounded as a function of network and demand parameters.

1.3 Empirical Study: Expanding Chicago's Cycling Infrastructure

Chapter 3 introduces a novel framework for estimating traffic equilibrium and optimizing bike lane planning, while Chapter 4 demonstrates its application through a comprehensive empirical study utilizing real-world data from the city of Chicago. This study combines seven distinct datasets, including Chicago's road network, existing bike lane network, observed traffic flows, travel mode choices, bike share trips, driving and cycling routes, and taxi trips.

Our estimation, based on this extensive dataset, achieves an out-of-sample relative travel time prediction error of 14.3%, which is competitive to previously proposed methods for predicting travel times using taxi trip data (Zhan et al. 2013). Once the travel time is estimated, we proceed to identify the optimal bike lane locations under various specifications of social planners' tolerance for congestion increase and financial budget, to cater to the needs of different municipalities. The outcomes of our analysis indicate that adding 25 miles of bike lanes recommended by our model could increase ridership from 3.9% to 6.9%, with a maximum increase in driving times of 8%. In addition to these findings, our research highlights the inadequacy of relying on intuitive heuristics for bike lane planning, which can lead to suboptimal outcomes in terms of cycling ridership and congestion. This emphasizes the importance of a holistic and data-driven approach to urban infrastructure planning.

CHAPTER 2

Partnerships in Urban Mobility: Incentive Mechanisms for Improving Public Transit Adoption

2.1 Introduction

The mission of the United States Federal Transit Administration (FTA) is to "*enhance citizens' mobility, accessibility, and economic well-being through the development and management of public transport services*". However, over the last decade, the FTA's mission has become increasingly threatened by a prolonged decline in public transit ridership across the United States. From 2008 to 2017, per capita ridership on buses, subways, and commuter trains saw a 4% drop in San Francisco, a 5% drop in New York City, and over a 25% drop in both the Los Angeles and Washington DC areas (The Economist 2019). Moreover, this decline is nationwide: bus ridership has decreased from 5.6 billion trips in 2008 to 4.7 billion trips in 2018 (Kamp 2020). The widespread decline in public transit ridership is believed to be driven by multiple factors, including lower gasoline prices, changing demographic patterns within cities, and the rise of alternative modes of transportation, such as ride-hailing (Mallett 2018). As a result of declining fare revenue, over 98% of the 2,200 public transit agencies in the U.S. are in financial crisis (Federal Transit Administration 2017). Covering these losses is challenging for many municipalities due to limited funding, as well as other competing priorities, such as public safety, education, and affordable housing.

To combat declining ridership and revenue, cities can certainly scale back public transit services. However, doing so would undermine the mission of the FTA for the following rea-

sons. Reducing access to public transit can hinder mobility, especially for those who cannot afford alternative modes of transportation (e.g., private vehicles, ride-hailing services, and taxis). Moreover, the American Public Transit Association (APTA) has found evidence suggesting that investments in public transit can generate economic returns by lifting business sales and promoting job growth (APTA 2019c). In addition to the economic benefits, public transit plays a vital role in reducing carbon emissions and alleviating traffic congestion (APTA 2019c).[1] Some cities have offered free bus rides to increase ridership (e.g., in 2019, Lawrence, Massachusetts and Olympia, Washington began offering free bus rides to improve mobility for the poor and the elderly (Kamp 2020)); however, the long-term financial viability of fully subsidizing bus fares is questionable. Therefore, there is an urgent need for municipalities to develop cost effective – ideally, budget neutral[2] – solutions for increasing transit ridership.

A major barrier to increasing public transit utilization is the "last mile" problem – a challenge caused by the lack of convenient and affordable transit services between an individual's home and a transit station (APTA 2019b, LA Metro 2016). Despite increased investments in public transit – total inflation-adjusted funding in the U.S. increased from $60 to $72 billion between 2007 and 2017 (APTA 2019a) – commuters continue to eschew public transit in favor of personal vehicles, in part due to the last mile problem. Indeed, in 2017, over 85% of US workers used personal vehicles to commute, with less than 6% relying on public transit (Bureau of Transportation Statistics 2018). Further, while offering subsidized parking near transit stations may appear to be an attractive solution to address the last mile problem, such programs are unlikely to be feasible on a large scale due to limited parking near transit stations, as well as the high cost of building, maintaining, and

[1]Single-occupancy vehicles emit 1.5 times the CO_2 emissions of buses, and four times the emissions of subways (Hodges et al. 2010). With respect to traffic, public transit has been found to be critical in reducing congestion during peak hours (Anderson 2014).

[2]The notion of delivering public services in a budget neutral manner has recently been proposed by the Centers for Medicare and Medicaid Services; the reader is referred to CMS (2018) for details.

subsidizing parking spaces. For example, in Los Angeles County, only 24,000 subsidized parking spaces are available, to support approximately 3.6 million daily commuters (Linton 2016, US Census Bureau 2018a).

To address the last mile problem, various municipal governments are forming partnerships with private transportation companies. For example, in 2019, Los Angeles partnered with the ride-hailing platform Via to launch a pilot project that offers commuters subsidized rides to and from public transit stations. The subsidy is supported by a grant from the Federal Transit Administration (Los Angeles County Metropolitan Transportation Authority 2019). Several other major cities (e.g., Atlanta, Austin, Detroit, Philadelphia, and Tampa) have also begun testing the use of ride-hailing subsidies as a potential solution to closing the last mile gap (Schwieterman and Livingston 2018, APTA 2019e). These subsidy programs have only recently become possible due to advances in information technology, which now enable on-demand ride-hailing, real time tracking of passenger locations, and online/mobile fare payments.[3] More generally, the integration between public transit systems and ride-hailing is accelerating – for example, in early 2020, Uber announced a new in-app feature that coordinates trip drop-off times with train schedules (Uber 2020). This novel integration between private transportation services and public transit opens the door to new forms of mutually beneficial partnerships.[4]

In this paper, we investigate two incentive mechanisms that aim to increase public transit ridership by addressing the last mile problem. Both mechanisms exhibit the following characteristics:

[3]Information technology has been successfully leveraged to help commuters plan and pay for multi-modal trips within a mobile app; examples include Whim in Helsinki, Citymapper in London, Moovel in Germany, UbiGo in Gothenburg, TAP in Los Angeles, Clipper in San Francisco, and Opal in Sydney (Goodall et al. 2017, Cole 2017).

[4]There is also evidence to suggest that ride-hailing may be contributing to the decline in public transit ridership, and adding to traffic congestion in city centers (Brown 2020). Partnerships between public transit agencies and private ride-hailing platforms may facilitate cooperation instead of competition, which can be beneficial to both parties (i.e., increased transit ridership in the city center and increased demand for ride-hailing in suburban areas).

1. **Both mechanisms have a common feature.** Both mechanisms rely on a strategic partnership between a public transit agency and a private ride-hailing platform. Specifically, in both mechanisms, commuters receive a subsidy for adopting a *mixed mode* of transportation, in which commuters use the ride-hailing service to travel the last mile distance ("x") between their homes and a transit station, and use public transit to travel between the transit station and a final destination (e.g. city center). In this setting, x can be interpreted as a measure of the *coverage*[5] of the public transit system: the last mile x is small (large) if commuters have convenient (inconvenient) access to transit stations from their homes.

2. **Both mechanisms share a common goal.** Both mechanisms are intended to increase the adoption of the mixed mode by a constant factor ("β").

The key differences between the mechanisms are: (a) the role of the local government; and (b) the source of funding for the ride-hailing subsidy. To elaborate, the first incentive program we consider is *direct* in the sense that the the government offers a subsidy ("s") directly to each commuter as an incentive for adopting the mixed mode. To defray the cost associated with the subsidy, the government charges a congestion fee ("e") to commuters who instead choose to drive a personal vehicle into the city.[6] Therefore, under the "direct mechanism" (denoted by [D]), the government uses two levers to entice commuters to adopt the mixed mode: the ride-hailing subsidy s, and the congestion fee e.

[5]In the context of public transit, "coverage" broadly refers to the accessibility of transit stations by the general population (i.e., commuters in our context) (Tomer et al. 2011).

[6]Congestion fees have been adopted by many municipal governments, including Singapore, Hong Kong, London, Milan, and Stockholm. Singapore was the first to introduce congestion pricing as a tool to control traffic volume, where citizens pay for fees when they enter city center areas (Development Asia 2018). London introduced congestion fees in 2003, where there is a charge for entering London's congestion charging zone (13 square miles) between 7 a.m. and 6 p.m. on weekdays (Transport for London 2019). Since the introduction of congestion fees, traffic congestion and private vehicle usage in London has dropped by 25% and 39%, respectively (Badstuber 2018). Similarly, in Milan, public ridership increased 12.5% from 2007 to 2013 (Croci 2016).

Unlike the direct mechanism [D] that relies on congestion fees collected from commuters who drive, the local government secures funding for the subsidy ("z") from the private sector in the second mechanism. This mechanism is *indirect* because the government's role is restricted to facilitating the transfer of the subsidy from a private enterprise to commuters. This scheme is motivated by the fact that many municipalities may not have sufficient funds to further subsidize public transit. To relieve this financial burden, Cole (2017) proposes the following partnership between a municipal government and a private enterprise. The government and public transit agency develop a mobile app to process different transactions, including fares collected from the commuters and subsidies provided by the private enterprise. The private enterprise benefits from higher transit ridership through an expanded customer base. Here, the customer base may represent foot traffic at stores near the transit station (e.g., coffee shops, bakeries, health clinics, spas, hair salons), or number of views on an in-app advertisement. Under this arrangement, the private enterprise can recover the cost associated with the subsidy from the extra revenue derived from the increased demand. In contrast to the direct mechanism, this "indirect mechanism" (denoted by [I]) involves just a single lever: the subsidy z, provided by the private enterprise.

Mechanism [I] is motivated by a growing interest among transit agencies in leveraging private sector partnerships to subsidize public transit trips. For example, since early 2020, commuters in Miami-Dade County can earn loyalty points (i.e., subsidies) by taking public transit, which can be redeemed for discounts on future trips through the transit agency's mobile app (known as "Go Miami-Dade Transit") (Corselli 2020). These loyalty points are funded by local businesses who post in-app video ads within the mobile app. The National Hockey League (NHL) and the Seattle Monorail formed a partnership in which NHL will subsidize rides on the Monorail for fans attending hockey games (Condor 2020). Both of partnerships can be viewed as a private enterprise (i.e., a single aggregate entity) subsidizing public transit trips in exchange for an expanded audience, which capture the main characteristics of mechanism [I].

2.1.1 Outline and contributions

Our intent is to investigate the impact of both subsidy mechanisms on multiple stakeholders: a municipal government, suburban commuters, city dwellers, a public transit agency, a ride-hailing platform, and a private enterprise. To do so, we examine the following questions:

1. How does each mechanism perform in terms of: (a) operating cost (borne by the government); (b) commuter welfare; (c) the public transit agency's revenue; (d) the ride-hailing platform's revenue; and (e) the private enterprise's profit?

2. How does the last mile distance (or coverage) x and the mixed mode adoption target β affect: (a) the optimal direct subsidy s^* and congestion fee e^* in mechanism [D], and (b) the optimal indirect subsidy z^* in mechanism [I] ?

3. Under what conditions, if any, should the municipal government adopt mechanism [I] over mechanism [D]?

We examine the above questions by developing a game-theoretic model that captures the interactions among all stakeholders. Specifically, in mechanism [D], we identify the optimal subsidy level s^* and congestion fee e^* that minimizes the government's net spending (subsidy cost less the revenue generated by the congestion fee), subject to improving adoption of the mixed mode by a factor of at least β. In addition, we restrict the set of feasible subsidies and congestion fees to those that satisfy participation constraints associated with commuters, the ride-hailing platform, the public transit agency, and the private enterprise. We also determine the optimal subsidy z^* under mechanism [I], subject to the same set of stakeholder participation constraints.

Our key findings are summarized as follows:

1. **The optimal incentive under mechanism [D] depends on the last mile distance x.** Specifically, the optimal incentive entails a large congestion fee and a small

ride-hailing subsidy when the transit coverage is high (i.e. when the last mile distance x is small). Conversely, when transit coverage is low (i.e., x is large), this prescription is reversed: the optimal incentive relies on aggressive ride-hailing subsidies and relatively smaller congestion fees. Moreover, we find that whether the optimal subsidy s^* increases or decreases in the last mile distance x depends on the commuter's relative preference between driving and taking public transit.

2. **Budget neutrality is not attainable under mechanism [D]:** It is not possible for the government to fully recoup the cost of the subsidy through congestion fees without decreasing total commuter welfare.

3. **Mechanism [I] can *dominate* mechanism [D] under certain conditions.** Despite the government having access to two levers in mechanism [D] (the subsidy s and the congestion fee e) and only a single lever in mechanism [I] (the subsidy z), mechanism [I] can *dominate* mechanism [D], but only when the adoption target β is modest. Specifically, when β is modest, the commuters, city dwellers, public transit agency, ride-hailing platform, and private enterprise are all better off under mechanism [I].

The remainder of the chapter is organized as follows. In Section 2.1.2, we discuss related literature. In Section 2.2, we present a general model for analyzing both mechanisms. In Section 2.3, we present mechanism [D] and analyze the optimal incentive scheme (e^*, s^*). In Section 2.4, we present mechanism [I], analyze the optimal subsidy z^*, and compare the performance of both mechanisms. In Section 2.5, we consider variants of mechanisms [D] and [I] that address the trade-off between operating cost and commuter welfare. In Section 2.6, we conclude and discuss potential future research directions.

2.1.2 Related literature

This chapter is related to three streams of literature: government subsidy programs, public-private partnerships, and budget neutral mechanisms.

Government subsidy programs. Within the operations literature, there is a growing focus on subsidy programs that promote the production or adoption of socially beneficial goods or services. This study can be categorized into subsidies for producers (e.g., farmers, manufacturers, healthcare providers) and consumers. Subsidies for producers have been studied in the context of healthcare (Taylor and Xiao 2014, Levi et al. 2017, Aswani et al. 2019), green technology (Ma et al. 2019, Bansal and Gangopadhyay 2003, Alizamir et al. 2016), and agriculture (Alizamir et al. 2019, Akkaya et al. 2019). This chapter is closer to the literature on consumer-facing subsidies. In the home appliances industry, Yu et al. (2018) determine whether the government should subsidize consumers only, manufacturers only, or both, where the goals of the government are to improve manufacturer profit and consumer welfare in rural areas. Xiao et al. (2019) examine the impact of the same home appliance subsidy program, which they conclude improves both affordability and accessibility for rural customers. In the context of solar energy technology, Chemama et al. (2019) compare the effectiveness of static and dynamic consumer subsidies on supplier behavior. This study differs from this existing literature in that we investigate whether the subsidy program can be budget neutral from the perspective of the subsidy provider.

Previous work in the operations literature on public transit subsidies is sparse. Lodi et al. (2016) consider a setting where operation of the public transit system is outsourced to the private sector, and the government provides a subsidy to offset the operating cost. Yang and Lim (2018) use a field experiment to show that temporarily subsidizing public transit can lead to long-term changes in commuter behavior. The paper that is most similar to ours in this line of research is by Xiao and Zhang (2014), who also consider congestion fees and transit subsidies simultaneously. They focus on the impact that commuter heterogeneity in value-of-time has on the optimal design of congestion fees, and show that transit subsidies can offset the loss in commuter welfare due to the congestion fee. Our work is different in that we focus on the role that congestion fees and transit subsidies can play in improving public transit ridership.

Public-private partnerships. Our work contributes to the modeling of public-private partnerships (PPP), which refers to a private sector partner "financing, constructing, and managing a project in return for a promised stream of payments directly from government or indirectly from users over the projected life of the project or some other specified period" (Weimer and Vining 2017). This study belongs to the stream of PPP literature where the government has all of the bargaining power. In the existing literature, this is often modeled as either a principal-agent problem or a Stackelberg game, where the government is the principal/leader and the firm is the agent/follower; applications include healthcare (So and Tang 2000, Lee and Zenios 2012, Gupta and Mehrotra 2015, Guo et al. 2019, Aswani et al. 2019), disaster management (Guan and Zhuang 2015, Guan et al. 2018) and risk management (Bakshi and Gans 2010). With respect to transportation, existing work on public-private partnerships has primarily focused on infrastructure. Lodi et al. (2016) address the government's incentive design problem when management of the public transit service is outsourced to private operators. Gagnepain and Ivaldi (2002) examine the effect on social welfare under different types of contracts between the regulator and the public transit operator. Hansson (2010) considers a multi-principal setting, where the local, regional and county governments interact to regulate public transit procurement. In contrast to these papers, the private partners play a different role in our work, namely, they provide a complementary transportation service (in the case of the ride-hailing platform) and funding for subsidies (in the case of the partner enterprise).

Previous work has also considered settings where the public-private partnership is formed through negotiation. In the transportation context, Kang et al. (2013) investigate royalty bargaining associated with underground railway station construction, and Wang and Zhang (2016) examine road pricing of transportation networks with both public and private roads. Other application areas include natural resource development (Anandalingam 1987), public procurement (Gur et al. 2017, Saban and Weintraub 2019) and global supply chain management (Cohen et al. 2018, Cho et al. 2019, Cohen and Lee 2020).

Budget neutral mechanisms. This study also contributes to the literature on budget neutral policies. This study has primarily appeared in the public policy literature, and has addressed issues such as social security reform (Burkhauser and Smeeding 1994), environmental taxation (Goulder 1995, Murray and Rivers 2015), and fiscal policy (Correia et al. 2013, D'Acunto et al. 2016). Within the operations management literature, previous work on budget neutral policies is scarce. Guo et al. (2014) optimize a two-tier queuing system with both a free server and fee-based server, in the setting where the system is self-financed by the costly server. Arifoglu and Tang (2019) develop a budget neutral incentive mechanism for coordinating a decentralized influenza vaccine supply chain.

With respect to transportation settings, most existing papers that focus on budget neutrality are based on schemes under which subsidies/rebates are funded by congestion fees, similar to mechanism [D] in our work (see, e.g., Guo and Yang (2010), Nie and Liu (2010), Chen and Yang (2012), Xiao and Zhang (2014)). Our work differs in that we also consider an indirect mechanism [I], where the government attains budget neutrality by obtaining funding from a private enterprise, instead of imposing congestion fees. Further, we compare mechanisms [D] and [I] in terms of their impact on the relevant stakeholders.

2.2 Model Preliminaries

A unit mass of commuters are located a (last mile) distance $x > 0$ from a transit station. All commuters must travel to a city center that is located an additional distance of 1 beyond the transit station. For tractability, we consider a parsimonious model where commuters choose between two modes of travel:[7] *driving* or a *mixed mode*. As depicted in Figure

[7]We focus on suburban commuters for whom walking to the transit station is prohibitively costly. Accordingly, we exclude the mixed mode that combines walking and public transit. Similarly, we exclude the case where commuters combine electric scooters or bicycles and public transit because (a) commuters' comfortable biking distance is limited (Rastogi and Krishna Rao 2003) and (b) less than 4% of commuters use scooters or bicycles to get to/from transit stations (Federal Highway Administration 2017)). We also exclude the mixed mode that combines driving and public transit, due to limited parking spaces near transit stations in the U.S. Commuters who exclusively use ride-hailing to commute are also not considered here(only 0.2%

2.1, x represents the length of the "last mile" that is not covered by public transit. Hence, commuters who choose the driving mode will drive a personal vehicle for a distance of $1+x$ to the city center. However, commuters who choose the mixed mode will first travel a distance of x to the transit station via a ride-hailing service, and then travel the remaining distance of 1 by public transit.

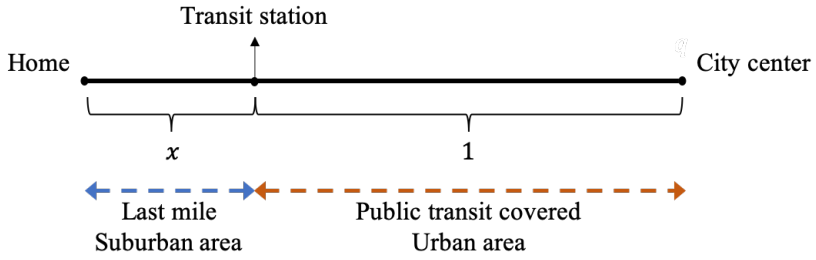


Figure 2.1: Travel distance of a commuter.

2.2.1 Base Model

We first describe our base model in the absence of any subsidies or congestion fees. In later sections, we extend our base model to incorporate the two incentive schemes. We note here that our focus throughout the chapter is on daily commuters who live in suburban areas outside the coverage area of the public transit system; therefore, all aspects of the model (e.g., transit ridership, operating cost, ride-hailing platform revenue) are defined with respect to this group of commuters only.

Commuter utility. Each commuter obtains a "base" utility M for commuting to the city center. A commuter has valuation (or willingness-to-pay) V per unit distance traveled via ride-hailing or public transit, or δV if she drives.[8] We assume $\delta > 1$ to reflect a higher

of suburban residents commute via taxi or ride-hailing services (Federal Highway Administration 2017)).

[8]Empirical evidence suggests that a commuter's willingness-to-pay for transportation is increasing in the travel distance (Van Ommeren et al. 2000, Jou et al. 2012). For tractability, we assume a linear relationship between willingness-to-pay and distance, given by the parameter V.

intrinsic utility for driving.[9] To capture heterogeneity among commuters, we assume $V \sim U[0,1]$.[10] (In Appendix A.3, we consider an alternative setting in which V is deterministic and identical for all commuters. Instead, we consider commuter heterogeneity in terms of x by assuming that x (i.e., the distance between a commuter's home and the transit station) is uniformly distributed. For this alternative setting, we obtain the same structural result.)

A commuter will travel a distance of $1+x$ if she chooses the driving mode of transportation. Similarly, under the mixed mode, the commuter will first travel a distance of x via ride-hailing and the remaining distance of 1 via public transit. In our model, we assume a commuter's cost (or fare) for driving, ride-hailing, and taking public transit are denoted by d, r, and p, respectively. We assume throughout that $r > d > p > 0$, which is supported empirically.[11] Therefore, the utilities associated with driving and the mixed mode in the absence of any incentives, denoted by U_d^0 and U_m^0, respectively, are given by

$$U_d^0 = M + (x+1)\delta V - (x+1)d, \tag{2.1a}$$

$$U_m^0 = M + (x+1)V - (rx+p). \tag{2.1b}$$

Travel mode demand. It follows from (2.1) that a commuter will adopt the mixed mode if and only if $U_m^0 \geq U_d^0$. We assume the base value M is large enough such that $U_d^0 \geq 0$ and $U_m^0 \geq 0$. Therefore, $U_m^0 \geq U_d^0$ if and only if:

$$V \leq v^0 \equiv \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}. \tag{2.2}$$

[9]Recent survey data has shown that commuters generally prefer driving to public transit and ride-hailing services (Zhu and Fan 2018). Our main results also persist in a model where commuters prefer ride-hailing to public transit, so that the additional value associated with ride-hailing is $l \cdot V$, where $\delta > \ell \geq 1$. For ease of exposition, we assume $\ell = 1$ throughout.

[10]Commuter valuation for each travel mode can vary based on demographic features such as age, income level, or health conditions (Zhu and Fan 2018). For simplicity, we assume the valuation $V \sim U[0,1]$.

[11]The cost of driving is estimated to be \$0.76/mile, assuming a mileage of 10,000 miles per year (American Automobile Association 2018). Ride-hailing is estimated to cost more than \$1.07/mile (Lyft 2021). The cost of commuting by public transit is typically much lower; for example, the cost of public transit in Los Angeles is estimated to be \$0.2/mile, assuming an average commute distance of 16 miles (Leonard 2019).

Let D_d^0 and D_m^0 denote the demand for driving and the mixed mode, respectively. Because $V \sim U[0,1]$, we can apply (2.2) to show that:

$$D_d^0 = \int_{v^0}^{1} 1 \cdot dV = 1 - \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}, \tag{2.3a}$$

$$D_m^0 = \int_{0}^{v^0} 1 \cdot dV = \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}. \tag{2.3b}$$

Note that D_m^0 represents the additional public transit ridership generated by the mixed mode commuters. Because $r > d > p > 0$, it is straightforward to verify that D_m^0 is decreasing in x, which implies that commuters that live far from the transit station are more likely to drive and less likely to adopt the mixed mode. Further, D_m^0 also measures the "reduction" in traffic congestion, because $D_m^0 = 1 - D_d^0$.

Let $[a]^+ = \max\{a, 0\}$. To exclude the degenerate cases where *all* commuters adopt the same travel mode, we assume that $x \in (\underline{x}, \bar{x})$, so that $D_m^0 > 0$ and $D_d^0 > 0$, where $\underline{x} = \left[\frac{(d-p)-(\delta-1)}{(\delta-1)+(r-d)}\right]^+$ and $\bar{x} = \frac{d-p}{r-d}$. We also assume that the public transit and the ride-hailing platform have sufficient capacity to accommodate demand for the mixed mode.[12]

Commuter welfare. Because $V \sim U[0,1]$, the commuter welfare in the base model is given by

$$W^0 = \int_{0}^{v^0} U_m^0 dV + \int_{v^0}^{1} U_d^0 dV. \tag{2.4}$$

Transit agency revenue. By noting that the unit fare of public transit is p, the revenue generated by the public transit agency from these commuters under the mixed mode demand D_m^0 is:

$$\Pi_p^0 = p \cdot D_m^0. \tag{2.5}$$

Ride-hailing platform revenue. Recall that each mixed mode commuter will travel a distance of x via ride-hailing to the transit station, at a unit fare of r. The revenue generated

[12]The average vehicle occupancy rate of public transit in 2017 was less than 30% (APTA 2019d). Ride-hailing services also have ample capacity: the average idle rate of Uber drivers is 30-40% (Currie 2018, Brown 2020).

by the ride-hailing platform from these commuters under the mixed mode demand D_m^0 is:

$$\Pi_r^0 = r \cdot x \cdot D_m^0. \tag{2.6}$$

Private enterprise profit. Consider a private enterprise who sells to public transit commuters. This private enterprise may be a business that sells goods or services to passengers that pass through a transit station, or one that places advertisements within a transit system's mobile app or physical walkways. We model the enterprise profit as $(k-c)[1-k\alpha]^+ D_m^0$, where k and c represent the unit price and cost, respectively, and $[1-k\alpha]^+$ represents the proportion of mixed mode commuters that purchase the product. Note that $\alpha > 0$ is the commuters' price sensitivity. To avoid the trivial case where the enterprise's optimal profit is non-positive, we assume $\alpha c < 1$.[13] Then, for any given mixed mode demand D_m^0, the optimal retail price in the base model is $k_b = \frac{\alpha c+1}{2\alpha}$ so that the corresponding optimal profit is equal to $(k_b - c)[1 - k_b\alpha]^+ D_m^0 = \frac{(1-\alpha c)^2}{4\alpha} D_m^0$. For ease of exposition, we let $K \equiv \frac{(1-\alpha c)^2}{4\alpha}$ to denote the profit margin so that the corresponding total profit can be expressed as:

$$\Pi_s^0 = K \cdot D_m^0. \tag{2.7}$$

2.2.2 A Generalized Mechanism: Government Subsidy, Congestion Fees and Private Subsidy

By using the quantities as defined above, we now present our model that incorporate the aforementioned incentives. To avoid repetition, we shall present a generalized mechanism that combines three interventions: (1) a congestion fee e charged to each commuter who travels by a personal vehicle; (2) a subsidy s for each commuter who takes the mixed mode, paid for by the government; and (3) a subsidy z for each commuter who takes the mixed mode, paid for by the private enterprise. We denote this generalized mechanism by (e, s, z). When $z = 0$, this mechanism reduces to mechanism [D]. Also, when $e = s = 0$, the mechanism simplifies to mechanism [I]. To isolate the effect of (e, s) under mechanisms [D] and the

[13] Note that if $\alpha c \geq 1$, then $(k - c)[1 - k\alpha]^+ D_m^0 \leq 0$ for any $D_m^0 \geq 0$ and $k \geq 0$.

effect of s under mechanism [I], we shall analyze mechanism [D] in Section 2.3 and mechanism [I] in Section 2.4 as two separate mechanisms. (Based on our knowledge, most transit agencies adopt either mechanism [D] or [I], but no both. However, we shall analyze the hybrid mechanism that combines both mechanisms [D] and [I] in Appendix A.1 (i.e., when (e, s) and z are permitted to be non-zero) and show that the structural results continue to hold.)

The sequence of events is as follows. First, the government chooses whether to adopt mechanism [D] or mechanism [I]. Second, the government sets the incentives that correspond to the chosen mechanism. Specifically, the government sets the congestion fee e and government subsidy s if mechanism [D] is chosen, or the private subsidy z if mechanism [I] is chosen. Next, once (e, s) or z is set, the private enterprise selects the unit price k to maximize its profit. Finally, each commuter chooses to commute by either driving or adopting the mixed mode.

By considering (2.1) along with the incentive (e, s, z), commuter utilities associated with different modes of transportation can be written as:

$$U_d(e, s, z) = U_d^0 - e, \tag{2.8a}$$

$$U_m(e, s, z) = U_m^0 + s + z, \tag{2.8b}$$

where U_d^0 and U_m^0 are given in (2.1).[14]

Hence, a commuter will adopt the mixed mode under incentive (e, s, z) if and only if $U_m(e, s, z) \geq U_d(e, s, z)$, or equivalently, if her valuation V satisfies:

$$V \leq v(e, s, z) \equiv \min\left\{1, \frac{(d-p) - (r-d)x + e + s + z}{(\delta - 1)(x+1)}\right\}. \tag{2.9}$$

Akin to the base case without incentives, the demand for each transportation mode under

[14] Observe from (2.1) that the utilities U_d^0 and U_m^0 are based on parameters x, d, r and p. Hence the lump-sum congestion fee e and subsidy s and z in (2.8) can also be expressed as a percentage fee and subsidy through an appropriate scaling of those parameters.

the incentive (e, s, z) is given by

$$D_d(e,s,z) = \left[1 - \frac{(d-p) - (r-d)x + e + s + z}{(\delta - 1)(x+1)}\right]^+, \tag{2.10a}$$

$$D_m(e,s,z) = 1 - \left[1 - \frac{(d-p) - (r-d)x + e + s + z}{(\delta - 1)(x+1)}\right]^+. \tag{2.10b}$$

Note that if $(e, s, z) = (0, 0, 0)$, then $v(e, s, z)$, $D_d(e, s, z)$, and $D_m(e, s, z)$ reduce to their base values v^0, D_d^0, and D_m^0 as stated in Section 2.2.1, respectively.

Performance metrics. Next, we define the metrics by which we evaluate the performance associated with the incentive (e, s, z). Because the government offers a subsidy s to each of the $D_m(e, s, z)$ mixed mode commuters and collects a fee e from each of the $D_d(e, s, z)$ commuters who drive, and meanwhile the government passes on a subsidy z from the partner enterprise to each mixed mode commuter, the total cost for the government to operationalize the incentive (e, s, z) is:

$$\begin{aligned} C(e,s,z) &= (s+z)D_m(e,s,z) - eD_d(e,s,z) - zD_m(e,s,z) \\ &= sD_m(e,s,z) - eD_d(e,s,z). \end{aligned} \tag{2.11}$$

Note that when $e = s = 0$, there is no operating cost $C(0, 0, z) = 0$. Similar to the base case, the total welfare accrued to all commuters is:

$$W(e,s,z) = \int_0^{v(e,s,z)} U_m(e,s,z)dV + \int_{v(e,s,z)}^1 U_d(e,s,z)dV, \tag{2.12}$$

the public transit agency's revenue generated from these commuters is:

$$\Pi_p(e,s,z) = p \cdot D_m(e,s,z), \tag{2.13}$$

and the ride-hailing platform's revenue generated from these commuters is:

$$\Pi_r(e,s,z) = r \cdot x \cdot D_m(e,s,z). \tag{2.14}$$

Lastly, the private enterprise's profit generated from these commuters is $((k-c)[1-k\alpha]^+ - z)D_m(e, s, z)$. Observe that the optimal price $k_g = \frac{\alpha c+1}{2\alpha}$ so that the corresponding optimal

profit is $(K - z)D_m(e, s, z)$, where K is the enterprise's profit margin in the base model as defined in Section 2.2.1. We assume the profit margin of a participating enterprise K satisfies $K > \underline{K}$ throughout the chapter, where $\underline{K} = (d - p) - (r - d)x$, because enterprises with low profit margin would not opt in the subsidy program. Hence, the enterprise's total profit under incentives (e, s, z) is

$$\Pi_s(e, s, z) = (K - z) \cdot D_m(e, s, z). \tag{2.15}$$

Note that when $(e, s, z) = (0, 0, 0)$, the metrics $W(e, s, z)$, $\Pi_p(e, s, z)$, $\Pi_r(e, s, z)$, and $\Pi_s(e, s, z)$ reduce to the base values W^0, Π_p^0, Π_r^0, and Π_s^0 (defined in Section 2.2.1), respectively.

Next, we present the evaluation criteria for the incentive program (e, s, z). In particular, we assume that (e, s, z) is set according to the following criteria:

I **Minimum operating cost.** Because most municipalities in the U.S. and around the world are budget constrained with respect to public transit,[15] we assume that the government has a strong desire to minimize the total operating cost of the incentive program, but not profit from it, where the government's operating cost is given by $[C(e, s, z)]^+$.

II **Increase transit ridership.** The intent of the incentive (e, s, z) is to increase public transit ridership among commuters by a factor of $\beta > 0$. We assume throughout that β is an exogenous adoption target, and that the incentive (e, s, z) must satisfy:

$$D_m(e, s, z) - D_m^0 \geq \beta D_m^0,$$

where D_m^0 is the baseline mixed mode adoption without any intervention as stated in Section 2.2.1. Note that the increase β is equivalent to a decrease in demand for

[15]Due to an anticipated budget deficit through 2026, the U.S. Department of Transportation is likely to further reduce transfers to states and transit agencies (Kirk and Mallett 2020). Similarly, Transport for London faces a grant cut from the government of £700m a year (Edwards 2018).

driving; therefore, β may be equivalently interpreted as the target reduction in traffic congestion. For this reason, the constraint above can also be interpreted as improving the well-being of city dwellers. Analogous to the assumption that $x \in (\underline{x}, \bar{x})$ in the base model, we assume throughout that the target β is restricted to the non-degenerate case where $D_d(e, s, z) > 0$ and $D_m(e, s, z) > 0$. To enforce the preceding inequalities, it suffices to assume that $\beta < \bar{\tau}(x) \equiv \frac{(\delta-1)(x+1)}{(d-p)-(r-d)x} - 1$. Note that if $\beta \geq \bar{\tau}(x)$, then upon implementation of incentive (e, s, z), no commuters will drive ($D_d(e, s, z) = 0$), which is unlikely to occur in practice.

III **Commuter participation.** To generate public support for the incentive program, commuters should not be worse off collectively: $W(e, s, z) \geq W^0$.

IV **Ride-hailing platform participation.** To form a viable partnership with the ride-hailing platform, the platform's revenue generated from daily commuters should not decrease: $\Pi_r(e, s, z) \geq \Pi_r^0$.

V **Public-transit participation.** To ensure that the public transit agency is not negatively impacted, its revenue generated from the mixed mode should not decrease: $\Pi_p(e, s, z) \geq \Pi_p^0$.[16]

VI **Enterprise participation.** Lastly, to prevent adverse effects on the local economy, it is desirable to ensure that the incentive (e, s, z) does not reduce the private enterprise's profit: $\Pi_s(e, s, z) \geq \Pi_s^0$. Further, in mechanism [I] in particular, to entice participation of the private enterprise, the private subsidy z is set to maximize the private enterprise's profit: $z = \arg\max_z \Pi_s(e, s, z)$.

Observe from criterion II that a feasible incentive program (e, s, z) must satisfy $D_m(e, s, z) \geq D_m^0$. Combine this observation with (2.13) and (2.14), criteria IV and V are satisfied (and

[16]Note that although transit agencies receive funding from local governments, they operate under separate budgets (Federal Transit Administration 2020). Therefore, we model the transit agency as a distinct stakeholder from the local government.

hence, redundant) immediately if $D_m(e, s, z) \geq D_m^0$. This result is intuitive: when the government increases the mixed mode adoption, both the ride-hailing company and the transit agency will be better off. Therefore, we shall omit criteria IV and V from the incentive design problems in the remainder of the chapter for conciseness.

2.3 Direct Mechanism: Commuter Subsidies and Congestion Fees

In the direct mechanism [D], the local government directly offers a subsidy s to each mixed mode commuter, and charges a congestion fee e to each commuter who drives. Hence, by letting $z = 0$ in the general formulation presented in Section 2.2.2, we can determine the commuter utility and the market demand associated with the driving mode and the mixed mode, along with all the performances metrics under mechanism [D]. Figure 2.2 depicts the relationship among the various stakeholders in this case.

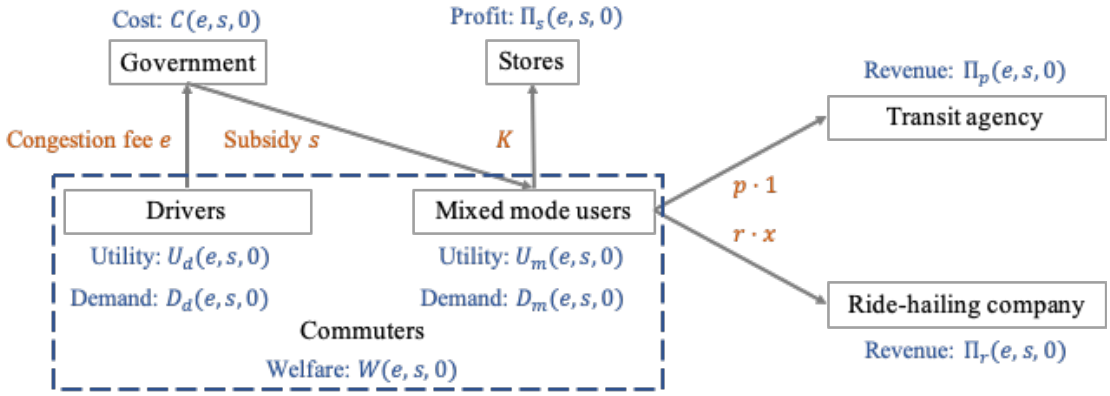


Figure 2.2: Strategic interactions among stakeholders under the indirect mechanism [D].

Based on the evaluation criteria I $\sim$ VI (with IV and V being omitted for conciseness),

the optimal incentives e^* and s^* are the solution to the following problem:

$$\min_{e,s\geq 0} \; [C(e,s,0)]^+ \tag{2.16a}$$

$$\text{s.t. } D_m(e,s,0) - D_m^0 \geq \beta D_m^0 \tag{2.16b}$$

Mechanism [D]:

$$W(e,s,0) \geq W^0 \tag{2.16c}$$

$$\Pi_s(e,s,0) \geq \Pi_s^0. \tag{2.16d}$$

2.3.1 Direct Mechanism: Optimal Incentive (e^*, s^*)

By solving problem (2.16), we obtain the optimal incentive (e^*, s^*) under mechanism [D].

Proposition 1. *The optimal congestion fee e^* and the optimal subsidy s^* under mechanism [D] are given by:*

$$\begin{aligned} e^* &= \frac{\beta(\beta+2)((d-p)-(r-d)x)^2}{2(\delta-1)(x+1)}, \\ s^* &= \frac{\beta((d-p)-(r-d)x)(2(\delta-1)(x+1)-(\beta+2)((d-p)-(r-d)x))}{2(\delta-1)(x+1)}. \end{aligned} \tag{2.17}$$

Further,

(i) The optimal congestion fee e^* strictly increases in β and strictly decreases in x.

(ii) The optimal subsidy s^* strictly increases in β. If δ is large, s^* strictly decreases in x. However, if δ is small, there exists $\tilde{x}$ such that s^* increases on $x < \tilde{x}$ and decreases on $x \geq \tilde{x}$.

Statements (i) and (ii) of Proposition 1 imply that, as the mixed mode adoption target β becomes more aggressive, the commuter subsidy s^* and congestion fee e^* also increase to meet the adoption target, as expected. Additionally, as the last mile distance x increases, the demand for driving also increases, and hence the congestion fee e^* decreases to maintain commuter welfare.

In contrast to the congestion fee e^*, the behavior of the optimal subsidy s^* is not monotonic in the last mile distance x. To see why, let us consider the impact on mechanism [D] when x increases. As x increases, the mixed mode adoption (before any government intervention) D_m^0 (as stated in (2.3)) decreases, which generates two competing effects. First, note that the transit ridership constraint can be rewritten as $D_m(e, s, 0) \geq (\beta+1)D_m^0$. Therefore, as D_m^0 decreases, the requirement for the mixed mode commuters $D_m(e, s, 0)$ becomes less stringent. Consequently, the government can afford to reduce s, which we refer to as the "*adoption effect*." Second, as D_m^0 decreases, commuter welfare decreases, due to fewer commuters receiving the subsidy. To ensure commuters are not worse off, the government needs to increase s, which we refer to as the "*welfare effect*." Hence, whether the optimal subsidy s^* increases or decreases in x depends on which of these two effects dominate.

To examine when one effect dominates the other, first consider the case when commuters strongly prefer driving over transit and ride-hailing (i.e., when δ is large). In this case, the welfare effect is weak (because the mixed mode adoption D_m^0 is already low before any intervention), and the adoption effect dominates – leading s^* to decrease in x. However, when commuters are relatively indifferent between the two modes (i.e., when δ is small), the mixed mode adoption D_m^0 is highly sensitive to small changes in x. As a result, as x increases, demand for the mixed mode drops sharply, which makes the welfare effect dominates – leading s^* to increase in x.[17]

Corollary 1. *There exists a threshold $\hat{x}$ such that $e^* \geq s^*$ in mechanism [D] if and only if $x \leq \hat{x}$. Further, there exists a threshold $\bar{\delta} > 1$ such that $s^* > e^*$ in mechanism [D] for all $x \in (\underline{x}, \bar{x})$ if and only if $\delta \geq \bar{\delta}$.*

Corollary 1 states that for any adoption target β, the optimal congestion fee e^* is comparatively larger than the optimal subsidy s^* when the last mile distance x is small. To see why, observe that when x is small, the mixed mode is already attractive to commuters, which

[17]Note that even when δ is small, the welfare effect ceases to dominate the adoption target effect when the last mile distance is large ($x \geq \tilde{x}$), due to low demand for the mixed mode.

makes providing a subsidy (per commuter) relatively costly; in this setting, the congestion fee e is the preferred lever for promoting the adoption of the mixed mode, due to its cost efficiency. Conversely, when the last mile distance x is large, more aggressive subsidies are required to meet the mixed mode adoption target. Corollary 1 also implies that if δ is large, then $\hat{x} < \underline{x}$, which yields $s^* > e^*$ for all $x \in (\underline{x}, \bar{x})$. This occurs because when δ is large, most commuters prefer to drive, and so it is sub-optimal for the government to set a high congestion fee, due to its degradation of commuter welfare.

Next, we examine whether the government can attain budget neutrality (or positive revenue) by using congestion fees to offset commuter subsidies under mechanism [D].

Corollary 2. *It is impossible for mechanism [D] to be budget neutral, i.e., $C(e^*, s^*, 0) > 0$. Further, the minimal operating cost $C(e^*, s^*, 0)$ is higher when β is large or when x is small.*

The intuition behind Corollary 2 is as follows. For any budget neutral incentive (e, s) (i.e., where the congestion fee e is selected to fully cover the cost of the subsidy s), it can be shown that the utility loss to drivers exceeds the utility gain to transit riders. In other words, for any subsidy s, there is no congestion fee e that can achieve budget neutrality without a net decrease in total commuter welfare. As a result, the requirement that commuter welfare be non-decreasing in mechanism [D] can only be satisfied through a positive operating cost. Next, to see why $C(e^*, s^*, 0)$ is higher when β is large or when x is small, observe that, in this case, the mixed mode adoption target βD_m^0 on the right hand side of constraint (2.16b) is large, which requires the mixed mode adoption $D_m(e, s, 0)$ to also be large. Therefore, as more mixed mode commuters receive the subsidy, mechanism [D] incurs a greater cost to the government. We can now summarize our main finding from mechanism [D] as follows.

Remark 1. *The direct mechanism [D] can enable the government to meet the public transit adoption target β. However, this mechanism is costly: attaining budget neutrality is impossible without undermining commuter welfare.*

Note that our commuter utility function implies that commuters are equally sensitive to the subsidy and congestion fee; that is, $|\partial U_m(e,s,0)/\partial s| = |\partial U_d(e,s,0)/\partial e|$. Note that this symmetry may not hold if commuters are loss averse. However, we find that the structure of our main results continues to hold under loss aversion as well. Intuitively, if commuters perceive losses from congestion fees more strongly than gains from subsidies, then the government would need to *increase* the subsidy and *decrease* the congestion fee to satisfy the non-decreasing commuter welfare constraint, resulting in a higher total operating cost.[18]

Remark 1 naturally raises the following question: Does there exist an alternate mechanism that can enable the government to attain budget neutrality, without negatively impacting commuter welfare? In the next section, we address this question by examining an alternative incentive program under which the subsidy is funded by a private enterprise, instead of congestion fees.

2.4 Indirect Mechanism: Funding Subsidies Through a Private Enterprise

Because most municipalities are financially constrained, self-funded or budget neutral projects are preferred. In this section, we present the *indirect* mechanism [I], where the local government secures funding for the subsidy entirely from a private enterprise,[19] and does not charge congestion fees. Note that the indirect mechanism is budget neutral by design. We will therefore focus on the performance of mechanism [I] on the remaining evaluation criteria.

[18]Similar to Homonoff (2018), we can extend our model to capture one aspect of loss aversion by defining the commuter utilities associated with the mixed mode and driving as $U_m(e,s,z) = U_m^0 + s + z$ and $U_d(e,s,z) = U_d^0 - \rho \cdot e$, respectively, where $\rho \geq 1$ captures the strength of the commuter's aversion to the congestion fee. The evaluation criteria I – VI can be re-defined accordingly. Using a parallel argument to the proof of Corollary 2, it is straightforward to verify that the optimal operating cost is strictly positive for all $\rho \geq 1$.

[19]As explained in Section 2.1, the transit agency can generate revenue from private companies who post in-app video ads via mobile apps, and the indirect mechanism can be operationalized through a mobile app that allows the government to collect fares from riders and subsidies from the private sector.

By using the same approach as in Section 2.3, we now present our model associated with mechanism [I], which corresponds to the special case of the general model framework presented in Section 2.2.2 when we set $(e, s) = (0, 0)$. In other words, by replacing (e, s, z) in the general formulation with $(0, 0, z)$, we can determine the performance metrics under mechanism [I]. The interactions among all stakeholders are depicted in Figure 2.3.

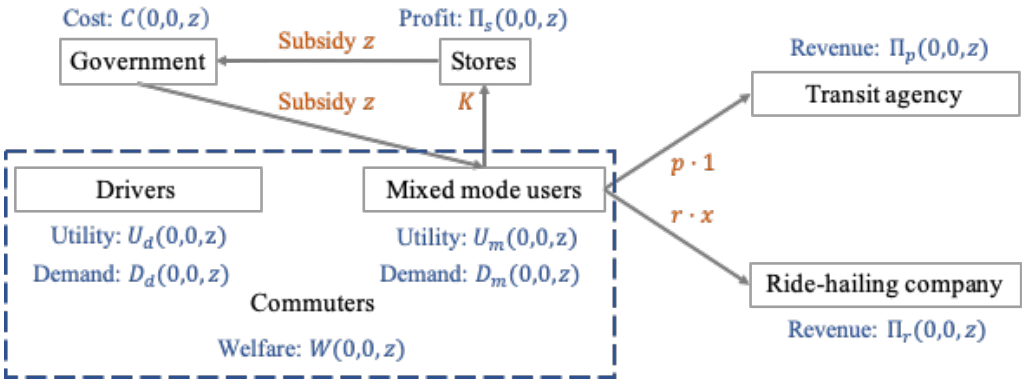


Figure 2.3: Strategic interactions among stakeholders under the indirect mechanism [I].

Because the viability of mechanism [I] depends on participation from the private enterprise, we assume the private subsidy z is set to maximize the enterprise's total profit. Based on the evaluation criteria I – VI, the government's incentive design problem associated with subsidy z under mechanism [I] can be formulated as the following optimization problem:

$$\min_{z \geq 0} \ [C(0,0,z)]^+ \tag{2.18a}$$

$$\text{s.t.} \ \ D_m(0,0,z) - D_m^0 \geq \beta D_m^0 \tag{2.18b}$$

Mechanism [I]:

$$W(0,0,z) \geq W^0 \tag{2.18c}$$

$$\Pi_s(0,0,z) \geq \Pi_s^0 \tag{2.18d}$$

$$z = \arg\max_z \Pi_s(0,0,z) \tag{2.18e}$$

Note that $C(0,0,z) = 0$ for all z, by the definition of $C(e,s,z)$ as given in (2.11). Therefore, under mechanism [I], the government attains budget neutrality as long as problem (2.18)

is feasible (in other words, mechanism [I] is budget neutral by design). Note also that formulation (2.18) may be infeasible if β is large, for the following reason. If β is large, then the private enterprise must issue a large subsidy to reach the adoption target. However, if the cost of providing a large subsidy exceeds the extra revenue earned by the enterprise (due to increased demand), then the participation constraint (2.18d) may be violated. Formally, let $\tau(x) = \min\{\frac{K}{((d-p)-x(r-d))} - 1, \bar{\tau}(x)\}$, where $\bar{\tau}(x) = \frac{(\delta-1)(x+1)}{(d-p)-(r-d)x} - 1$ as defined in Section 2.2.2. Note that $0 < \tau(x) \leq \bar{\tau}(x)$.

2.4.1 Indirect Mechanism: Optimal Subsidy z^*

By solving problem (2.18), we obtain the optimal subsidy z^* in the following proposition.

Proposition 2. *Mechanism [I] is feasible if and only if $\beta \leq \tau(x)$. If $\beta \leq \tau(x)$, then the optimal subsidy is given by:*

$$z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\}, \tag{2.19}$$

where $\underline{z} = \beta((d-p) - x(r-d))$, $\hat{z} = \frac{1}{2}(K - ((d-p) - x(r-d)))$ and $\bar{z} = \min\{K, (\delta-1)(x+1)\} - ((d-p) - x(r-d))$. The minimal cost is 0, $C(0, 0, z^*) = 0$. Further,

(i) If the enterprise's profit margin K is high, the optimal subsidy z^* strictly increases in x.

(ii) If the enterprise's profit margin K is low, there exists $\tilde{x}$ such that the optimal subsidy z^* strictly decreases on $x < \tilde{x}$ and strictly increases on $x \geq \tilde{x}$.

The intuition behind Proposition 2 is similar to that of Proposition 1. To elaborate, note that as the last mile distance x increases, the mixed mode adoption (before any government intervention) D_m^0 decreases, which generates two competing effects. First, the "adoption effect" (as described in Section 2.3.1) continues to play a role here, meaning the requirement for the corresponding mixed mode commuters $D_m(0, 0, z)$ becomes less stringent as D_m^0

decreases. Consequently, the private enterprise can afford to reduce the subsidy z. Second, as D_m^0 decreases, the private enterprise has incentive to increase z, so that it can recoup the subsidy cost by increasing demand (i.e, a higher $D_m(0,0,z)$). We shall refer to the latter effect as the "*customer effect*." Note that the customer effect plays a role in mechanism [I] – because the subsidy z is paid by the private enterprise – but is absent in mechanism [D]. Hence, whether the optimal subsidy z^* increases or decreases in x depends on which effect dominates.

To examine which effect dominates, let us consider the case where the enterprise's profit margin K is high. In this case, the enterprise's profit is highly sensitive to changes in D_m^0. Therefore, the private enterprise has a stronger incentive to offer a higher subsidy to boost the customer base $D_m(0,0,z)$, which makes the customer effect dominate the adoption effect. As a result, the subsidy z^* increases in x when K is large. The intuition behind the decrease in z^* when K is small is similar, which we omit for conciseness.

2.4.2 Comparison: Direct Mechanism [D] versus Indirect Mechanism [I]

By considering Proposition 1 and Proposition 2 presented in Section 2.3.1 and Section 2.4.1, we can compare the performance of the direct and indirect mechanisms according to each of the six metrics discussed in Section 2.2.2. For ease of reference, we shall use superscript D and I to denote quantities obtained at optimal solutions in mechanisms [D] and [I], respectively. Also, in preparation, let $\tilde{\tau}(x) = \frac{\hat{z}^2-(z^*-\hat{z})^2}{K((d-p)-(r-d)x)}$. The following result compares the performance of the two mechanisms assuming $\beta \leq \tau(x)$.

Theorem 1. *Mechanism [I] outperforms mechanism [D] in terms of operating cost ($C^I = 0 < C^D$), mixed mode adoption ($D_m^I \geq D_m^D$), commuter welfare ($W^I > W^D$), ride-hailing platform profit ($\Pi_r^I \geq \Pi_r^D$), and transit agency revenue ($\Pi_p^I \geq \Pi_p^D$). Further,*

(i) if $\beta \leq \tilde{\tau}(x)$, mechanism [I] outperforms mechanism [D] on enterprise profit ($\Pi_s^I \geq \Pi_s^D$),

(ii) if $\beta \in (\tilde{\tau}(x), \tau(x)]$, mechanism [D] outperforms mechanism [I] on enterprise profit

($\Pi_s^I < \Pi_s^D$*).*

Theorem 1 states that the indirect mechanism [I] dominates the direct mechanism [D] in *all* performance metrics when the adoption target is not too large, $\beta \leq \tilde{\tau}(x)$. However, in the intermediate case where $\tilde{\tau}(x) < \beta \leq \tau(x)$, the direct mechanism [D] will enable the private enterprise to generate a higher profit.

Recall that the main difference between mechanisms [I] and [D] is the funding source for the commuter subsidy. In mechanism [D], the government funds the subsidy by imposing congestion fees on drivers; in mechanism [I], the government simply coordinates the funding from the private sector partner, and is otherwise passive and budget neutral. Theorem 1 highlights the benefits of a public-private partnership over congestion fees: in addition to achieving budget neutrality, the government can both improve commuter welfare (by avoiding congestion fees on drivers) and support the private sector partner (by boosting demand). As a result, all stakeholders are left better off under mechanism [I] compared to [D], assuming the adoption target β is not too large. Moreover, note that mechanism [I] can dominate mechanism [D] despite containing only a single lever (the subsidy z). This result highlights that the gains from mechanism [I] are due to the coordination with the private enterprise.

Also, recall from Proposition 2 that when the adoption target is high $\beta > \tau(x)$, formulation (2.18) is infeasible, meaning there is no subsidy level z that can satisfy all of the criteria I – VI in Section 2.2.2. This observation enables us to make the following summarizing remark:

Remark 2. *If the public transit adoption target is aggressive,* $\beta > \tau(x)$*, then the government should adopt mechanism [D] over mechanism [I]. If the adoption target is conservative,* $\beta \leq \tau(x)$*, the government should adopt mechanism [I] over mechanism [D] (even though the private enterprise can earn higher profit under mechanism [D] when* β *satisfies* $\tilde{\tau}(x) < \beta \leq \tau(x)$*).*

The above remark has the following implications. When the municipal government has a

very tight or no budget, the government should set a conservative adoption target $\beta \leq \tau(x)$. In this case, the government should adopt the indirect mechanism [I], which leaves all parties better off, and critically, does not require increased spending by the government. However, if the local government has ample funding, then it is feasible to set an aggressive transit adoption target level $\beta \geq \tau(x)$, in which case the government should adopt and fund mechanism [D].

2.5 Alternative Formulation: Maximizing Commuter Welfare Under Budget Neutrality

In Section 2.3 and Section 2.4, we analyse the optimal incentive schemes under mechanisms [D] and [I] by minimizing operating cost $C(e, s, z)$. In this section, we analyze mechanisms [D] and [I] by considering an alternative objective: maximizing commuter welfare while maintaining budget neutrality. (In Appendix A.2, we consider a different alternative model formulation for the case when the transit agency has an exogenous given earmark budget B that is reserved for supporting various schemes for increasing public transit ridership. Specifically, we consider two different incentive design problems associated with mechanisms [D] and [I] by maximizing public transit adoption subject to a budget constraint. As shown in Appendix A.2, we obtain the same structural results under both mechanisms.)

2.5.1 Alternative Direct Mechanism [D-A]

Recall from Remark 1 in Section 2.3.1 that it is not possible to implement the direct mechanism [D] in a budget neutral manner without undermining commuter welfare. However, when a financially constrained municipality finds it too costly to implement mechanism [D], some loss in commuter welfare may be tolerated (especially if there is public support for policies that lower carbon emissions and traffic congestion). This observation motivates us to modify the original incentive design problem (2.16) by maximizing commuter welfare

subject to a budget neutrality constraint (2.20c). Specifically, the modified problem can be formulated as:

$$\max_{e,s\geq 0} \; (W(e,s,0) - W^0) \tag{2.20a}$$

$$\text{s.t. } D_m(e,s,0) - D_m^0 \geq \beta D_m^0 \tag{2.20b}$$

Mechanism [D-A]:

$$C(e,s,0) = 0 \tag{2.20c}$$

$$\Pi_s(e,s,0) \geq \Pi_s^0, \tag{2.20d}$$

where D_m^0, W^0 and Π_s^0 are the demand for the mixed mode, commuter welfare and profit of the private enterprise in the base model as defined in (2.3), (2.4) and (2.7), respectively. Analogous to Proposition 1, the next result characterizes the optimal incentive scheme under mechanism [D-A].

Proposition 3. *The optimal congestion fee e^* and the optimal subsidy s^* under the alternative direct mechanism [D-A] are given by:*

$$e^* = \frac{\beta(\beta+1)((d-p)-(r-d)x)^2}{(\delta-1)(x+1)}$$
$$s^* = \frac{\beta((d-p)-(r-d)x)((\delta-1)(x+1)-(\beta+1)((d-p)-(r-d)x))}{(\delta-1)(x+1)}$$

Further,

(i) The optimal congestion fee e^* increases in β and decreases in x.

(ii) If δ is large, s^* strictly decreases in x. However, if δ is small, there exists $\tilde{x}$ such that s^* increases on $x < \tilde{x}$ and decreases on $x \geq \tilde{x}$.

(iii) There exists $\tilde{\beta}$ such that s^* increases on $\beta < \tilde{\beta}$ and decreases on $\beta \geq \tilde{\beta}$.

The above results resemble Proposition 1, except that the optimal subsidy s^* under the alternate mechanism [D-A] is no longer always increasing in β. Instead, the optimal subsidy s^* decreases β when the adoption target β is large. This difference is driven by the budget

neutrality constraint (2.20c). To elaborate, consider the case when β is large. In this case, the right hand side of (2.20b) is large, which requires the mixed mode demand $D_m(e, s, 0)$ to be large. As more mixed mode commuters collect the subsidy, the government must reduce the subsidy s in order to maintain budget neutrality, which leads s^* to decrease in the adoption target β.

Next, similar to Corollary 1, we compare the relative size of the optimal congestion fee and subsidy.

Corollary 3. *There exists $\hat{x}$ such that $e^* \geq s^*$ in mechanism [D-A] if and only if $x \leq \hat{x}$. Further, there exists $\bar{\delta} > 1$ such that $s^* > e^*$ in mechanism [D-A] for all $x \in (\underline{x}, \bar{x})$ if and only if $\delta \geq \bar{\delta}$.*

Corollary 3 is analogous to Corollary 1: for any β, the government should set a higher congestion fee and a lower subsidy when the last mile distance x is small, and vice versa when x is large.

Recall from Corollary 2 in Section 2.3.1 that the direct mechanism [D] is costly (i.e., $C(e^*, s^*, 0) > 0$) due to the commuter welfare constraint (2.16c). Next, we present a counterpart to Corollary 2: to maintain budget neutrality, commuter welfare degradation is unavoidable under the alternative direct mechanism [D-A].

Corollary 4. *The alternative direct mechanism [D-A] will always result in commuter welfare degradation; i.e., $W(e^*, s^*, 0) < W^0$. Further, the reduction in commuter welfare, $W^0 - W(e^*, s^*, 0)$, under mechanism [D-A] is higher when the adoption target β is large and when the last mile distance x is small.*

Corollary 4 reveals that mechanism [D-A] degrades commuter welfare more severely when the adoption target β is large. To elaborate, consider the case when the adoption target β is large. In this case, the corresponding mixed mode demand $D_m(e, s, 0)$ has to be large in order to meet the adoption target. As more mixed mode commuters collect the subsidy, the

government must reduce the subsidy in order to maintain budget neutrality, which causes the commuter welfare to deteriorate. Next, we consider why commuter welfare reduction is higher when x is small. In this case, the mixed mode adoption (before any government intervention) D_m^0 as stated in (2.3) is large, which makes the requirement for the mixed mode adoption $D_m(e, s, 0)$ also large. Similar to the large β case, as more mixed mode commuters receive the subsidy, the government must increase the congestion fee e in order to maintain budget neutrality, which reduces commuter welfare.

2.5.2 Alternative Indirect Mechanism [I-A]

In the same spirit as Section 2.5.1, we now consider a variation of mechanism [I], which we shall refer to as mechanism [I-A], where the objective is to maximize commuter welfare while maintaining budget neutrality. To do so, we modify our original program (2.18) presented in Section 2.4 as follows:

$$\max_{z \geq 0} \ (W(0,0,z) - W^0) \tag{2.21a}$$

$$\text{s.t. } D_m(0,0,z) - D_m^0 \geq \beta D_m^0 \tag{2.21b}$$

Mechanism [I-A]:

$$C(0,0,z) = 0 \tag{2.21c}$$

$$\Pi_s(0,0,z) \geq \Pi_s^0 \tag{2.21d}$$

$$z = \arg\max_z \Pi_s(0,0,z) \tag{2.21e}$$

In parallel with Proposition 2, the following proposition characterizes the optimal subsidy under the alternative indirect mechanism [I-A].

Proposition 4. *The alternative indirect mechanism [I-A] is feasible if and only if $\beta \leq \tau(x)$. When $\beta \leq \tau(x)$, the optimal subsidy is given by*

$$z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\},$$

where $\tau(x)$ is defined in Section 2.4 and $\underline{z}, \hat{z}, \bar{z}$ are defined in Proposition 2. The maximal welfare increment is strictly positive so that $W(0,0,z^*) > W^0$.

Note the optimal subsidy z^* under mechanism [I-A] is equal to that under mechanism [I], as presented in Proposition 2. This is because constraints (2.21b), (2.21d) and (2.21e), which are the same as constraints (2.18b), (2.18d) and (2.18e), uniquely determine feasible private subsidy z^* and $C(0, 0, z^*) = 0$ by definition in (2.11). Thus z^* is feasible and thus optimal to both problems (2.18) and (2.21). Also, unlike mechanism [D-A] with deteriorating commuter welfare as stated in Corollary 4, Proposition 4 exerts that, when $\beta \leq \tau(x)$, mechanism [I-A] will always enhance commuter welfare.

2.5.3 Comparison: Two Alternative Mechanisms [D-A] and [I-A]

We now compare the performance of the two alternative mechanisms [D-A] and [I-A]. Using the same approach as presented in Section 2.4.2, we can retrieve the relevant performance metrics from (2.10)– (2.15) by substituting (e, s, z) with $(e^*, s^*, 0)$ and $(0, 0, z^*)$.

Theorem 2. *The alternative indirect mechanism [I-A] and the alternative direct mechanism [D-A] both attain budget neutrality ($C^I = C^D = 0$). However, alternative indirect mechanism [I-A] outperforms alternative direct mechanism [D-A] in terms of mixed mode adoption ($D_m^I \geq D_m^D$), commuter welfare ($W^I > W^0 > W^D$), ride-hailing platform profit ($\Pi_r^I \geq \Pi_r^D$), and transit agency revenue ($\Pi_p^I \geq \Pi_p^D$). Further,*

(i) if $\beta \leq \tilde{\tau}(x)$, mechanism [I-A] outperforms mechanism [D-A] in terms of enterprise profit ($\Pi_s^I \geq \Pi_s^D$),

(ii) if $\beta \in (\tilde{\tau}(x), \tau(x)]$, mechanism [D-A] outperforms mechanism [I-A] in terms of enterprise profit ($\Pi_s^I < \Pi_s^D$),

where $\tau(x)$ and $\tilde{\tau}(x)$ are defined in Section 2.4 and Section 2.2.2, respectively.

Using the above proposition, we summarize our observations in the following remark.

Remark 3. *Suppose in addition to budget neutrality, the government prioritizes commuter welfare. Then the government should adopt the alternative direct mechanism [D-A] when the*

adoption target satisfies $\beta > \tau(x)$, and adopt the alternative indirect mechanism [I-A] when $\beta \leq \tau(x)$.

Remark 2 and Remark 3 imply that the dominance of the indirect mechanism [I] is robust, regardless of whether the goal is to maximize commuter welfare or to minimize operating cost. Therefore, when the municipal government is financially constrained, it is advisable for the government to set a conservative transit adoption target $\beta \leq \tau(x)$, and adopt the indirect mechanism [I]. However, if the government has sufficient funding, or if the commuters are willing to accept lower welfare (e.g., in support of a reduction in carbon emissions and traffic congestion), then the government can afford to set an aggressive adoption target $\beta > \tau(x)$ and adopt the direct mechanism [D].

2.6 Conclusion

Motivated by the increased focus on public transit and urban mobility by municipal governments, we have analyzed two mechanisms for improving public transit ridership. Both mechanisms are intended to address the "last mile gap" by providing subsidies to commuters who adopt a mixed mode of transportation that combines ride-hailing with public transit. The main differences between the two mechanisms are the role of the municipal government and the source of funding for the subsidy. In the direct mechanism [D], the government provides a ride-hailing subsidy to commuters that adopt the mixed mode, and charges a congestion fee to commuters who travel by personal vehicle. The congestion fee is used to offset the cost of the subsidy, and also serves as an additional incentive for transit adoption (by making driving more costly). In the indirect mechanism [I], the government partners with a private enterprise that provides the funding for the subsidy, and does not charge a congestion fee. We present analytical results that characterize the optimal incentives under these two mechanisms, as well as the impact on all involved stakeholders.

Our findings offer several prescriptions for policy makers who are interested in increasing

public transit ridership through partnerships with the private sector. First, in the direct mechanism, it's advisable to set large congestion fee and small subsidy when public transit coverage is high (i.e. when the last mile distance is small), and conversely, when the public transit coverage is low, the optimal incentive requires more aggressive subsidies. This result is driven by the opposing effects that subsidies and congestion fees have on commuter welfare and operating cost. We also find that the dependence of the optimal subsidy on the transit coverage level depends on the relative preference between driving and taking public transit. In particular, we find that when commuters strongly prefer driving, jurisdictions with larger last mile distances should set lower subsidies, but this behavior can be reversed if commuters only slightly prefer driving over public transit.

Second, we find that the government cannot fully recover the cost of providing subsidies by collecting congestion fees. Specifically, we show that the direct mechanism cannot be budget neutral unless the government is willing to accept a decrease in commuter welfare. This suggests that attempting to implement commuter subsidies and congestion fees in a budget neutral manner may be ill-advised if the government is sensitive to commuter welfare. However, in the event that the government can obtain subsidy funding from a private enterprise (who benefits from increased foot traffic at the transit station or increased views of an advertisement on a mobile app) in lieu of charging congestion fees (as in the indirect mechanism), we show that public transit ridership can be increased without degrading commuter welfare. However, because the implementation of an indirect mechanism requires the participation of the private enterprise, we find that this indirect mechanism is only viable if the mixed mode adoption target is modest. In other words, only the direct mechanism can enable the government to achieve an ambitious adoption target, and the implementation of the direct mechanism is not budget neutral: it will require extra funding.

Third, although the government has one less "lever" in the indirect mechanism [I] (due to the absence of the congestion fee), it can dominate the direct mechanism [D] if the adoption target is modest. Specifically, when the adoption target is modest, the indirect mechanism

can benefit all stakeholders: the government, commuters, the public transit agency, the ride-hailing platform, and the private enterprise. This finding suggests that, for jurisdictions that wish to increase transit ridership, but are severely budget constrained, it may be more fruitful to fund ride-hailing subsidies through partnerships with the private sector than by charging congestion fees. Moreover, this dominance result is robust to an alternate specification of mechanism [I] (namely, when the objective is to maximize commuter welfare instead of enterprise profit).

As an initial attempt to explore different incentive mechanisms for improving public transit ridership, our model has several limitations that we wish to highlight. First, for tractability, we restrict attention to two modes of travel – driving and the mixed mode. Although other means of commuting are less common (such as commuting by ride-hailing), including them as travel modes may affect our results. Second, subsidies for ride-hailing trips may cause commuters to switch from other last-mile travel modes, such as walking, biking, or bussing, which may have consequences not discussed in this chapter. Further, to focus on the economic viability of different incentive mechanisms, we have abstracted away from operational issues such as capacity, traffic congestion, and the spatial nature of public transit infrastructure. Considering these features may yield additional insight into the design of incentive mechanisms for improving transit ridership. We shall defer these issues to future research.

CHAPTER 3

Planning Bike Lanes: Ridership, Congestion, and Path Selection

3.1 Introduction

Urbanization is accelerating globally. By 2050, an estimated 68% of the world's population is expected to live in cities, up from 55% in 2018 (UN 2019). Urbanization brings numerous benefits, including improved access to employment opportunities, education, and social services. However, these benefits are typically accompanied by increased traffic congestion, which threatens the sustainability of cities (Bertinelli and Black 2004, Çolak et al. 2016). In addition to its negative effects on the environment and health, congestion is estimated to incur economic losses of $85 billion annually across U.S. cities (US DOT 2009).

Alleviating traffic congestion is likely to require a multi-faceted solution, including increased adoption of sustainable travel modes (e.g., walking, cycling, public transit) and associated investments in urban infrastructure (US DOT 2017). In alignment with these goals, in 2021 the U.S. Congress enacted the Infrastructure Investment and Jobs Act (IIJA), a $1.2 trillion spending package that, among other priorities, allocates funding to promote sustainable transportation and invest in the necessary infrastructure (The White House 2021, 117th Congress 2021). Urban infrastructure is also a concern globally – one of the 17 Sustainable Development Goals proposed by the United Nations is to make cities "inclusive, safe, resilient and sustainable", which requires a robust transportation infrastructure (UN 2015).

Increasing cycling ridership has been proposed as a solution to reducing traffic congestion (Thorsten and Rudolph 2016, Hamilton and Wichman 2018). Cycling is one of the most sustainable modes of urban transport – it causes minimal environmental damage, is cost effective compared to car ownership, and promotes health through physical activity. Widespread cycling ridership is also associated with substantially increased road safety for all users, with fatal crash rates being estimated to be 44% lower than the US average in cities with abundant bike lanes (Marshall and Ferenchak 2019). Demand for cycling is also growing rapidly – from 2001 to 2017, the number of cycling trips in the US is estimated to have doubled from 1.7 billion to 3.6 billion (National Household Travel Survey 2017), and U.S. cities saw a 21% surge in cycling trips during the COVID-19 pandemic (LA Times 2020).

To meet growing demand and promote cycling adoption, municipal governments have recently made major investments into installing new bike lanes.[1] For example, between 2014 to 2019, New York City increased total bike lane mileage from 900 to 1,240 (DOT 2019), and in 2021 Chicago announced a plan to expand their bike lane network by 100 miles over the following two years (Chicago DOT 2021). The U.S. federal government has also committed to improving cycling infrastructure nationwide – the Transportation Alternatives Program, which receives federal funding for bike lanes and pedestrian paths, has seen its budget increase from $850 million to $1.44 billion annually under the IIJA (Mills 2021).

Despite the benefits of boosting cycling ridership, the expansion of bike lane networks remains contentious because it necessarily reduces road capacity for automotive vehicles. Intuition suggests this re-allocation may have the undesirable effect of *amplifying* congestion (BBC 2021), making bike lane planning a thorny issue for city planners (Khany 2022). Despite major investments in bike lanes and the related controversy, few previous studies have attempted to rigorously model or estimate the effect of bike lanes on congestion. Doing

[1]There is evidence that installing bike lanes increases cycling adoption (Parks et al. 2012, Strauss and Miranda-Moreno 2013, Monsere et al. 2014, Mitra et al. 2017, Khany 2022).

so is non-trivial due to the opposing forces alluded to above: While bike lanes may increase the attractiveness of cycling and take cars off the road, they also risk exacerbating congestion by reducing capacity in critical segments of the road network. As a result, modeling the net impact of bike lanes on congestion requires careful accounting of both of these effects.

The effect of bike lanes on congestion and cycling ridership also depends on planning decisions. Although bike lanes may not necessarily worsen congestion (Johnson and Johnson 2014), they may only have a minimal effect on cycling ridership if the locations are poorly chosen. However, data-driven models for bike lane planning remain scarce, leading city planners to rely primarily on ad-hoc approaches that do not rigorously account for impacts on ridership (Khany 2022).

3.1.1 Contributions

Where should bike lanes be built to promote cycling ridership while mitigating congestion effects? We address this question by developing an estimation procedure for quantifying the effect of bike lanes on cycling ridership and congestion, which we extend to a prescriptive model that generates recommendations for bike lane expansion. Our modeling approach is informed by discussions with city planners,[2] and is illustrated through an extensive empirical study using data from the City of Chicago[3].

Our main methodological contribution is a new estimation technique for uncovering parameters of a traffic equilibrium model from observational data. Specifically, we aim to estimate a congestion (i.e., driving time) function from features of the road network and observed vehicle flows. While there is a rich literature on traffic equilibrium models (following the work of Wardrop (1952), Beckmann et al. (1956), and Dafermos (1980)), few previous studies attempt to estimate the congestion functions that lie at the foundation of

[2]Department of Transportation, City of Chicago, and Transportation Planning Division, City of Vancouver, British Columbia.

[3]The readers are referred to Chapter 4 for more details

such models.

Specifically, our method estimates parameters of a congestion function by assuming the traffic data are noisy observations of Wardrop equilibrium network flows. We show that the estimator is asymptotically optimal in that it provably recovers the ground-truth model as the size of the network grows large (under appropriate conditions). The main computational hurdle is that enforcing the Wardrop equilibrium conditions makes the estimator a challenging non-convex optimization problem. To obtain solutions, we show that an approximation based on Lagrangian relaxation admits a multi-convex structure, which allows us to generate reliable estimates through an iterative solution procedure. We then estimate a multinomial logit mode choice model in which driving and cycling utilities depend on both network structure and predicted equilibrium travel times.

Next, we embed the estimated congestion and mode choice parameters into a prescriptive *path selection* model that identifies road segments for bike lane installation. The model endogenizes the mode choice of all commuters, the route choice of drivers, and driving travel times, making it a mixed-integer nonlinear program. To obtain solutions, we employ a piecewise linear approximation technique that yields a mixed-integer linear program, and we show that the suboptimality due to the linearization is bounded as a function of network and demand parameters.

The remainder of the chapter is organized as follows. Section 3.1.2 reviews related literature. Section 3.2 outlines model fundamentals, including the traffic equilibrium and mode choice models. Section 3.3 develops an estimator for learning parameters of the traffic model from network features and observed vehicle flows. Section 3.4 presents a prescriptive model for bike lane expansion, an approximation scheme, and a bound on the approximation error.

3.1.2 Related Literature

This work contributes to the literature on inverse optimization, transportation network design, and cycling infrastructure planning. Each of these topics has a large extant literature, so we focus here on the most closely related work.

Inverse optimization on networks. Inverse optimization methods aim to learn parameters of an optimization model from observations of its optimal solutions. Many inverse optimization methods take a statistical learning perspective, where the model to be fit to observational data is itself a constrained optimization model (e.g., Bertsimas et al. (2015), Aswani et al. (2018), Esfahani et al. (2018), Chan et al. (2019)). Our estimator takes the form of an inverse optimization problem because we interpret vehicle traffic data to be noisy observations of equilibrium network flows, which can be posed as the solution to a convex optimization problem (Beckmann et al. 1956). Below, we focus on prior work in inverse optimization in the context of traffic equilibria and network problems, and refer the reader to Chan et al. (2021b) for a comprehensive review of the broader inverse optimization literature.

Given a network with a set of origin-destination (OD) demands and a known congestion function on each edge, a *traffic assignment* model describes the flow of traffic through the network (Patriksson 2015). The standard equilibrium concept in traffic models is Wardrop equilibrium, which states that no user can unilaterally decrease their travel time by taking an alternative route (Wardrop 1952). While there is an extensive literature on traffic modeling dating back to Wardrop (1952), the inverse problem of estimating traffic models from observational data has only recently been considered (Chow et al. 2014, Bertsimas et al. 2015, Thai et al. 2015, Zhang and Paschalidis 2017, Zhang et al. 2018, Allen et al. 2021).

Among this work, this chapter is closest to Bertsimas et al. (2015) and Thai et al. (2015), who also propose methods for estimating congestion functions on a network from observed traffic flows. Our method differs from Bertsimas et al. (2015) primarily in the specification of the loss function – Bertsimas et al. (2015) assume that the observed flows

approximately satisfy Wardrop equilibrium, and minimize a loss function that measures the degree to which equilibrium conditions are violated. In contrast, we enforce the Wardrop equilibrium conditions exactly, and assume that traffic flows are noisy observations of true (unobserved) equilibrium flows, resulting in a loss function that measures the error between the observed and true vehicle flows. The main consequence of these differing loss functions is that Bertsimas et al. (2015) obtain an estimation problem that is convex (due to relaxing the equilibrium conditions), whereas our estimator is non-convex. However, we show in this chapter that our estimator is statistically consistent, which (as shown by Aswani et al. (2018)) is not the case for the convex approach in Bertsimas et al. (2015). In other words, in the trade-off between computational tractability and optimality, the method of Bertsimas et al. (2015) prioritizes tractability of the estimator, whereas we pursue asymptotic optimality of the estimates.

Our estimation procedure also has similarities to Thai et al. (2015) – we highlight two key differences. First, we estimate a congestion function that can depend on an arbitrary number of road features (e.g., length, width, number of lanes, location), whereas Thai et al. (2015) estimate a single road capacity parameter. Second, Thai et al. (2015) use a *link-flow* formulation of traffic assignment, whereas we use a *path-flow* formulation, which restricts the number of possible paths users can take between each OD pair (Patriksson 2015). Our approach yields far fewer decision variables in the resulting inverse optimization problem, which allows it scale more gracefully to realistically sized road networks.

This chapter also contributes to a broader literature on inverse network flow problems (Burton and Toint 1992, Xu and Zhang 1995, Zhang et al. 1995, Zhang and Ma 1996, Zhang and Cai 1998, Ahuja and Orlin 2001, Zhao et al. 2015, Chan et al. 2021a). To the best of our knowledge, ours is the first inverse optimization method specifically for network flow problems that provably recovers ground-truth, data-generating parameters.

Transportation network planning. The design of transportation networks is a fundamental area of research in the transportation literature. Seminal papers include Dantzig

et al. (1979), Abdulaal and LeBlanc (1979) and Magnanti and Wong (1984); we refer the reader to Farahani et al. (2013) for a comprehensive review. Our work is closest to the subset of literature that considers multiple travel modes, which often introduces a discrete choice component to the traffic assignment problem (Lee and Vuchic 2005, Fan and Machemehl 2006, Cipriani et al. 2006, Beltran et al. 2009, Szeto et al. 2010, Gallo et al. 2011, Miandoabchi et al. 2012, Bertsimas et al. 2020).

Our work differs from the majority of the multi-modal network design literature in two ways. First, demand endogeneity plays a significant role in our work – we assume the addition of bike lanes increases cycling utility, and can either increase or decrease driving utility depending on the net impact on congestion. In contrast, few previous papers consider a setting where a central planner modifies network topology and the modal-split depends on the constructed network (Lee and Vuchic (2005) and Cipriani et al. (2006) are notable exceptions). A related paper in this regard is by Wei et al. (2021), who consider the impact of adjusting transit schedules on both traffic congestion and demand. The main difference is we focus on altering network structure (i.e., bike lane locations), whereas Wei et al. (2021) address scheduling of transit times on a given network.

Second, because network design problems are usually non-convex, they are often solved using fast heuristics that return local optimal solutions, such as genetic algorithms (Farahani et al. 2013). In contrast, our solution technique is based on a mixed-integer linear programming (MILP) approximation of the network design problem. The advantage of our approach is that it allows us to bound the approximation error and obtain provably near-optimal solutions.

Cycling infrastructure planning. Cycling infrastructure planning has gained attention due to the worldwide adoption of bike-share systems and increased emphasis on sustainability by municipal governments. In particular, operational data made available by bike-share systems enables a better understanding of biking demand and travel patterns (Singhvi et al. 2015, Scott et al. 2021) and improved design of bike-share station networks

(Kabra et al. 2020, He et al. 2021). However, in the field of bike lane planning, where the computational challenges are prominent and different heuristic methods have been explored (Mauttone et al. 2017, Liu et al. 2019), few papers integrate real world data sets into the planning models. Bao et al. (2017) develop a bike lane planning model built on bike trajectory data, in which an exponential scoring function is used to guide bike lane planning. Recently, Liu et al. (2021) propose an integer optimization model to combine cycling demand data with bike lane network design in a way that balances demand coverage and lane continuity. However, both papers ignore the endogeneity of cycling demand and congestion effects. Our work presents a systematic method for bike lane planning that brings together travel-time estimation and network optimization to rigorously account for the interaction between vehicle traffic and bike lane construction.

3.2 Model

This section outlines a traffic model where congestion and travel mode choice both depend on the presence of bike lanes in the road network. Specifically, Section 3.2.1 describes the road network, Section 3.2.2 presents a model of traffic congestion, and Section 3.2.3 describes how commuters choose among alternative travel modes (e.g., driving vs. cycling) based on driving travel times and the bike lane network.

3.2.1 Road and Bike Lane Network

Let $\mathcal{I}$ be a set of *nodes*, each of which can be an *origin*, *destination*, or both. Let $\mathcal{W}$ index the set of all *origin-destination (OD) pairs*. Each OD pair is a pair of nodes (i, j) such that $i \neq j$ that commuters wish to travel between; following the literature, for conciseness we index all pairs using $w \in \mathcal{W}$. Let $\mathcal{S}$ be the set of *segments*, where each segment $s \in \mathcal{S}$ connects two nodes. Let $\mathcal{M} = \{C, D, O\}$ be the set of travel modes, where C, D, and O represent *cycling*, *driving*, or an *outside option* capturing all other modes (e.g., walking

and public transit). We call a concatenation of adjacent segments a *path*; let $\mathcal{P}_w^D$ be the set of possible driving paths that connect OD pair w. For tractability, we assume there is a single cycling path p_w^C that connects OD pair w. Let $\mathcal{S}_p^D$ be the set of all segments that lie on driving path $p \in \mathcal{P}_w$, $w \in \mathcal{W}$ and $\mathcal{S}_w^C$ be the segments on cycling path $p = p_w^C$, $w \in \mathcal{W}$. For ease of reference, we use $\mathcal{P}^D = \bigcup_{w\in\mathcal{W}} \{\mathcal{P}_w^D\}$ and $\mathcal{P}^C = \bigcup_{w\in\mathcal{W}} \{p_w^C\}$ to denote all possible driving and cycling paths, and $\mathcal{S}^D = \bigcup_{p\in\mathcal{P}^D} \{\mathcal{S}_p^D\}$ and $\mathcal{S}^C = \bigcup_{w\in\mathcal{W}} \{\mathcal{S}_w^C\}$ to denote all road segments used for driving and cycling, respectively.

Let $\mathbf{x} \in \{0,1\}^{|\mathcal{S}|}$ be a binary vector denoting the existence of bike lanes on each segment of the network, where $x_s = 1$ if and only if segment s contains a bike lane. We let $\tilde{\mathbf{x}}$ denote the status quo bike lane. We assume each commuter is infinitesimal throughout. Let $\mathbf{d}^D \in \mathbb{R}^{|\mathcal{W}|}$, $\mathbf{d}^C \in \mathbb{R}^{|\mathcal{W}|}$ and $\mathbf{d}^O \in \mathbb{R}^{|\mathcal{W}|}$ be the demand for each mode over all OD pairs, and let $\mathbf{d} = (\mathbf{d}^C, \mathbf{d}^D, \mathbf{d}^O)$. Next, we consider how a fixed driving demand vector $\mathbf{d}^D$ induces vehicle traffic through the network.

3.2.2 Route Selection and Traffic Congestion

Commuters who travel by driving for OD pair w choose among the paths in $\mathcal{P}_w^D \subseteq \mathcal{P}^D$. Accordingly, let ϕ_p be the mass of commuters who choose path $p \in \mathcal{P}^D$, and let v_s be mass of commuters on segment $s \in \mathcal{S}^D$. We refer to ϕ_p and v_s as path and segment *flows*, respectively. For conciseness, let $\boldsymbol{\phi} \in \mathbb{R}^{|\mathcal{P}^D|}$ and $\mathbf{v} \in \mathbb{R}^{|\mathcal{S}|}$ be the vectors of ϕ_p and v_s, respectively. Then for fixed $\mathbf{d}^D$, we say $(\boldsymbol{\phi}, \mathbf{v})$ are *feasible* flows if they satisfy

$$\sum_{p\in\mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{3.1a}$$

$$v_s = \sum_{\{p\in\mathcal{P}^D | s\in\mathcal{S}_p^D\}} \phi_p, \quad s \in \mathcal{S}, \tag{3.1b}$$

$$(\boldsymbol{\phi}, \mathbf{v}) \geq \mathbf{0}. \tag{3.1c}$$

For convenience, we define

$$\Omega\left(\mathbf{d}^D\right) = \{(\boldsymbol{\phi}, \mathbf{v}) \,|\, (\boldsymbol{\phi}, \mathbf{v}) \text{ satisfies (3.1)}\} \tag{3.2}$$

to be the set of all feasible flows given driving demand $\mathbf{d}^D$. The equations in (3.1) assume a restricted set of candidate driving paths $\mathcal{P}_w^D$ for each OD pair w, meaning drivers cannot travel along an arbitrary sequence of segments. This *path-flow* formulation is widely used in the literature to improve tractability (e.g., Fisk (1980), Dafermos (1980, 1982)).[4]

Next, we describe how driving travel times depend on vehicle flows, the road network, and the presence of bike lanes. With a slight abuse of notation, let $\mathbf{q}_s(\mathbf{x})$ be a vector of features on the road segment s (e.g., the road's length, width, number of vehicle lanes, or location), where the features may depend on the presence of bike lanes. We define for each segment a *congestion function* α_s, which depends linearly on segment features:

$$\alpha_s\left(\mathbf{x}, \boldsymbol{\theta}\right) = \boldsymbol{\theta}^\top \mathbf{q}_s\left(\mathbf{x}\right). \tag{3.3}$$

We refer to $\boldsymbol{\theta}$ as the *congestion parameter* that influences travel times. The travel time on segment s is then given by

$$t_s(\mathbf{x}, \boldsymbol{\theta}, v_s) = \alpha_s\left(\mathbf{x}, \boldsymbol{\theta}\right) \cdot v_s + T_s, \tag{3.4}$$

where T_s is the *free-flow* travel time. We impose the mild assumption that $\alpha_s\left(\mathbf{x}, \boldsymbol{\theta}\right) > 0$, which implies driving times strictly increase in traffic flow v_s.[5] The total travel time along each path $p \in \mathcal{P}^D$ is then given by

$$t_p\left(\mathbf{x}, \boldsymbol{\theta}, \mathbf{v}\right) = \sum_{s \in \mathcal{S}_p^D} t_s\left(\mathbf{x}, \boldsymbol{\theta}, v_s\right). \tag{3.5}$$

Next, we consider how drivers on OD pair w select among the paths in $\mathcal{P}_w^D$. Following Wardrop's first principle, we assume that drivers select paths such that no driver on any OD

[4] An alternative approach is the *link-flow* formulation, which requires variables for each segment and OD pair in the network – see Beckmann et al. (1956), Nguyen (1974), Patriksson (2015). However, the link-flow representation leads to a less tractable estimation problem in large networks as a result of using $|\mathcal{W}| \cdot |\mathcal{S}|$ variables to represent traffic flows, whereas the path-flow formulation leads to $|\mathcal{P}| + |\mathcal{S}|$ variables.

[5] The most well-known congestion function takes the form of $t_s\left(\mathbf{x}, v_s\right) = T_s \cdot (1 + \alpha(v_s/m_s)^4)$, where typically $\alpha = 0.15$ (United States Bureau of Public Roads 1964). Our use of a linear congestion function is primarily to improve the tractability of the estimator. Linear congestion functions have appeared elsewhere in the literature (Roughgarden and Tardos 2002, Lin et al. 2011, Swamy 2012).

pair w can reduce their travel time by unilaterally selecting an alternative path (Wardrop 1952). To capture this equilibrium condition, let $\mathbf{v}^*(\mathbf{x}, \boldsymbol{\theta})$ and $\boldsymbol{\phi}^*(\mathbf{x}, \boldsymbol{\theta})$ denote equilibrium segment and path flows, respectively, and let $t_w(\mathbf{x}, \boldsymbol{\theta})$ denote the equilibrium travel time on OD pair w. Then Wardrop's first principle is equivalent to the following set of conditions:

$$t_w\left(\mathbf{x}, \boldsymbol{\theta}\right) - t_p\left(\mathbf{x}, \boldsymbol{\theta}, \mathbf{v}^*(\boldsymbol{\theta})\right) \left\{ \begin{array}{ll} = 0, & \text{if } \phi_p^*(\mathbf{x}, \boldsymbol{\theta}) > 0 \\ \leq 0, & \text{if } \phi_p^*(\mathbf{x}, \boldsymbol{\theta}) = 0 \end{array} \right\}, \quad p \in \mathcal{P}_w,\ w \in \mathcal{W}. \tag{3.6}$$

For ease of exposition in the remainder of the chapter, we define

$$\Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta}) = \{(\boldsymbol{\phi}, \mathbf{v}) | (\boldsymbol{\phi}, \mathbf{v}) \in \Omega\left(\mathbf{d}^D\right) \text{ and } (\boldsymbol{\phi}, \mathbf{v}) \text{ satisfies (3.6)}\} \tag{3.7}$$

to be the set of feasible flows that satisfy Wardrop's first principle. That is, $(\boldsymbol{\phi}, \mathbf{v})$ are *equilibrium flows* if and only if $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta})$. Note that $\Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta})$ depends on the observable road segment features $\mathbf{q}_s(\mathbf{x})$ and the parameters $\boldsymbol{\theta}$ via the congestion functions $\alpha_s(\mathbf{x}, \boldsymbol{\theta})$.

3.2.3 Travel Mode Choice

We now describe how commuters select among alternate travel modes. Following the literature (e.g., Domencich and McFadden (1975), Abdulaal and LeBlanc (1979)), we assume the mode-specific demands d_w^m are determined by a multinomial logit (MNL) choice model. Let $u_w^m(\mathbf{x}, \boldsymbol{\theta})$ be the disutility of traveling by mode m on OD pair w. Note that the equilibrium conditions (3.6) imply for each $w \in \mathcal{W}$, the driving time is equal for all $p \in \mathcal{P}_w^D$. Accordingly, we assume all drivers on OD pair w share the same disutility regardless of their selected path, which depends primarily on the travel time:[6]

$$u_w^D(\mathbf{x}, \boldsymbol{\theta}) = \beta_0^D + \beta_1^D \cdot t_w(\mathbf{x}, \boldsymbol{\theta}). \tag{3.8}$$

[6]Travel time is typically the key determinant of utility in mode choice models (Ben-Akiva and Bierlaire 1999). Our method can be extended to accommodate other route features (e.g., number of turns) provided sufficient data is available.

We define the disutility of cycling next. For tractability, we assume that for each OD pair $w \in \mathcal{W}$, all cyclists travel along the same path p_w^C. We assume cycling disutility depends on both cycling distance and bike lane *coverage* (i.e., the proportion of the path that is covered by a bike lane), which have both been observed to be significant factors for cycling adoption (Hood et al. 2011, Broach et al. 2012).[7] To that end, let ℓ_s be the length of segment s, and let $\mathcal{S}_w^C$ be the set of segments that lie on cycling path p_w^C. The cycling distance for OD pair w is then given by

$$L_w = \sum_{s \in \mathcal{S}_w^C} \ell_s, \tag{3.9}$$

and the bike lane coverage for OD pair w is given by

$$\rho_w(\mathbf{x}) = \frac{1}{L_w} \sum_{s \in \mathcal{S}_w^C} \ell_s \cdot x_s. \tag{3.10}$$

The disutility of cycling for each w is then

$$u_w^C(\mathbf{x}) = \beta_0^C + \beta_1^C \cdot \rho_w(\mathbf{x}) + \beta_2^C \cdot L_w. \tag{3.11}$$

Note that cycling disutility does not depend on the congestion parameter $\boldsymbol{\theta}$, which influence driving times only. Without loss of generality, we normalize the disutility of the outside option to $u_w^O = 0$ for all $w \in \mathcal{W}$.

Let $\tilde{d}_w$ be the total demand across all travel modes for OD pair w; following the literature, we assume $\tilde{d}_w$ is exogenous and does not depend on $\mathbf{x}$ or $\boldsymbol{\theta}$. The demand for mode m on OD pair w is then given by

$$d_w^m(\mathbf{x}, \boldsymbol{\theta}) = \tilde{d}_w \cdot \frac{e^{-u_w^m(\mathbf{x}, \boldsymbol{\theta})}}{e^{-u_w^C(\mathbf{x})} + e^{-u_w^D(\mathbf{x}, \boldsymbol{\theta})} + e^{-u_w^O}}, \tag{3.12}$$

Note that the MNL model (3.12) implies $\tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m(\mathbf{x}, \boldsymbol{\theta})$. Next, having defined how the equilibrium traffic assignment depends on $\mathbf{x}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, we now consider how to estimate the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ from observational data.

[7] Cycling traffic has not yet been found to be a significant factor in cycling utility (Hood et al. 2011, Broach et al. 2012), so we assume cycling is free from congestion effects.

3.3 Estimation

Our estimation method consists of two steps. In the first step, we estimate the congestion parameter $\boldsymbol{\theta}$ from observed road network features and associated traffic flows. In the second step, we use the estimate of $\boldsymbol{\theta}$ and the status quo bike lane network $\tilde{\mathbf{x}}$ as inputs for estimating the mode choice parameters $\boldsymbol{\beta}$. Our primary methodological contribution is in the first step; given $\boldsymbol{\theta}$, the second step follows from well-known methods.

In Section 3.3.1, we present an estimator for $\boldsymbol{\theta}$ and provide conditions under which it asymptotically recovers the ground-truth model. In Section 3.3.2, we first show the estimator can be expressed as a non-convex optimization problem, and then develop a corresponding solution technique. In Section 3.3.3, we describe how the estimated $\boldsymbol{\theta}$ can be used as an input to estimating the mode choice parameters $\boldsymbol{\beta}$.

3.3.1 Estimator for Congestion Parameter

The estimator for $\boldsymbol{\theta}$ takes input data of the form

$$(\mathbf{q}_s, z_s), \quad s = 1, \ldots, n, \tag{3.13}$$

where $\mathbf{q}_s$ are the features of road segment s, z_s is the associated flow on segment s, and n is the number of segments in the network (equivalently, $|\mathcal{S}|$). Additionally, the estimator takes as input the observed driving demand $\mathbf{d}^D$ and the path sets $\mathcal{P}_w^D$, $w \in \mathcal{W}$. Because our goal is to fit a structural model of traffic congestion to observational data, we assume there exists a true, unknown congestion parameter $\boldsymbol{\theta}_0$ such that the observed flows z_s are noisy measurements of equilibrium segment flows $\mathbf{v}^*(\boldsymbol{\theta}_0)$. Additionally, because our focus here is on estimating the congestion parameter $\boldsymbol{\theta}$ for a given road and bike lane network, we assume $\mathbf{x}$ and $\mathbf{d}$ are fixed and suppress dependence on them throughout this section.

We first define the following loss function:

$$L_n(\boldsymbol{\theta}) = \underset{\boldsymbol{\phi},\mathbf{v}}{\text{minimize}} \;\; \|\mathbf{v}-\mathbf{z}\|_2^2 \tag{3.14a}$$

$$\text{subject to} \;\; (\boldsymbol{\phi},\mathbf{v}) \in \Psi(\boldsymbol{\theta}), \tag{3.14b}$$

where $\|\cdot\|_2$ is the ℓ_2-norm. Intuitively, the loss $L_n(\boldsymbol{\theta})$ measures the error between the observed flows $\mathbf{z}$ and the flows $\mathbf{v}$, where the constraint $(\boldsymbol{\phi},\mathbf{v}) \in \Psi(\boldsymbol{\theta})$ forces $\mathbf{v}$ to satisfy Wardrop equilibrium under $\boldsymbol{\theta}$. In other words, the loss $L_n(\boldsymbol{\theta})$ is a measure of distance between the observed traffic flows and the equilibrium flows implied by $\boldsymbol{\theta}$. The estimate is then obtained by minimizing of the loss function over a parameter set Θ:

$$\hat{\boldsymbol{\theta}}_n \in \underset{\boldsymbol{\theta}\in\Theta}{\text{argmin}} \;\; L_n(\boldsymbol{\theta}). \tag{3.15}$$

An important property of any estimator is statistical *consistency*, which refers to the asymptotic recovery of unknown model parameters (Bickel and Doksum 2015). If the observed flows $\mathbf{z}$ are generated by a true, unknown parameter $\boldsymbol{\theta}_0$, then estimator consistency implies that the sequence of estimates $\hat{\boldsymbol{\theta}}_n$ converges to the true parameter $\boldsymbol{\theta}_0$. Our main result in this section is to provide conditions under which the estimator (3.15) is indeed consistent:

Theorem 3 (**Estimator consistency**). *Suppose the following conditions hold:* (i) $z_s = v_s^*(\boldsymbol{\theta}_0) + \varepsilon_s$ *for* $s = 1,\ldots,n$, *where the errors* ε_s *are drawn i.i.d. from a distribution with* $\mathbb{E}[\varepsilon] = 0$ *and unknown variance* $\sigma^2 > 0$, (ii) *the parameter set* Θ *is finite and compact, and* (iii) $\lim_{n\to\infty} \sum_{s=1}^{n} (\boldsymbol{v}^*(\boldsymbol{\theta}) - \boldsymbol{v}^*(\boldsymbol{\theta}_0))^2 = \infty$ *for all* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. *Then* $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ *as* $n \longrightarrow \infty$.

Theorem 3 provides conditions under which the sequence of estimates $\hat{\boldsymbol{\theta}}_n$ converges in probability to the true parameter $\boldsymbol{\theta}_0$. Conditions (i) and (ii) are standard assumption imposed on the data and parameter set in non-linear least squares estimation. Condition (iii) is an *identifiability* condition, which ensures that the ground truth model can be learned from data. Intuitively, the identifiability condition states that as the size of the network grows large, any two distinct values of $\boldsymbol{\theta}$ should lead to two distinct equilbrium flows $\mathbf{v}^*(\boldsymbol{\theta})$. In

the absence of model identifiability, two different parameters $\boldsymbol{\theta}$ may induce the exact same equilibrium flows over the network, in which case consistency cannot be attained by any estimator, because the observed data is insufficient for pinpointing the true parameter $\boldsymbol{\theta}_0$.

The estimation problem (3.15) can be viewed as an analog to non-linear least squares regression (Jennrich (1969), Wu (1981)). The main distinction between the estimator (3.15) and traditional non-linear least squares regression is that the response variables $\mathbf{z}_s$ are not observations of a non-linear function in closed form, but are instead noisy observations of equilibrium segment flows $\mathbf{v}^*(\boldsymbol{\theta})$ which are implicitly defined. Further, as we discuss in Section 3.3.2 below, the equilibrium flows $(\boldsymbol{\phi}^*(\boldsymbol{\theta}), \mathbf{v}^*(\boldsymbol{\theta}))$ can be attained at the solution to a convex optimization problem, which allows (3.15) to be expressed as an *inverse optimization* problem, which refers to the the estimation of an optimization model's parameters from observations of its optimal solutions.

Statistical consistency in inverse optimization was first addressed by Aswani et al. (2018). The main difference with our approach is that Aswani et al. (2018) assume the feature and response data are independent draws of a common joint distribution across all n observations. However, independence fails to hold in our setting because each observation corresponds to a road segment in a network, which have correlated traffic flows. As a result, the sufficient conditions for consistency presented in Theorem 3 differ from those given in Aswani et al. (2018).

3.3.2 Solution Procedure

Next, we present a procedure for solving the estimation problem (3.15). We first express (3.15) as a mathematical program by developing an explicit representation of the equilibrium condition $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$. To that end, consider the following "Wardrop least-squares"

formulation:

$$\underset{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b}}{\text{minimize}} \quad \|\mathbf{v}-\mathbf{z}\|_2^2 \tag{3.16a}$$

$$\text{subject to} \quad \nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^{\top}\boldsymbol{\phi} - \left(\mathbf{d}^D\right)^{\top}\mathbf{b} = 0, \tag{3.16b}$$

$$\sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w^D\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right), \quad p\in\mathcal{P}, \tag{3.16c}$$

(WLS)

$$\sum_{p\in\mathcal{P}_w^D} \phi_p = d_w^D, \quad w\in\mathcal{W}, \tag{3.16d}$$

$$v_s = \sum_{\{p\in\mathcal{P}^D|s\in\mathcal{S}_p^D\}} \phi_p, \quad s\in\mathcal{S}, \tag{3.16e}$$

$$\boldsymbol{\theta}\in\Theta, \tag{3.16f}$$

where

$$g(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{s\in\mathcal{S}}\left(\frac{1}{2}\cdot\alpha_s(\mathbf{x})\cdot\left[\sum_{\{p\in\mathcal{P}^D|s\in\mathcal{S}_p^D\}}\phi_p\right]^2 + T_s\cdot\left[\sum_{\{p\in\mathcal{P}^D|s\in\mathcal{S}_p^D\}}\phi_p\right]\right). \tag{3.17}$$

The following result establishes an equivalence between the estimator given in (3.15) and the optimization problem WLS:

Proposition 5. *Let $\boldsymbol{\theta}^*$ be attained at an optimal solution to WLS. Then $\boldsymbol{\theta}^*\in\operatorname{argmin}_{\theta\in\Theta}L_n(\boldsymbol{\theta})$.*

At a high level, Proposition 5 relies on a well-known result which shows that equilibrium network flows $(\boldsymbol{\phi},\mathbf{v})\in\Psi(\mathbf{d}^D,\boldsymbol{\theta})$ correspond to the solution to a convex optimization problem (Beckmann et al. 1956). Expressing the optimality conditions of that convex program as a variational inequality and invoking linear programming duality[8] then yields the constraints (3.16b) and (3.16c), which are together equivalent to the Wardrop equilibrium conditions (3.6). This implies that a solution $(\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b})$ that satisfies (3.16b)-(3.16e) must also satisfy $(\boldsymbol{\phi},\mathbf{v})\in\Psi(\mathbf{d}^D,\boldsymbol{\theta})$, which allows us to express the estimator WLS explicitly as (3.16).

[8]We refer the reader to Aghassi et al. (2006) for details on the on the reformulation of variational inequalities using linear programming duality.

One challenge in obtaining solutions to (3.16) is that constraint (3.16b) is a non-linear equality constraint, which are known to pose numerical difficulties with respect to finding feasible solutions (Hearn and Ramana 1998). To overcome this, we dualize constraint (3.16b) by introducing a penalty parameter $\lambda > 0$ and the auxiliary variable ϵ, which yields the following approximation of WLS:

$$\underset{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b},\epsilon}{\text{minimize}} \quad \|\mathbf{v}-\mathbf{z}\|_2^2 + \lambda \cdot \epsilon \tag{3.18a}$$

(WLS-A)
$$\text{subject to } \nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} \leq \epsilon, \tag{3.18b}$$

$$(3.16c) - (3.16f).$$

The loss (3.18a) can be interpreted as a linear combination of the "flow error" $\|\mathbf{v}-\mathbf{z}\|_2^2$ and the "Wardrop error" ϵ.[9]

In addition to circumventing the numerical issues of the exact estimator WLS, the approximate estimator WLS-A has the advantage of being a multi-convex optimization problem, meaning the decision variables can be partitioned into separate blocks such that the optimization problem is convex in one block of variables when the others are held fixed (Xu and Yin 2013). Specifically, using (3.17), it can be shown that fixing $(\boldsymbol{\theta}, \mathbf{b})$ in WLS-A yields a convex subproblem in $(\boldsymbol{\phi}, \mathbf{v})$, and vice versa. As a result, locally optimal solutions can be obtained using an iterative block coordinate descent method, which is the standard approach to solving multi-convex optimization problems (see e.g., Bertsekas (1997), Xu and Yin (2013), and Hong et al. (2017)). Details on our implementation of block coordinate descent on the approximate estimator WLS-A are provided in Appendix B.1.2.

3.3.3 Travel Mode Choice Estimation

It is straightforward to estimate the parameter $\boldsymbol{\beta}$ that defines the mode choice model outlined in Section 3.2.3. Specifically, let $\tilde{\boldsymbol{\theta}}$ be the estimate from solving WLS-A. Then the

[9]Note that multiplying both sides of (3.16c) by $\boldsymbol{\phi}$ yields $\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} \geq 0$, which implies the Wardrop error ϵ is always non-negative.

corresponding equilibrium segment flows $\mathbf{v}^*(\tilde{\boldsymbol{\theta}})$ are given by solving a convex optimization problem (see formulation (B.20) in the proof of Proposition 5). Let $\tilde{\mathbf{x}}$ encode the status quo bike lane network under which the observed mode shares d_w^m are generated. It follows from (3.6) that the equilibrium travel time for each OD pair $w \in \mathcal{W}$ is given by

$$t_w\left(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}\right) = \min_{p \in \mathcal{P}_w^D} \left\{ t_p\left(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}, \mathbf{v}^*(\tilde{\boldsymbol{\theta}})\right) \right\}. \tag{3.19}$$

Therefore, the status quo bike lane network $\tilde{\mathbf{x}}$ and the estimate $\tilde{\boldsymbol{\theta}}$ together pin down the travel times $t_w\left(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}\right)$, and thus the driving disutilities $u_w^D(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$. Similarly, the cycling disutility under the existing bike lane network is given by $u_w^C(\tilde{\mathbf{x}})$, where the coverage $\rho_w(\tilde{\mathbf{x}})$ and cycling distance L_w are determined by (3.9) and (3.10).[10]

Given the functions $u_w^D(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$ and $u_w^C(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$, as well as observed mode shares d_w^C, d_w^D and d_w^0 for $w \in \mathcal{W}$, the parameters $\boldsymbol{\beta}$ in the mode choice model (3.12) are then readily estimable by standard MNL estimation methods (McFadden et al. 1973).

3.4 Bike Lane Path Selection with Ridership and Congestion Effects

Equipped with estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, we now present a prescriptive model for planning a bike lane network $\mathbf{x}$ (i.e., from the perspective of a city planner). We assume $\boldsymbol{\theta}$ are the estimates returned by WLS-A, and generally suppress dependence on it for brevity.

3.4.1 Model

Continuity. Continuity of bike lanes is often a priority of city planners (Federal Highway Administration (FHA) 2006, Bao et al. 2017). To that end, we consider bike lane design

[10] In practice, estimating the mode choice model requires us to specify, for each $w \in \mathcal{W}$, the set of possible driving paths $\mathcal{P}_w^D$ and the cycling path p_w^C. This can be done by querying route planning software such as Google Maps, which we do in Chapter 4.

decisions at the path level: let $\mathbf{y} \in \{0,1\}^{|\mathcal{W}|}$ represent binary decision variables where $y_w = 1$ if and only if a bike lane is constructed on cycling path p_w^C. For notational convenience, let S_w^C be the set of segments that lie on path p_w^C. Because $x_s = 1$ if a bike lane is located on segment s, we enforce coherence between $\mathbf{x}$ and $\mathbf{y}$ using the constraints

$$x_s \geq y_w, \quad s \in S_w^C,\ w \in \mathcal{W}, \tag{3.20a}$$

$$x_s \leq \sum_{\{w \in \mathcal{W} | s \in \mathcal{S}_w^C\}} y_w, \quad s \in \mathcal{S}. \tag{3.20b}$$

Bike lane budget. Let $L = \sum_{s \in \mathcal{S}} \ell_s$ be the total length of all road segments eligible for bike lane construction. We assume the city planner can expand the bike lane network by at most B miles where $B \in [0, L]$. This bike lane budget may be due to financial, scheduling, or other logistical constraints – we abstract away from those specifics and focus on total mileage, which yields the constraint

$$\sum_{s \in \mathcal{S}} \ell_s \cdot x_s \leq B. \tag{3.21}$$

Mode choice and congestion. For ease of exposition, let $\mathbf{d} = (\mathbf{d}^C, \mathbf{d}^D, \mathbf{d}^O)$ represent all mode-specific demands, and define

$$\Gamma(\mathbf{x}) = \{(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \mid \mathbf{d} \text{ satisfies (3.12) and } (\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta})\}. \tag{3.22}$$

We call the tuple $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v})$ a *traffic assignment*. The set $\Gamma(\mathbf{x})$ represents all equilibrium traffic assignments under the bike lane network $\mathbf{x}$ (and congestion parameter $\boldsymbol{\theta}$). We can then capture the effect of a candidate bike lane plan $\mathbf{x}$ on the ensuing commuter mode choice and congestion by enforcing the equilibrium condition

$$(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x}). \tag{3.23}$$

Cycling ridership. Lastly, we assume the city planner's goal is to maximize total cycling ridership (alternative objectives are discussed in Section 4.7). This is consistent with the strategic goals of several municipal transportation planning agencies – for example, Chicago's

DOT states that "ridership is the key criterion for evaluating the success of bicycle infrastructure" (Chicago DOT 2020), and New York City aims to have "1 out of every 10 trips in NYC be taken by bicycle by 2050" (DOT 2019). Noting that total cycling ridership is given by $\sum_{w\in\mathcal{W}} d_w^C(\mathbf{x})$, the model for *bike lane path selection* (BLP) can now be written as

$$\text{(BLP)} \qquad \begin{aligned} \underset{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v}}{\text{maximize}} \quad & \sum_{w\in\mathcal{W}} d_w^C && \text{(3.24a)}\\ \text{subject to} \quad & x_s \geq y_w, \quad s \in S_w^C,\ w \in \mathcal{W}, && \text{(3.24b)}\\ & x_s \leq \sum_{\{w\in\mathcal{W}|s\in\mathcal{S}_w^C\}} y_w, \quad s \in \mathcal{S}, && \text{(3.24c)}\\ & \sum_{s\in\mathcal{S}} \ell_s \cdot x_s \leq B, && \text{(3.24d)}\\ & (\mathbf{d},\boldsymbol{\phi},\mathbf{v}) \in \Gamma(\mathbf{x}), && \text{(3.24e)}\\ & \mathbf{x} \in \{0,1\}^{|\mathcal{S}|}, && \text{(3.24f)}\\ & \mathbf{y} \in \{0,1\}^{|\mathcal{W}|}. && \text{(3.24g)} \end{aligned}$$

The formulation above is computationally challenging because the cycling ridership objective (3.24a) and the traffic equilibrium condition (3.24e) are both non-convex as a result of depending on multinomial logit choice probabilities (see (3.12)). In the next section, we show that BLP can be approximated as a mixed-integer linear program (MILP), which yields a formulation that is amenable to commercial optimization solvers, and we characterize the approximation error explicitly.

3.4.2 Linear Approximation and Suboptimality Bound

The approximation of BLP consists of two steps: First, we reformulate (3.24) exactly by leveraging an equivalence between the equilibrium condition (3.24e) and a related convex optimization problem. Second, we present a linearization technique to approximate the remaining non-convex terms as piecewise linear functions, which yields a MILP.

Reformulation of equilibrium conditions. The following theorem establishes that the

set of equilibrium traffic assignments $\Gamma(\mathbf{x})$ is equivalent to the set of optimal solutions to a convex optimization problem.

Proposition 6 (**User equilibrium**). *For any bike lane network* $\boldsymbol{x}$, $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\boldsymbol{x})$ *if and only if* $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v})$ *solves the convex optimization problem*

$$\underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v}}{\text{minimize}} \;\; S(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{v}) \tag{3.25a}$$

$$\text{subject to} \;\; \tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m, \;\; w \in \mathcal{W}, \tag{3.25b}$$

$$(\boldsymbol{\phi}, \boldsymbol{v}) \in \Omega\left(\mathbf{d}^D\right), \tag{3.25c}$$

where

$$S\left(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{d}, \boldsymbol{v}\right) = \sum_{w \in \mathcal{W}} \left(\sum_{m \in \{O,C\}} \int_0^{d_w^m} u_w^m(\boldsymbol{x}) \cdot dd + \int_0^{d_w^D} \beta_0^D \cdot dd \right) + \beta_1^D \sum_{s \in \mathcal{S}} \left(\int_0^{v_s} t_s\left(\mathbf{x}, \boldsymbol{\theta}, v\right) \cdot dv \right) \tag{3.26a}$$

$$+ \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} d_w^m \cdot \log\left(d_w^m\right). \tag{3.26b}$$

Further, the equilibrium mode demands $\boldsymbol{d}$ *and traffic flows* $\boldsymbol{v}$ *attained at a solution to* (3.25) *are unique.*

For ease of reference, we suppress the dependence of function $S\left(\cdot\right)$ on $\boldsymbol{\theta}$ throughout this section. Proposition 6 follows closely from Abdulaal and LeBlanc (1979), Dafermos (1982) and Fisk (1980).[11] Following the literature, we refer to formulation (3.25) as the *user equilibrium* problem because it specifies equilibrium mode choices and driving routes. The objective $S\left(\mathbf{x}, \mathbf{d}, \mathbf{v}\right)$ is composed of two parts: a system-wide generalized cost term (3.26a), which is the sum of disutilites of all commuters, and an entropy term (3.26b). Note that $S\left(\mathbf{x}, \mathbf{d}, \mathbf{v}\right)$ is related to the observable road segment features $\mathbf{q}(\mathbf{x})$ and traffic flow v

[11] Our formulation differs in that (1) Abdulaal and LeBlanc (1979) only considers two modes, public transit and driving, without specifying mode-split functions, (2) Dafermos (1982) assumes consumers adopt the travel mode with the lowest cost, and (3) Fisk (1980) focuses on the mode of driving only.

via the driving time function $t_s(\mathbf{x}, \boldsymbol{\theta}, v)$. Further, because $t_s(\mathbf{x}, \boldsymbol{\theta}, v)$ strictly increases in v, $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is strictly convex in (d, v).

Proposition 6 allows us to replace the equilibrium condition (3.24e) in the path selection model with the equivalent condition

$$(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \underset{\mathbf{d},(\boldsymbol{\phi},\mathbf{v})\in\Omega(\mathbf{d}^D)}{\operatorname{argmin}} S(\mathbf{x}, \mathbf{d}, \mathbf{v}), \tag{3.27}$$

and transform problem (BLP) into a *bi-level* optimization problem (Dempe 2002). At a high level, the advantage of this substitution is that it allows us to sidestep use of the non-convex function (3.12) in the path selection model and instead work with the convex problem (3.27), which is more tractable with respect to linearization. Further, as discussed below, the linearization of (3.27) yields a bi-level program with a linear lower level problem, which can be reformulated exactly as a mixed-integer linear program (Fortuny-Amat and McCarl 1981).

The uniqueness result in Proposition 6 is important in that it ensures that the equilibrium mode shares (i.e., cycling ridership) are identified from the bike lane network $\mathbf{x}$ alone. This uniqueness allows us to address the network design problem without requiring additional assumptions on the ensuing equilibrium.

Linearization and MILP reformulation. Next, we describe a linearization technique that approximates BLP as a MILP. Note that the upper-level objective function and constraints (3.24a)-(3.24d) and (3.24f), (3.24g) are all linear, so we focus on approximating the user equilibrium problem (3.25). First, note that by evaluating each term in $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ defined in (3.26), it can be written as

$$\sum_{w\in\mathcal{W}} \left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D\right) + \beta_1^D \sum_{s\in\mathcal{S}} (T_s \cdot v_s) + \beta_1^D \sum_{s\in\mathcal{S}} \underbrace{\frac{1}{2}\alpha_s(\mathbf{x}) \cdot v_s^2}_{\xi_s(v_s)} + \sum_{w\in\mathcal{W}} \sum_{m\in\mathcal{M}} \underbrace{d_w^m \log(d_w^m)}_{\psi(d_w^m)}, \tag{3.28}$$

where $\xi_s(v_s)$ and $\psi(d_w^m)$ are both (convex) non-linear terms to be approximated via piecewise linear functions. To linearize $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$, we approximate $\xi_s(v_s)$ with a piecewise linear

function, where

$$\xi_s^r(v_s) = \alpha_s(\mathbf{x}) \cdot v_s^r \cdot v_s - \frac{1}{2} \cdot \alpha_s(\mathbf{x}) \cdot (v_s^r)^2 \tag{3.29}$$

is the r^{th} line segment and v_s^r is the r^{th} sample point. We choose each sample point v_s^r such that the slopes $\alpha_s(\mathbf{x}) \cdot v_s^r$ are equidistant in the interval $\alpha_s(\mathbf{x}) \cdot [\underline{v}, \bar{v}]$, where $\bar{v} = \max_{s \in \mathcal{S}} \{v_s\}$ and $\underline{v} = \min_{s \in \mathcal{S}} \{v_s\}$. Note that linearity of $\alpha_s(\mathbf{x})$ implies $\xi_s^r(\mathbf{x}, v_s)$ is also linear in $\mathbf{x}$, is important for approximating BLP as a MILP. Similarly, we approximate $\psi(d_w^m)$ with a piecewise linear function, where

$$\psi_w^r(d_w^m) = [1 + \log(d_w^r)] \cdot d_w^m - d_w^r \tag{3.30}$$

is the r^{th} line segment and d^r is the r^{th} sample point. We choose each sample point d_w^r such that the slopes are equidistant in the interval $\left[1 + \log(\underline{d}_w), 1 + \log\left(\bar{d}_w\right)\right]$, where $\bar{d}_w = \max_{m \in \mathcal{M}} \{d_w^m\}$ and $\underline{d}_w = \min_{m \in \mathcal{M}} \{d_w^m\}$. Next, let $\mathcal{R}_\xi$ and $\mathcal{R}_\psi$ index the linear segments used to approximate $\xi_s(v_s)$ and $\psi(d_m^w)$, respectively. Then using (3.29) and (3.30), the linear approximation of the user equilibrium problem (3.25) can be written as

$$\underset{\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}, \boldsymbol{\psi}, \boldsymbol{\xi}}{\text{minimize}} \quad S_L(\mathbf{x}, \mathbf{d}, \mathbf{v}, \boldsymbol{\psi}, \boldsymbol{\xi}) \tag{3.31a}$$

$$\text{subject to} \quad \tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m, \quad w \in \mathcal{W}, \tag{3.31b}$$

$$(\boldsymbol{\phi}, \mathbf{v}) \in \Omega(\mathbf{d}^D), \tag{3.31c}$$

$$\xi_s \geq \xi_s^r(\mathbf{x}, v_s), \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{3.31d}$$

$$\psi_w^m \geq \psi_w^r(d_w^m), \quad r \in \mathcal{R}_\psi, w \in \mathcal{W}, m \in \mathcal{M}, \tag{3.31e}$$

where

$$S_L(\mathbf{x}, \mathbf{d}, \mathbf{v}, \boldsymbol{\psi}, \boldsymbol{\xi}) = \sum_{w \in \mathcal{W}} \left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D\right) + \beta_1^D \sum_{s \in \mathcal{S}} T_s \cdot v_s + \beta_1^D \sum_{s \in \mathcal{S}} \xi_s + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \psi_w^m. \tag{3.32}$$

The linear approximation of BLP is then

$$\underset{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}}{\text{maximize}} \quad \sum_{w\in\mathcal{W}} d_w^C \tag{3.33a}$$

(BLP-A) $$\text{subject to (3.24b)} - \text{(3.24d)}, \tag{3.33b}$$

$$(\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) = \underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}}{\text{argmin}}\ S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}), \tag{3.33c}$$

$$\text{subject to (3.31b)} - \text{(3.31e)}.$$

Theorem 4 presents our main result of this section. Let $C(\mathbf{x}) = \sum_{w\in W} d_w^C(\mathbf{x})$ be the cycling ridership under bike lane plan $\mathbf{x}$, i.e., the objective of BLP. Further, define $\underline{\alpha} = \min_{s\in\mathcal{S},\mathbf{x}\in\{0,1\}^{|\mathcal{S}|}} \{\alpha_s(\mathbf{x})\}$, $\underline{d} = \min_{w\in\mathcal{W}}\{d_w\}$, and $\bar{d} = \max_{w\in\mathcal{W}}\{d_w\}$.

Theorem 4 (**Suboptimality bound**). *Let* $\mathbf{x}^*$ *and* $\bar{\mathbf{x}}$ *be optimal solutions to the exact bike lane planning formulation BLP and the approximation BLP-A, respectively. Then*

$$C(\boldsymbol{x}^*) - C(\bar{\boldsymbol{x}}) \leq 2\cdot\mu_1\cdot\sqrt{|\mathcal{W}|}\left(\mu_2\cdot\frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi|-1} + \mu_3\cdot\frac{\sqrt{3\cdot|\mathcal{W}|}}{|\mathcal{R}_\psi|-1}\right), \tag{3.34}$$

where $\mu_1 = \max\left\{\bar{d}, \frac{1}{\beta_1^D\cdot\underline{\alpha}}\right\}$, $\mu_2 = \beta_1^D\cdot\alpha_s(\mathbf{x})\cdot(\bar{v}-\underline{v})$, *and* $\mu_3 = \log\left(\bar{d}\right) - \log\left(\underline{d}\right)$.

From Theorem 4, it can be seen that the error bound depends on the size of the road network (via $|\mathcal{S}|$ and $|\mathcal{W}|$), variation in traffic flows (μ_2) and variation in mode demands (μ_3), and that the dependence on the number of linear segments is $O\left(\frac{1}{|\mathcal{R}_\xi|} + \frac{1}{|\mathcal{R}_\psi|}\right)$. For practically-sized road networks, the optimality gap in Theorem 4 can be made small using a modest number of linear segments. For example, in our empirical study in Chapter 4, we use $|\mathcal{R}_\psi| = |\mathcal{R}_\xi| = 12$, which yields an upper bound on the optimality gap of 1.25% (see Appendix B.3.1).

Finally, to solve BLP-A, we leverage the fact that a bi-level program with a linear lower-level problem can be expressed exactly as a MILP (Fortuny-Amat and McCarl 1981). The complete formulation and additional technical details are contained in Appendix B.3.

3.4.3 Limiting Congestion

In practice, city planners are sensitive to the effect of bike lanes on increasing congestion (Khany 2022). Our model can accommodate these considerations via additional constraints on the increase in driving times. While there does not exist a universal measure of traffic congestion, we provide two possible examples of how congestion may be limited within our framework. The first approach is to ensure that the total demand-weighted driving time is increased by at most a factor of τ, which can be enforced by adding the following constraint to BLP-A:

$$\sum_{w \in \mathcal{W}} d_w^D \cdot t_w(\mathbf{x}, \mathbf{v}) \leq (1+\tau) \sum_{w \in \mathcal{W}} \tilde{d}_w^D \cdot t_w(\tilde{\mathbf{x}}, \tilde{\mathbf{v}}),$$

where $\tilde{d}_w^D$, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{v}}$ are the current driving demand, bike lane network and traffic flows, prior to the addition of any new bike lanes.

Alternatively, we may require that the worst case increase in driving times over all possible routes is at most a factor of τ. This can be enforced via the following constraints:

$$t_p(\mathbf{x}, \mathbf{v}) \leq (1+\tau) \cdot t_p(\tilde{\mathbf{x}}, \tilde{\mathbf{v}}), \quad p \in \mathcal{P}^D. \tag{3.35}$$

The choice of congestion control depends on the priorities of city planners. In particular, constraint (3.35) may be especially useful if planners wish to avoid sharp spikes in congestion in *any* part of the network, which can generate backlash and erode public support for bike lanes. In our empirical study in Chapter 4, we adopt this latter set of constraints, which allows for more granular control of congestion effects.

CHAPTER 4

Empirical Study: Expanding Chicago's Cycling Infrastructure

4.1 Introduction

In the preceding chapter, we have developed a framework to model the traffic equilibrium and mode choices of commuters for designing bike lanes, estimate the parameters of the equilibrium model, and optimize the bike lane expansion plan. In this chapter, we build upon the theoretical foundation by conducting an empirical study of bike lane expansion in the City of Chicago and examining its policy implications.

We select Chicago for our study because it is the one of the most congested cities in the United States (Inrix 2020), and also because expanding cycling infrastructure is a major policy priority in Chicago, which in 2021 announced plans to add 100 miles of bike lanes in the following two years (Chicago DOT 2021). Chicago is also experiencing rapid growth in the demand for cycling – from 2000 to 2018, the share of daily commuters that cycled more than tripled from 0.5% to 1.8% (US Census Bureau 2020).

New bike lines are typically planned based on current ridership and population density, without accounting for potential congestion effects (Chicago DOT 2020). Further, there are few extant methods for estimating the effect of changes to urban infrastructure on traffic congestion. To that end, the goal of our empirical study is two-fold: (1) to estimate the effect of expanding Chicago's cycling infrastructure on cycling ridership and traffic congestion, and (2) to compare the performance of our recommended bike lane network with alternative

methods that ignore ridership and congestion effects.

Our study combines seven detailed datasets: i) Chicago's road network, ii) the existing bike lane network, iii) observed traffic flows, iv) travel mode choices, v) bike share trips, vi) driving and cycling routes, and vii) taxi trips. We estimate that adding 25 miles of bike lanes as prescribed by our model can increase cycling ridership from 3.9% to 6.9% in downtown Chicago, without increasing the travel time of any driving route by more than 8%.

We evaluate our method by benchmarking against three intuitive heuristics for bike lane planning. In particular, we quantify the value of endogenizing congestion effects by comparing against a "traffic-agnostic" model, and find that ignoring traffic dynamics when designing bike lanes can needlessly worsen congestion. We also quantify the value of optimization in our setting by evaluating two heuristics that discard the network structure of the problem, and find that they can lead to smaller gains in cycling ridership with worse congestion outcomes.

The remainder of this chapter is organized as follows: Section 4.3 summarizes the datasets used in the study, Section 4.4 specifies the model for traffic congestion and mode choice and presents the associated estimation results for $(\boldsymbol{\theta}, \boldsymbol{\beta})$, Section 4.5 estimates the impact of several of our model's prescriptions on cycling ridership and traffic congestion, and Section 4.6 validates our method through a comparison with three intuitive benchmark methods for bike lane planning. Section 4.7 discusses policy implications and concludes.

4.2 Study Region

Our study focuses on a contiguous region of downtown Chicago, depicted as the shaded area in Figure 4.1. We select this region due to the avaiability of detailed cycling demand data which we use to estimate the mode choice parameters $\boldsymbol{\beta}$. When estimating the congestion parameter $\boldsymbol{\theta}$, we make use of vehicle traffic flow data from the entire City of Chicago (larger area in Figure 4.1), so to maximize the data available to our estimator.

4.3 Data Description and Sources

Here we provide an overview of the datasets and sources of data used in our empirical study, which is summarized in Table 4.1. Our study makes use of the following seven datasets:

1. Road network topology and features. To construct the road network, we first retrieve the the locations of 798 census tracts within the City of Chicago from 2010 U.S. census data (Chicago Data Portal 2010). We construct the node set $\mathcal{I}$ by assuming a node exists at the geometric centroid of each census tract. We construct the set of road segments $\mathcal{S}$ by querying OpenStreetMap (OSM 2021), which resulted in a total of 31,815 road segments within Chicago's city boundaries.

In addition to network structure, we also obtained data on three road features at the segment level: length (ℓ_s), number of lanes (n_s), and lane width (w_s). Segment length and number of lanes were extracted from OpenStreetMap. Lane widths are imputed based on guidelines from the Illinois Department of Transportation Bureau of Local Roads and Streets Manual (Illinois DOT 2018), which provides recommendation of lane widths based on road classification and traffic counts.

2. Bike lane network. We extract Chicago's bike lane network as of 2018 (Chicago Data Portal 2018). We match the segments of the bike lane network with the constructed road network, which allows us to construct the status quo bike lane network $\tilde{\mathbf{x}}$, representing a total of length of 265 miles. We assume bike lanes have a width of 1.5 meters in accordance with the Local Roads and Streets Manual (Illinois DOT 2018).

3. Vehicle traffic flows. We collect vehicle flow data through the network from the Highway Performance Monitoring System (HPMS) (FHA 2018). This dataset contains information on the Average Annual Daily Traffic (AADT) that flows through major road segments. We focus on average traffic flows during the morning rush hour (6am-10am) in the year 2018.[1] We match the road segments in the HPMS dataset onto the OpenStreetMap

[1]We focus on morning rush hours because they represent a sizeable share of all traffic (24% of all ve-

network using the ST-Matching algorithm proposed by Lou et al. (2009). Because HPMS does not contain traffic data for low volume road segments, we obtain AADT data for a subset of 12,447 segments in the network.

4. Travel mode choices. To construct the mode-specific demands, we first obtain commuter flow data from the Origin-Destination Employment Statistics dataset maintained by the Longitudinal Employer-Household Dynamics (LEHD) (US Census Bureau 2018b). The dataset specifies commuting flows between pairs of census tracts. We only focus on OD pairs with a strictly positive number of commuters and where the origin and destination census tracts are distinct. This resulted in a total of 146,847 OD pairs in our study region. We then retrieve census tract-level mode shares from the 2014 - 2019 benchmarking (ACS) (US Census Bureau 2020), which specifies mode shares of driving, cycling, and all other options for each census tract. Multiplying the mode shares with the total commuter flows provides an estimate of the total demand for each mode of transport *originating* at each census tract.[2]

5. Bike share trips. We computed the cycling demand d_w^C for each OD pair w using detailed cycling trip data made available by Divvy, Chicago's bike share system, which is operated by Lyft (Divvy 2021). The Divvy trip data contains information on individual bike rentals, including timestamps for the rental's start and end, and the addresses of the Divvy station that each bike was rented from and returned to. We first selected all trips taken in 2018 during the morning rush hour period (6am to 10am), resulting in a total of 487,632 cycling trips. We then geocoded the addresses of the Divvy stations and matched them to Chicago's census tracts to determine, for each origin census tract, the share of Divvy trips that arrived at each destination tract. Finally, for each origin census tract, we combined the computed destination trip-shares from Divvy with the total commuter flows and travel

hicle flows, Illinois Department of Transportation (2018)), and commuters are the main beneficiaries of improvements to cycling infrastructure (Chicago DOT 2020).

[2]Because the data from US Census Bureau (2020) contains demand at the origins, it only partially defines the driving demand d_w^D. In our study, we use this information to jointly estimate d_w^D along with $\boldsymbol{\theta}$.

mode shares from the 2014 - 2019 ACS data to determine the cycling demand d_w^C.

6. Driving and cycling routes. For each OD pair w, we constructed the path sets $\mathcal{P}_w^D$ and the cycling path p_w^C by querying the Google Direction Application Programming Interface (API) (Google Maps 2021). Specifically, we constructed $\mathcal{P}_w^D$ by obtaining the top three Google recommended driving routes for the given OD pair – using census tract centroids for the start and end locations – and matching the returned routes to the road network. In addition to providing routes, the Google Direction API also returns the travel time along each route; we queried the API at 12am to obtain the free-flow driving time parameters T_s, $s \in \mathcal{S}$.

Similarly, for each $w \in \mathcal{W}$, we constructed the cycling path p_w^C by retrieving the top-most recommended cycling path by Google Maps. We then constructed $\mathcal{S}_w^C$ (the set of segments that constitute the cycling path p_w^C) and the total biking distance L_w by matching the recommended cycling route with the bike lane network $\tilde{\mathbf{x}}$. This matching process also allowed us to compute the bike lane coverage $\rho_w(\tilde{\mathbf{x}})$ for each path p_w^C.

7. Taxi trips. We validate our estimates of the congestion parameter $\boldsymbol{\theta}$ using taxi trip data from Chicago. These records report the origin, destination, and duration of all taxi trips made within city boundaries (Chicago Data Portal 2021). For consistency with the traffic flow data used during estimation, we select all trips taken during morning rush hours (6am to 10am) from 2018, resulting in 1,649,214 trips between 4,400 OD pairs. The taxi trip data serves as an out-of-sample dataset for evaluating our model's ability to predict driving travel times in the network.

Table 4.1: Data summary and sources.

Source	**Quantity**	**Description**	**Size**
OSM (2021)	$\mathcal{I}$	Nodes	798
	$\mathcal{W}$	Origin-destination (OD) pairs	146,847
	$\mathcal{S}$	Road segments	31,815
	l_s	Road segment length	31,815
	n_s	Number of lanes	31,815
Google Maps (2021)	$\mathcal{P}^D$	Driving paths	399,356
	T_s	Free-flow travel times	31,815
	p_w^C	Cycling paths	146,847
Divvy (2021)	–	Individual cycling trips	487,632
	d_w^C	Cycling demand	5,272
Chicago Data Portal (2021)	$\mathcal{W}^{\text{taxi}}$	OD pairs for taxi trips	4,400
	N_w^{taxi}	Individual taxi trips	1,649,214
FHA (2018)	$\mathbf{z}$	Vehicle traffic flows (AADT)	12,447
	$\mathcal{S}^{\text{obs}}$	Segments with AADT data	12,447
Chicago Data Portal (2018)	$\tilde{\mathbf{x}}$	Existing bike lane network	1,289
Illinois DOT (2018)	w_s	Vehicle lane widths	31,815
US Census Bureau (2018b, 2020)	d_w^D	Driving demand	146,847

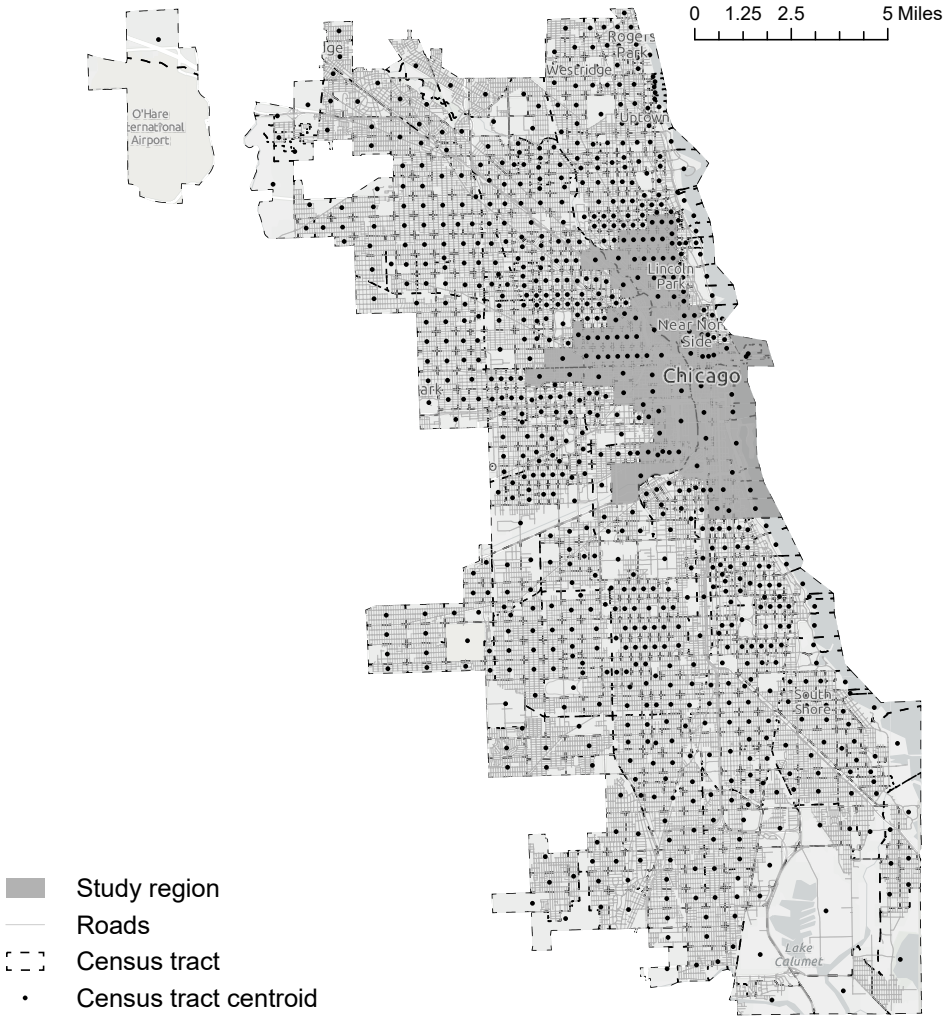


Figure 4.1: Study region: City of Chicago.

4.4 Model Estimation and Validation

We begin by specifying the congestion function. Traffic assignment models typically assume the travel time on a road segment is inversely related to its capacity (Patriksson 2015). To that end, for each segment $s \in \mathcal{S}$ we construct a feature vector related to the road's dimensions:

$$\mathbf{q}_s(x_s) = \begin{cases} \left(1, \ell_s, n_s \cdot w_s, \frac{\ell_s}{n_s \cdot w_s}\right), & \text{if } x_s = 0, \\ \left(1, \ell_s, n_s \cdot w_s - \Delta w, \frac{\ell_s}{n_s \cdot w_s - \Delta w}\right), & \text{if } x_s = 1, \end{cases} \tag{4.1}$$

where ℓ_s is the segment's length, $n_s \cdot w_s$ is the total width, $\ell_s/(n_s \cdot w_s)$ captures an inverse relationship with capacity, and the 1 permits an intercept in the congestion function. Following the Illinois Local Roads and Streets Manual (Illinois DOT 2018), we assume the presence of a bike lane ($x_s = 1$) narrows the road segment by 1.5 meters on each side, which implies $\Delta w = 3$ meters. For each $s \in \mathcal{S}$, the congestion function is then given by

$$\alpha_s(x_s, \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{q}_s(x_s) = \theta_0 + \theta_1 \cdot \ell_s + \theta_2 \left(n_s \cdot w_s - \Delta w \cdot x_s\right) + \theta_3 \left(\frac{\ell_s}{n_s \cdot w_s - \Delta w \cdot x_s}\right). \tag{4.2}$$

The segment travel time is then

$$t_s(x_s, \boldsymbol{\theta}, v_s) = \alpha_s(x_s, \boldsymbol{\theta}) \cdot v_s + T_s, \tag{4.3}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ is the parameter vector be estimated. As noted in Section 4.3, only the driving demand at the origin of each OD pair is available from the American Community Survey data. We therefore slightly modify the estimator WLS-A by jointly imputing the driving demand d_w^D with the congestion parameter $\boldsymbol{\theta}$ (see Appendix B.1 for details). In Appendix B.5, we present numerical evidence that this modification of WLS-A continues to produce reasonable estimates of $\boldsymbol{\theta}$.

Because AADT counts are not available for all possible road segments, we measure the flow error $\|\mathbf{v} - \mathbf{z}\|_2^2$ only over the segments for which traffic flow data is available ($|\mathcal{S}^{\text{obs}}| = 12,447$), while assuming Wardrop equilibrium holds over the full network ($|\mathcal{S}| = 31,815$). To evaluate the trade-off between the flow error $\|\mathbf{v} - \mathbf{z}\|_2^2$ and the Wardrop error ϵ, we re-solved

the estimator for each value of the penalty parameter λ in $\{1.0, 1.5, 2.0, 2.5, 3.0\} \times 10^3$. We validate the model fit at each value of λ by measuring how well it predicts the the travel times of a large number of taxi trips in Chicago (see Section 4.3).[3] Among the values of λ tested, we find $\lambda = 2500$ yields the minimal prediction error in taxi travel times, with an average absolute error of 1.34 minutes and an average relative error weighted by number of taxi trips of 14% (see Appendix B.1.3 for additional details on the selection of λ). Figure 4.2 depicts predictive performance on the taxi trip data. The correlation depicted in Figure 4.2(a) is 0.85, which suggests a sensible model fit. [4]

Table 4.2 reports the estimates for congestion parameter $\boldsymbol{\theta}$ under $\lambda = 2500$. The sign of the estimates suggest that travel time increases in segment length and decreases in total segment width (due to increased road capacity), which aligns with intuition.

Table 4.2: Congestion parameter estimates ($\times 10^{-7}$).

Intercept	ℓ_s	$n_s \cdot w_s$	$\ell_s/(n_s \cdot w_s)$
θ_0	θ_1	θ_2	θ_3
6.51	2.53	-0.26	75.7

Next, we estimate the mode choice parameters $\boldsymbol{\beta}$, using the estimated congestion parameter $\boldsymbol{\theta}$ as input. Table 4.3 summarizes the data used. To estimate $\boldsymbol{\beta}$, we restrict attention to the sub-network for which Divvy trip data is available (shaded region of Figure 4.1), which

[3]The penalty λ is a user-specified tuning parameter, which are typically chosen through cross validation. However, because cross-validation assumes independent observations, it is inapplicable in our setting due to the data being generated by a single connected network. We therefore use taxi trip data as our out-of-sample validation dataset.

[4]Our estimator's out-of-sample performance is comparable to extant methods for travel time prediction on road networks. For example, Zhan et al. (2013) present an MNL choice model for driving route selection, which has mean relative errors ranging from 17% to 41% in predicting the travel times of taxis in New York City. Furthermore, a naive prediction based on free-flow travel times from Google Maps yields an out-of-sample error of 27%, which underscores the value of accounting for traffic congestion in travel time prediction.

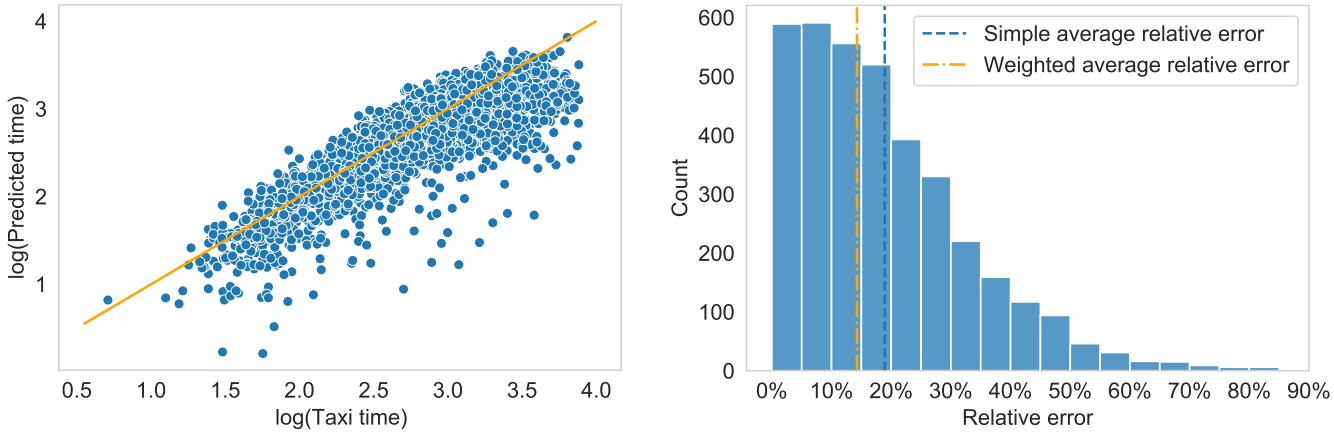


Figure 4.2: Driving time prediction errors on out-of-sample taxi trip data.

corresponds to in 5,272 OD pairs, 6,101 road segments, and 14,632 driving paths. Table 4.4 shows the estimation results, which show all parameter estimates to be statistically significant ($p < 0.05$). The signs of the estimates agree with intuition: The positive coefficient for driving time (β_1^D) indicates shorter travel times increase the attractiveness of driving. For cycling, the negative coefficient for bike lane coverage (β_1^C) implies the attractiveness of cycling increases in bike lane coverage; similarly, the positive coefficient for cycling distance (β_2^C) indicates cycling demand decreases in distance. Lastly, the estimated mode choice model yields $R^2 = 0.54$, which suggests a reasonable fit to the data.

Table 4.3: Summary of data for estimating mode choice parameters $\boldsymbol{\beta}$.

		Units	Mean	SD	Min.	25%	Median	75%	Max.
Bike lane coverage	ρ_w	%	50.8	26.0	0.0	31.6	54.1	71.9	100.0
Cycling distance	L_w	kilometers	5.90	3.34	0.31	3.45	5.50	7.65	21.26
Predicted driving times	t_w	minutes	14.14	4.62	3.64	10.86	14.15	17.30	30.78
Cycling mode share	$d_w^C/\tilde{d}_w$	%	4.78	11.92	0.00	0.20	0.51	3.42	98.17

Table 4.4: Estimates (and standard errors) of mode choice parameters.

Driving, u_w^D		Cycling, u_w^C			
intercept	t_w	intercept	ρ_w	L_w	R^2
β_0^D	β_1^D	β_0^C	β_1^C	β_2^C	
-3.341**	0.250**	0.263*	-1.172**	0.691**	0.54
(0.075)	(0.005)	(0.131)	(0.194)	(0.014)	

Note: $^*p < 0.05$; $^{**}p < 0.01$.

4.5 Bike Lane Expansion: Impact on Cycling Ridership and Traffic Congestion

Using the estimates $(\boldsymbol{\theta}, \boldsymbol{\beta})$, we use our model to develop bike lane plans for downtown Chicago and estimate the associated increase to cycling ridership and congestion. Our study focuses on expanding the existing bike lane network instead of redesigning the entire network, which aligns with the city's plan (Chicago DOT 2021). For tractability, we solve the path selection model over the largest OD pairs capturing 80% of commute demand in the study region (shaded area of Figure 4.1). This resulted in 887 candidate paths for bike lane installation, representing 4,127 road segments.

We consider 12 combinations of the bike lane budget B and the congestion tolerance τ: $(B, \tau) \in \{10, 25, 50, 75\} \times \{0.05, 0.1, 0.15\}$. For each combination of (B, τ), we add the congestion tolerance constraint (3.35) to the path selection model BLP-A given in (3.33), which is solved to obtain a recommended network $\mathbf{x}^*$. We then evaluate the model's recommendation by solving the (exact) user equilibrium problem (3.25) under $\mathbf{x}^*$ and computing the relative increase in cycling ridership over the status quo network $\tilde{\mathbf{x}}$:

$$\frac{\sum_{w \in \mathcal{W}} d_w^C(\mathbf{x}^*) - d_w^C(\tilde{\mathbf{x}})}{\sum_{w \in \mathcal{W}} d_w^C(\tilde{\mathbf{x}})}. \tag{4.4}$$

Consistent with the model BLP-A, we evaluate congestion using the worst-case increase in driving times over all driving paths:

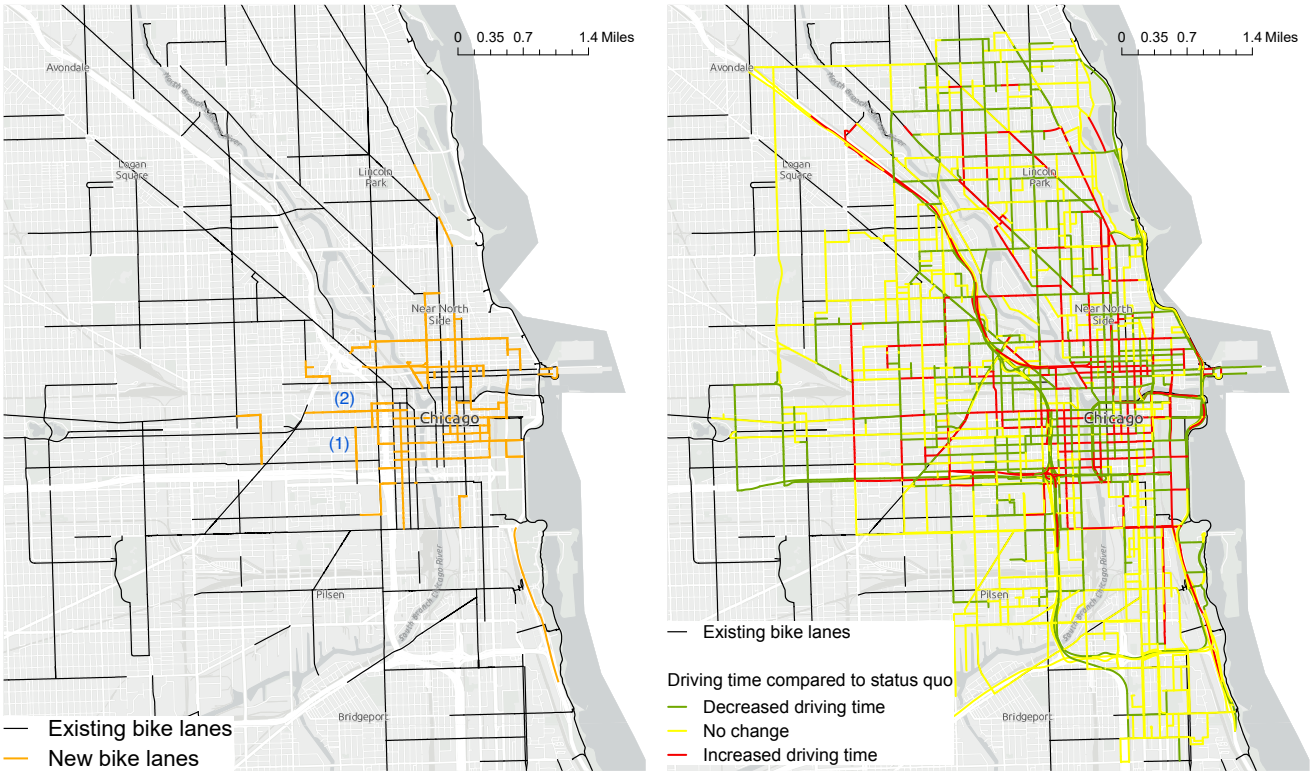
$$\max_{p \in \mathcal{P}^D} \left\{ \frac{t_p(\ \mathbf{x}^*) - t_p(\tilde{\mathbf{x}})}{t_p(\tilde{\mathbf{x}})} \right\}. \tag{4.5}$$

Our choice of the congestion constraint (3.35) (and the corresponding performance metric (4.5)) is motivated by discussions with city planners, who are sensitive to public perceptions regarding the installation of bike lanes and their effect on congestion (Khany 2022). Accordingly, to avoid "spikes" in driving times throughout the network, we focus on limiting the worst-case increase in driving times over all paths. As an alternative measure of congestion, we also estimate the impact of the recommended bike lane plans on the total system-wide driving time in Appendix B.4.

As an illustrative example, Figure 4.3 visualizes the model's recommended bike lane plan and the estimated effects on driving times for the case where $(B, \tau) = (25,\ 10\%)$. Figure 4.3a depicts existing and recommended bike lanes. Interestingly, we observe that the recommended bike lanes tend to close gaps in the existing bike lane network (e.g., (1) in Figure 4.3a) and also bridge distant parallel bike lanes (e.g., (2)), both of which are priorities for the Chicago DOT, even though we do not explicitly enforce these requirements.

Figure 4.3b shows the change in driving time on each road segment as compared to the status quo. We make the following observations. First, among the 25 miles for which new bike lanes are recommended, 77.6% have longer driving times and 8.2% have shorter driving times (no change is observed in the remaining 14.2% of segments). The intuition is as follows. Building bike lanes reduces road width and thus increases the driving time, which we refer to as the *road effect*. On the other hand, building bike lanes increases cycling utility and thus boosts biking adoption, which reduces driving demand and thus alleviates congestion, which we refer to as the *demand effect*. Figure 4.3b visualizes the segments where the road effect dominates the demand effect (red) and vice versa (green).

We also observe spillover effects of new bike lanes to the entire network. Specifically,



(a) Bike lane expansion.

(b) Change in driving times.

Figure 4.3: Recommended bike lane expansion and congestion effects in downtown Chicago for $(B, \tau) = (25, 10\%)$.

among the road segments on which no new bike lanes are built, 11.8% have longer driving times and 40.0% have shorter driving times. Because the demand effect spills over to the rest of the network, the driving time reduction is expected. The congestion-increasing road effect on some segments is due to the fact that adding bike lanes causes drivers to adopt alternative paths in equilibrium, which increases driving times elsewhere in the network. These spillover effects are contingent on network structure and traffic dynamics, which highlights the necessity of a system-wide approach to bike lane planning.

Figure 4.4 summarizes model performance under each of the 12 combinations of (B, τ). In general, we estimate notable increases to cycling ridership even when the congestion tolerance and the budget are minimal – for example, under $(B, \tau) = (10, 5\%)$, cycling shares

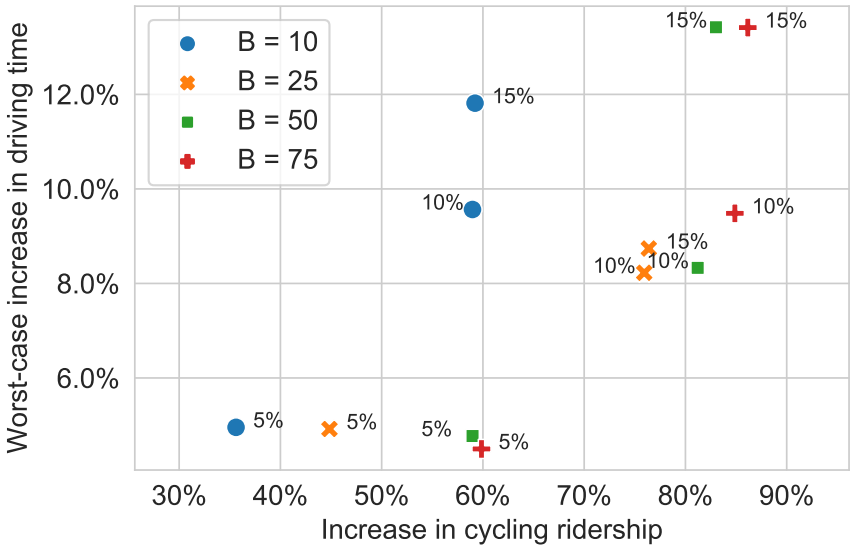


Figure 4.4: Increase in cycling ridership and worst-case driving travel times under 12 combinations of (B, τ).

increase by 36% (from 3.9% to 5.3%). Further, our results suggest diminishing returns on increasing the congestion tolerance τ – increasing τ from 5% to 10% boosts cycling ridership by 22% to 34%, depending on the bike lane mileage budget B, wheres increasing τ from 10% to 15% is associated with at most a 2% relative increase in cycling ridership. We observe a similar behavior as B increases, with most of the gains to cycling ridership occurring within the first 50 miles of bike lane expansion.

4.6 Comparison with Alternative Bike Lane Planning Methods

To further evaluate the performance of the bike lane plans recommended by our model, we benchmark its performance against three alternative methods for bike lane expansion.

Fixed-time model. In this approach, we assume driving times are unaffected by the addition of bike lanes. Specifically, we assume segment driving times remained fixed at their status quo values, $t_s(\tilde{\mathbf{x}})$, which are returned by the congestion parameter estimator described

in Section 3.3. We then solve a variant of the path selection model BLP-A with fixed driving times. The precise formulation for the fixed-time model is given in Appendix B.4.1.

Greedy heuristic. We also present a greedy heuristic for solving BLP-A, which is based on selecting promising OD pairs between which the city planner should build bike lanes while satisfying the congestion tolerance constraints. The greedy heuristic can be viewed as a congestion-aware bike lane planning method that does not take a system-wide perspective. Intuitively, the greedy heuristic consists of three components. First, the OD pairs are sorted from highest to lowest based on their potential for improving cycling ridership per unit of new bike lane built. Second, we screen the OD pairs by their anticipated increase in travel times due to bike lane construction, and then greedily select OD pairs such that the budget constraint is not violated. Third, the algorithm verifies whether the congestion constraint for the selected OD pairs is satisfied; if not, the heuristic is repeated with a more conservative selection (with respect to travel time increase) of OD pairs in the second step. Implementation details for the greedy heuristic are presented in Appendix B.4.2.

Demand heuristic. The last benchmark we consider is a naive planning approach that adds bike lanes on the paths for the OD pairs with the highest current commute demand, which aligns with one of Chicago's stated principles for building bike lanes (Chicago DOT 2020). Implementation details for the demand heuristic are presented in Appendix B.4.3.

Comparison results. We again vary $B \in \{10, 25, 50, 75\}$ and $\tau \in \{5\%, 10\%, 15\%\}$ to obtain 12 combinations of (B, τ) for comparing BLP-A with the three benchmark methods. For brevity, we present detailed results for $B = 25$ miles in Figure 4.5, and defer the cases for for $B \in \{10, 50, 75\}$ to Appendix B.4.4. Note that the fixed-time model and demand heuristic are not parameterized by τ, so they each correspond to only one point in Figure 4.5. We start interpreting the results by comparing our model's prescriptions (blue 'o') with the demand heuristic's bike lane plan (red '+'). As expected, because the demand heuristic places bike lanes between OD pairs with the most commuters, it obtains substantial growth in bike demand (increased by 59%), although the resulting increase in congestion

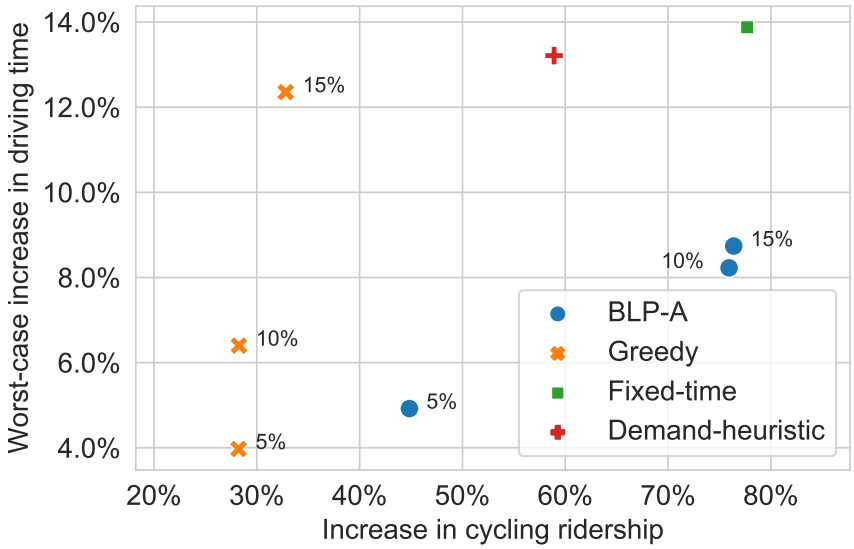


Figure 4.5: Increase in ridership and worst-case driving time for bike lane path selection model BLP-A and three benchmark methods ($B = 25$ miles).

is relatively high (13.2% increase in worst-case driving time). In particular, the demand heuristic is dominated by the recommendations from BLP-A when $\tau = 10\%$ or $\tau = 15\%$. With respect to the fixed-time model (green '□'), it increases cycling ridership by 78% (from 3.9% to 7.0%), which is comparable to our model's recommended bike lane expansion under $\tau = 10\%$ or 15%. However, the increase in congestion under the fixed-time model is significantly higher than under our approach (13.9% vs. 8.7%). This difference underscores the importance of endogenizing congestion effects in bike lane planning, which permits city planners to manage driving times throughout the network. Finally, we examine the greedy approach that considers congestion effects (orange '×') but discards the network structure of the problem. We find that our model's prescriptions achieve significantly higher bike demand across all values of $\tau \in \{5\%, 10\%, 15\%\}$. Specifically, when $\tau = 15\%$, the relative gain in cycling ridership under our method is more than twice that of the greedy heuristic (76% vs. 33%), while achieving lower increases in driving times as well (12.3% vs 8.7%). Taken together, these results demonstrate the value of endogenizing congestion and ridership

effects, as well as taking a holistic (i.e., network) perspective to urban bike lane planning.

4.7 Policy Implications and Conclusion

The expansion of cycling infrastructure has become a priority for municipal governments seeking to promote urban sustainability. While the addition of bike lanes has generated controversy due to their perceived effects on traffic congestion, few previous studies have attempted to rigorously estimate the impact of bike lane expansion on congestion from a system-wide perspective. Our findings suggest that adding bike lanes does not necessarily worsen congestion if done mindfully: We estimate that adding 25 miles of bike lanes to downtown Chicago as prescribed by our model can increase cycling ridership from 3.9% to 6.9%, while increasing driving times by at most 7.5% over all routes. We obtain qualitatively similar results for different bike lane budgets and congestion tolerances. Further, our results suggest that congestion may in fact be alleviated in many segments of the network following the addition of bike lanes. Taken together, our results cast doubt on the perception that bike lanes dramatically increase traffic congestion.

We also find that alternative planning methods that ignore congestion effects can amplify congestion without necessarily improving cycling ridership beyond what is attained under our model's prescriptions. To the extent that city planners wish to expand bike lane networks while mitigating congestion and the associated consequences, prescriptive models like ours that take a system-wide perspective may be useful. Our framework may be especially valuable for financially-constrained municipalities that wish to maximize the benefits of bike lane expansion under a limited budget, or in cities where the cycling infrastructure has matured past the point of "low hanging fruit" and would benefit from more rigorous planning methods. We also note that while sophisticated simulation models for traffic modeling are widely adopted by city planners, there is a scarcity of models for bike planning, which makes it challenging for city planners to systematically leverage traffic data when developing bike

lane plans (Khany 2022). As a result, data-driven methods like ours have the potential to improve the modeling support available to city planners.

There are many additional considerations to bike lane planning that are not captured by our model, including safety, implementation cost, closing gaps in the existing network, meeting needs of underserved communities, and land use. There may also be on-the-ground constraints that make some aspects of our model's recommendations impractical. To that end, our method should be viewed as a complement and not a substitute to existing approaches used by city planners for bike lane planning.

This study has limitations. First, although our method can be applied to other cities where similar data is available, our empirical findings may not generalize beyond Chicago due to differences in road network topology, commuters' preferences, and traffic patterns. Second, due to limitations in data availability, our analysis relies on demand data that is aggregated at the census-tract level – future analyses using more fine-grained data may yield different results. Our estimates of driving times are based on observable road segment features, such as length and width, and under the assumption of Wardrop equilibrium. In practice, traffic flows may also be influenced by additional factors such as traffic signal timing, permissibility of left and right turns, speed limits, and zone type (e.g. commercial vs. residential), all of which are unaccounted for in our study. Lastly, other factors that may influence cycling utility that we do not consider include safety, continuity of the bike lane network, and access to shops, services, green space, and public transit.

CHAPTER 5

Conclusions

In this thesis, we shed light on practical problems public sector face in smart city operations and urban mobility, using economic modeling, optimization, and data analytics as our guiding tools. Our primary focus is on the problems in promoting the adoption of public transit and cycling. Below we conclude each chapter and list potential future research directions.

Chapter 2 scrutinizes two mechanisms existing in practice aimed at promoting public transit ridership. The direct mechanism involves the government providing a ride-hailing subsidy to commuters who use a mix of ride-hailing and public transit, while also charging a congestion fee to those who drive personal vehicles. The direct mechanism involves partnering with a private enterprise to provide subsidies without charging a congestion fee. We show that the optimal incentives under these mechanisms depend on the level of public transit coverage, commuter preferences, and the adoption target. We also find that the direct mechanism is not budget neutral without undermining commuter welfare, while the indirect mechanism leaves all stakeholders better off and is only viable for modest adoption targets.

As one of the pioneering research studies in the literature investigating this issue, we offer several potential directions for future work. First, it may be fruitful to examine other forms of partnership between the transit agency, ride-hailing platform, and the private enterprise. For example, because the ride-hailing platform benefits from the commuter subsidy (due to increased demand), it may be worthwhile to examine partnerships where the cost of the subsidy is shared between the municipal government and the ride-hailing platform. Second, our study is relevant to other settings where two firms with "complementary capacity"

engage in a mutually beneficial partnership. For example, FedEx offers a "SmartPost" delivery service in which FedEx maintains responsibility for the long haul transportation of goods, while the United States Postal Service (USPS) handles the last mile delivery between a USPS center and a customer's home. Our study can serve as a springboard for analyzing partnerships of this nature. Third, we have modeled commuter welfare at an aggregate level; it may also be worthwhile to investigate mechanisms that can improve welfare for *all* commuters (e.g., by reducing traffic congestion for drivers). Lastly, budget neutral incentive mechanisms have received relatively little attention in the operations literature, but may be valuable in other settings where government funds are limited. Our findings suggest that an appropriately designed public-private partnership may be a cost-effective solution for governments that wish to improve adoption of other socially beneficial goods or services (e.g. electric vehicles).

Chapter 3 and 4 focus on the bike lane planning problem considering endogenous transportation mode demand and travel times. Our study provides rigorous estimate for the impact of bike lane expansion on congestion from a system-wide perspective. We find that adding 25 miles of bike lanes to downtown Chicago as prescribed by our model can increase cycling ridership from 3.9% to 6.9%, while increasing driving times by at most 7.5% over all routes. Our study demonstrates that careful implementation of bike lanes does not inevitably lead to worsened traffic congestion, and in many segments of the network, congestion may actually be alleviated.

We list potential directions for future work below. First, our method can be extended to other considerations such as the effect of bike lanes on greenhouse gas emissions or safety. It may also be worth examining how bike lane expansion affects different populations, such as low-income and underserved neighborhoods, or groups for whom cycling may not feasible, including seniors or people with disabilities. Finally, our modeling framework may also be useful in planning other types of transportation infrastructure where congestion impacts are a salient concern, such as the addition of dedicated bus or carpool lanes.

As sustainable urban mobility continues to receive increased attention, a plethora of research questions arise, such as inquiries into the design of transportation systems to enhance fairness and accessibility, as well as into the nature of public-private partnerships in pursuit of sustainability goals. Our proposed methodologies, which incorporate modeling and data analytics, offer a promising avenue for addressing these research questions.

APPENDIX A

Partnerships in Urban Mobility: Incentive Mechanisms for Improving Public Transit Adoption

A.1 General Hybrid Mechanism: Optimal Incentive (e^*, s^*, z^*)

In the previous sections, we find that mechanisms [D] and [I] capture two types of incentives that have already been implemented in practice: the funding of transit subsidies through congestion fees and partnerships with a private enterprise, respectively. By comparing both mechanisms, we found that mechanism [D] is costly to implement, but is capable of achieving higher public transit adoption target. Conversely, mechanism [I] is budget neutral, but only viable for small β. This raises a natural question: Is it possible to maintain budget neutrality while achieving a higher adoption target by combining both mechanisms?

In this section, we propose a general mechanism – denoted by [G] – that combines mechanisms [D] and [I]. In mechanism [G], all three levers, namely, the congestion fees e, the government subsidy s and the private subsidy z, are available. In other words, we allow $e \geq 0$, $s \geq 0$ and $z \geq 0$ in the general formulation presented in §2.2.2. Figure A.1 depicts the relationships among the various stakeholders under the general mechanism [G].

Based on criteria I – VI, the optimal incentive (e^*, s^*, z^*) under mechanism [G] is thus given by the solution to the following optimization problem. Note that similar to mechanism [I], because mechanism [G] requires the participation of the private enterprise, the subsidy

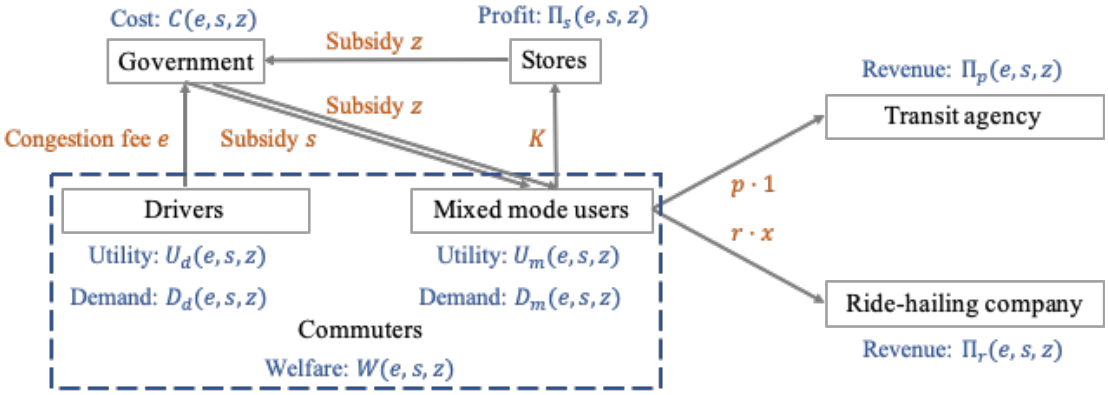


Figure A.1: Strategic interactions among stakeholders under the general mechanism [G].

z is set to be profit maximizing.

$$\min_{e,s,z\geq 0} \ [C(e,s,z)]^+ \tag{A.1a}$$

$$\text{s.t. } D_m(e,s,z) - D_m^0 \geq \beta D_m^0 \tag{A.1b}$$

Mechanism [G]:

$$W(e,s,z) \geq W^0 \tag{A.1c}$$

$$\Pi_s(e,s,z) \geq \Pi_s^0 \tag{A.1d}$$

$$z = \arg\max_z \Pi_s(e,s,z). \tag{A.1e}$$

We now characterize the optimal incentive (e^*, s^*, z^*) attained at a solution to (A.1). For ease of exposition, let $\gamma_1 \equiv (\delta-1)(x+1)$, $\gamma_2 \equiv (d-p) - x(r-d)$, $\breve{x} \equiv \frac{(d-p)-(\delta-1)+2K}{(r-d)+(\delta-1)}$, and $\beta_1(x) \equiv \frac{2\gamma_1 - 3\gamma_2 - \sqrt{(2\gamma_1-\gamma_2)^2 - 8\gamma_1 K}}{2\gamma_2} < \bar{\tau}(x)$, $\bar{\beta}(x) \equiv \frac{2K}{\gamma_2}$, where $\bar{\beta}(x) < \bar{\tau}(x)$ if and only if $x > \breve{x}$.

Proposition 7. *(i) If the last mile distance is small, $x \leq \breve{x}$, then for any $\beta \in (0, \bar{\tau}(x))$, there exists an optimal incentive program (e^*, s^*, z^*) that incurs no operating cost for the government, $[C(e^*, s^*, z^*)]^+ = 0$. Further,*

(a) if $0 \leq \beta \leq \tau(x)$, then $e^* \geq 0$, $s^* \geq 0$, and $z^* > 0$;

(b) if $\tau(x) < \beta < \bar{\tau}(x)$, then $e^* > 0$, $s^* \geq 0$, and $z^* > 0$.

(ii) Suppose the last mile distance is large, $x > \breve{x}$. Then

(a) if $0 \leq \beta \leq \bar{\beta}(x)$*, there exists an optimal incentive program* (e^*, s^*, z^*) *that incurs no operating cost for the government,* $[C(e^*, s^*, z^*)]^+ = 0$*. Further,*

i. if $0 \leq \beta \leq \tau(x)$*, then* $e^* \geq 0$*,* $s^* \geq 0$*, and* $z^* > 0$*;*

ii. if $\tau(x) < \beta \leq \beta_1(x)$*, then* $e^* > 0$*,* $s^* \geq 0$ *and* $z^* > 0$*;*

iii. if $\beta_1(x) < \beta \leq \bar{\beta}(x)$*, then* $e^* > 0$*,* $s^* > 0$*, and* $z^* > 0$*.*

(b) if $\bar{\beta}(x) < \beta < \bar{\tau}(x)$*, then the optimal incentive program* (e^*, s^*, z^*) *incurs a positive operating cost for the government,* $[C(e^*, s^*, z^*)]^+ > 0$*. Further,* $e^* > 0$*,* $s^* > 0$*, and* $z^* > 0$*.*

We offer a few remarks on Proposition 7. First, under mechanism [G], when the adoption target is modest, $\beta \leq \bar{\beta}(x)$, budget neutrality is attainable; when the adoption target is aggressive, $\beta > \bar{\beta}(x)$, the government has to bear a positive operating cost. Therefore, analogous to Remark 2, we can conclude that when adopting mechanism [G], a financially strained government should conservatively set the adoption target β to attain budget neutrality, while a sufficiently funded government can set a more aggressive target and further improve public transit adoption. Second, because $\tau(x) < \bar{\beta}(x)$ for any fixed x, it follows that mechanism [G] allows the government to remain budget neutral while achieving higher transit adoption target β than mechanism [I] and leaving all stakeholders (weakly) better off. Third, because the optimal private subsidy is always strictly positive, namely, $z^* > 0$ for all $\beta > 0$, we conclude that it is essential for the municipal government to partner with the private sector if the objective is to minimize operating cost. In addition, as β increases, to achieve more a ambitious adoption target, more interventions should be utilized, that is, $e^* > 0$ and $s^* > 0$.

Using the optimal solutions to mechanism [G], we are able to compare the optimal cost with mechanisms [D] and [I], which is summarized in Corollary 5.

Corollary 5. *For any adoption target* $\beta \in (0, \bar{\tau}(x))$*, the government should always adopt mechanism [G] over mechanisms [D] and [I]. Further, the minimal operating cost* $[C(e^*, s^*, z^*)]^+$

is higher when β is large.

The above corollary is driven by the fact that the optimal government cost under mechanism [G] is always less than (or equal to) mechanisms [D] and [I]. Intuitively, this occurs because the government can fund the commuter subsidy from both congestion fees and the private enterprise, instead of from just a single source. Specifically, when the adoption target is conservative, $\beta \leq \tau(x)$, both mechanisms [G] and [I] are budget neutral; when the target is ambitious, $\tau(x) < \beta < \bar{\tau}(x)$, the optimal cost under mechanism [G] is strictly less than that of mechanism [D]. The reason why the minimal operating cost under mechanism [G] increases in β is similar to intuition for mechanism [D], which we omit for conciseness.

To summarize, we find that when combining both mechanisms, mechanism [G] attains budget neutrality when β is modest, and incurs a strictly positive cost when β is aggressive. Specifically, in the budget neutral regime, mechanism [G] can satisfy a higher adoption target β than mechanism [I], and in the positive cost regime, mechanism [G] is less costly than mechanism [D]. To the best of our knowledge, there is no incentive mechanism that exists in practice that resembles mechanism [G]. Therefore, it may be worthwhile for municipal governments to investigate incentive programs that combine multiple sources of funding for last-mile subsidies (e.g., both congestion fees and private sector partnerships).

A.2 Alternative Formulation: Maximizing Public Transit Adoption as Objective Function

In addition to minimizing operating cost and maximizing commuter welfare, some municipal governments may have an earmarked budget for incentivizing public transit usage. To address this, in this section, we consider the incentive design problem of maximizing public transit adoption with a strictly positive budget $B > 0$, under both mechanisms [D] and [I].

A.2.1 Mechanism [D-O] with Transit Adoption Objective

In this section, we consider a variation of mechanism [D] where the objective is to maximize public transit adoption with a budget constraint, which we refer to as mechanism [D-O]. We then modify the original incentive design problem (2.16) as follows:

$$\min_{e,s\geq 0} \; D_m(e,s,0) - D_m^0 \tag{A.2a}$$

$$\text{s.t. } C(e,s,0) \leq B \tag{A.2b}$$

Mechanism [D-O]:

$$W(e,s,0) \geq W^0 \tag{A.2c}$$

$$\Pi_s(e,s,0) \geq \Pi_s^0. \tag{A.2d}$$

Note that when the budget B is very large, the government can induce all commuters to adopt the mixed mode by simply setting a high subsidy and not charging a congestion fee. To avoid this degenerate case, we assume $B < \bar{B}$, where $\bar{B} \equiv (\delta-1)(x+1) - ((d-p) - x(r-d))$. Then by solving problem (A.2), we obtain the optimal incentives e^* and s^* under Mechanism [D-O], given in Proposition 8 below. In preparation, let $\tilde{B} = \frac{((\delta-1)(x+1)-((d-p)-x(r-d)))^2}{2(\delta-1)(x+1)}$. It is straightforward to verify that $\tilde{B} < \bar{B}$.

Proposition 8. *For fixed x, the optimal incentives (e^*, s^*) under mechanism [D-O] depends on the size of the budget B. In particular,*

(i) When the budget is small, $B \leq \tilde{B}$, there exists a unique solution

$$\begin{aligned} e^* &= \frac{\sqrt{2B}((d-p) - x(r-d))}{\sqrt{(\delta-1)(x+1)}} + B, \\ s^* &= B\left(\frac{\sqrt{2}((\delta-1)(x+1) - ((d-p) - x(r-d)))}{\sqrt{B(\delta-1)(x+1)}} - 1\right). \end{aligned} \tag{A.3}$$

Upon implementation of the incentive program $(e^*, s^*, 0)$, mixed mode demand will increase to $D_m(e^*, s^*, 0) = \frac{\sqrt{2B(\delta-1)(x+1)}+(d-p)-x(r-d)}{(\delta-1)(x+1)}$.

(ii) When the budget is large, $B > \tilde{B}$, the optimal subsidy is in the interval $s^* \in [\tilde{B}, B]$. Further, for every $s^* \in [\tilde{B}, B]$, the corresponding optimal congestion fee satisfies $e^* \geq \bar{B} - s^*$, and all commuters take the mixed mode, $D_m(e^*, s^*, 0) = 1$.

Note that the mixed mode demand under the optimal incentive program is non-decreasing in the government budget B. This result is intuitive - as the government's budget increases, so does the size of the subsidy, which increases transit adoption. Because the number of commuters are limited, when the budget B is sufficiently large, it enables the government to provide a large enough subsidy such that all the commuters choose public transit, in which case the mixed mode demand is constant at $D_m(e^*, s^*, 0) = 1$.

A.2.2 Mechanism [I-O] with Transit Adoption Objective

Next, we consider a variant of mechanism [I] with the objective of maximizing public transit adoption, which we call mechanism [I-O]. The optimal incentive z^* under mechanism [I-O] is given by the solution to the following optimization problem:

$$\max_{z \geq 0} \; D_m(0, 0, z) - D_m^0 \tag{A.4a}$$

$$\text{s.t. } C(0, 0, z) \leq B \tag{A.4b}$$

Mechanism [I-O]:

$$W(0, 0, z) \geq W^0 \tag{A.4c}$$

$$\Pi_s(0, 0, z) \geq \Pi_s^0 \tag{A.4d}$$

$$z = \arg\max_{z} \Pi_s(0, 0, z) \tag{A.4e}$$

By solving problem (A.4), we obtain the optimal subsidy z^* under mechanism [I] as follows.

Proposition 9. *The optimal subsidy z^* under mechanism [I-O] is given by*

$$z^* = \min\left\{\hat{z}, (\delta - 1)(x + 1) - ((d - p) - x(r - d))\right\},$$

where $\hat{z} = \frac{1}{2}(K - ((d-p) - x(r-d)))$ *as defined in Proposition* (2)*. The resulting mixed mode demand is* $D_m(0,0,z^*) = \frac{\min\{\frac{1}{2}(K+((d-p)-x(r-d))),(\delta-1)(x+1)\}}{(\delta-1)(x+1)}$.

Proposition 9 reveals that the resulting transit demand under the optimal private subsidy program z^* is non-decreasing in the partner enterprise's unit profit K. To see why, note that when the enterprise's profit margin K is high, the enterprise's profit is highly sensitive to changes in D_m^0, so the private enterprise has a stronger incentive to boost the customer base $D_m(0,0,z)$. As a result, the optimal transit demand $D_m(0,0,z^*)$ is non-decreasing in K. However, when the profit margin K is sufficiently large, the enterprise sets a large enough subsidy z such that all commuters adopt the mixed mode, in which case the mixed mode demand is constant, $D_m(0,0,z^*) = 1$.

To summarize, the optimal mixed mode adoption under mechanism [D-O] (where the private enterprise is passive) depends on the government's budget B, and the optimal mixed mode adoption under mechanism [I-O] (where the government is passive) depends on the enterprise's profit margin K. Note that due to the additional budget parameter B, it is difficult to compare the relative effectiveness of mechanism [D-O] and mechanism [I-O] with respect to the government's operating cost and commuter welfare.

A.3 Commuter Heterogeneity Captured by Last Mile Length x

In the main mode, we capture commuter heterogeneity by assuming that commuters' unit valuation for the mixed mode V is uniformly distributed between 0 and 1, for any fixed last mile distance x. In this section, we assume V is deterministic and identical for all commuters, and instead the last mile x distance is uniformly distributed. Specifically, we assume commuters are uniformly located in the suburban area of length σ with density 1[1], so the mass of commuters is σ and their last-mile distance is $x \sim U[0,\sigma]$. In this case, we

[1]Our results also persist when the commuter density is generalized to be a constant λ. We assume the commuter density to be 1 for ease of exposition and in parallel with the original model.

also aim to find the optimal incentives under both mechanisms [D] and [I] based on criteria I – VI as defined in §2.2. The related metrics are defined as follows.

Note that in this case, commuter utilities for the mixed mode and driving are the same as defined in (2.8). Using (2.8), we have that under the incentive (e, s, z), a commuter will adopt the mixed mode if and only if $U_m(e, s, z) \geq U_d(e, s, z)$, or equivalently, if her last mile distance x satisfies:

$$x \leq x_1(e, s, z) \equiv \min \left\{ \sigma, \frac{(d-p) - (\delta - 1)V + e + s + z}{(r-d) + (\delta - 1)V} \right\}. \tag{A.5}$$

The demand for each mode is then derived as:

$$D_d(e, s, z) = \left[\sigma - \frac{(d-p) - (\delta - 1)V + e + s + z}{(r-d) + (\delta - 1)V} \right]^+, \tag{A.6a}$$

$$D_m(e, s, z) = \sigma - \left[\sigma - \frac{(d-p) - (\delta - 1)V + e + s + z}{(r-d) + (\delta - 1)V} \right]^+. \tag{A.6b}$$

Let $D_d^0 = D_d(0, 0, 0)$ and $D_m^0 = D_m(0, 0, 0)$. To exclude the degenerate cases where all commuters adopt the same travel mode, we assume that $V \in (\underline{V}, \bar{V})$, so that $D_m^0 > 0$ and $D_d^0 > 0$, where $\underline{V} = \max\{0, \frac{d-p-(r-d)\sigma}{(\delta-1)(\sigma+1)}\}$ and $\bar{V} = \frac{d-p}{\delta-1}$. Analogously, based on the transit adoption criterion II, we also assume throughout that the target β is restricted to the non-degenerate case where $D_d(e, s, z) > 0$ and $D_m(e, s, z) > 0$. To enforce the preceding inequalities, it suffices to assume that $\beta < \bar{\tau}(x) \equiv \frac{\sigma((r-d)+(\delta-1)V)}{(d-p)-(\delta-1)V} - 1$.

Next, we define the performance metrics to evaluate incentive (e, s, z) in the same fashion as §2.2.2. The total welfare accrued to all commuters is:

$$W(e, s, z) = \int_0^{x_1(e,s,z)} U_m(e, s, z) dV + \int_{x_1(e,s,z)}^{\sigma} U_d(e, s, z) dV, \tag{A.7}$$

and the operating cost of the government, public transit agency's revenue, ride-hailing platform's revenue and the total profit of the partner enterprise are the same as defined in (2.11), (2.13), (2.14), and (2.15). We denote the various metrics in the absence of any incentives (i.e., when $(e, s, z) = 0$) in the same way as in §2.2, namely by using superscript 0: $W(0, 0, 0) = W^0$, $\Pi_p(0, 0, 0) = \Pi_p^0$, $\Pi_r(0, 0, 0) = \Pi_r^0$, and $\Pi_s(0, 0, 0) = \Pi_s^0$.

A.3.1 Mechanism [D-H] with Heterogeneous x

In this section, we investigate the optimal congestion fees and government subsidies under mechanism [D] when assuming last mile distance $x \sim U[0, \sigma]$, which we shall refer to as mechanism [D-H]. Based on the same criteria given in I – VI, the optimal incentives e^* and s^* are given by the solution to the following problem:

$$\begin{aligned} & \min_{e,s\geq 0} \ [C(e,s,0)]^+ && \text{(A.8a)} \\ & \text{s.t. } D_m(e,s,0) - D_m^0 \geq \beta D_m^0 && \text{(A.8b)} \\ \text{Mechanism [D-H]:} \quad & W(e,s,0) \geq W^0 && \text{(A.8c)} \\ & \Pi_s(e,s,0) \geq \Pi_s^0. && \text{(A.8d)} \end{aligned}$$

Proposition 10. *The optimal congestion fee e^* and the optimal subsidy s^* under mechanism [D-H] are given by:*

$$\begin{aligned} e^* &= \frac{\beta(\beta+2)((d-p)-(\delta-1)V)^2}{2\sigma((r-d)+(\delta-1)V)}, \\ s^* &= \beta((d-p)-(\delta-1)V) - \frac{\beta(\beta+2)((d-p)-(\delta-1)V)^2}{2\sigma((r-d)+(\delta-1)V)}. \end{aligned} \tag{A.9}$$

Further, the government incurs positive operating cost: $C(e^*, s^*, 0) = \frac{\beta^2((d-p)-(\delta-1)V)^2}{2((r-d)+(\delta-1)V)} > 0$.

The above proposition resembles the results given in §2.3.1. Similar to mechanism [D], mechanism [D-H] enables the government to achieve the public transit adoption target β, but it takes a positive cost to implement, and the minimal operating cost $C(e^*, s^*, 0)$ increases in the adoption target β.

A.3.2 Mechanism [I-H] with Heterogeneous x

Analogous to §A.3.1, when assuming $x \sim U[0, \sigma]$, we denote mechanism [I] as mechanism [I-H]. Similar to the previous sections, we assume the participating enterprise has sufficient profit margin $K > \underline{K} = (d-p) - (\delta-1)V$. The optimal incentive z^* under mechanism [I-H]

is then given by the solution of the following optimization problem:

$$\max_{z\geq 0}[C(0,0,z)]^+ \tag{A.10a}$$

$$\text{s.t. } D_m(0,0,z) - D_m^0 \geq \beta D_m^0 \tag{A.10b}$$

Mechanism [I-H]:

$$W(0,0,z) \geq W^0 \tag{A.10c}$$

$$\Pi_s(0,0,z) \geq \Pi_s^0 \tag{A.10d}$$

$$z = \arg\max_z \Pi_s(0,0,z) \tag{A.10e}$$

Similar to formulation (2.18), the optimization problem (A.10) may be infeasible if β is large. The intuition is presented in §2.4, which we omit for conciseness. Define $\tau(x) = \min\{\frac{K}{(d-p)-(\delta-1)V} - 1, \bar{\tau}(x)\}$, where $\bar{\tau}(x) = \frac{\sigma((r-d)+(\delta-1)V)}{(d-p)-(\delta-1)V} - 1$. Note that $0 < \tau(x) \leq \bar{\tau}(x)$.

Proposition 11. *Mechanism [I-H] is feasible if and only if $\beta \leq \tau(x)$. When $\beta \leq \tau(x)$, then the optimal subsidy z^* is given by:*

$$z^* \quad = \quad \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\}, \tag{A.11}$$

where $\underline{z} = \beta((d-p)-(\delta-1)V)$, $\hat{z} = \frac{1}{2}(K - ((d-p)-(\delta-1)V))$ and $\bar{z} = \min\{K, \sigma((r-d)+(\delta-1)V)\} - ((d-p)-(\delta-1)V)$. Further, the government incurs no operating cost: $C(0,0,z^*) = 0$.

Analogous to §2.4.2, using the optimal incentives under mechanisms [D-H] and [I-H], we can compare the the performance of these two mechanisms. We use superscript D and I to denote quantities obtained at optimal solutions in mechanisms [D-H] and [I-H], respectively. In preparation, define $\tilde{\tau}(x) = \frac{\hat{z}^2 - (\hat{z}-z^*)^2}{K((d-p)-(\delta-1)V)}$.

Theorem 5. *The indirect mechanism [I-H] outperforms the alternative direct mechanism [D-H] on operating cost ($C^I = 0 < C^D$), mixed mode adoption ($D_m^I \geq D_m^D$), commuter welfare ($W^I > W^D$), ride-hailing platform profit ($\Pi_r^I \geq \Pi_r^D$), and transit agency revenue ($\Pi_p^I \geq \Pi_p^D$). Further,*

(i) if $\beta \leq \tilde{\tau}(x)$, mechanism [I-H] outperforms mechanism [D-H] on enterprise profit ($\Pi_s^I \geq \Pi_s^D$),

(ii) if $\beta \in (\tilde{\tau}(x), \tau(x)]$, mechanism [D-H] outperforms mechanism [I-H] on enterprise profit ($\Pi_s^I < \Pi_s^D$),

Comparing Theorem 5 to Theorem 1, we find that the dominance of the indirect mechanism is robust, regardless of whether the commuters heterogeneity is captured by random unit utility V or random last mile distance x. We draw the following conclusions in parallel with §2.4.2: When the adoption target is conservative, $\beta \leq \tau(x)$, the government should adopt the indirect mechanism [I-H]; when the adoption target is aggressive, $\beta > \tau(x)$, the government should adopt the direct mechanism [D-H].

A.4 Proofs

Lemma 1. *In mechanism [D], the following inequality holds for all (e, s) such that $D_m(e, s, 0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, or equivalently, $e + s \leq (\delta - 1)(x + 1) - ((d - p) - (r - d)x)$:*

$$[C(e, s, 0)]^+ \geq C(e, s, 0) \geq \frac{(\beta((d - p) - (r - d)x))^2}{2(\delta - 1)(x + 1)}.$$

Proof Note that when requiring $D_m(e, s, 0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, constraint (2.16b) implies

$$D_m(e, s, 0) - D_m^0 = \frac{e + s}{(\delta - 1)(x + 1)} \geq \beta D_m^0 = \beta \frac{(d - p) - (r - d)x}{(\delta - 1)(x + 1)}.$$

Because $\delta > 1$ and $x > 0$, it follows that

$$e + s \geq \beta((d - p) - (r - d)x), \tag{A.12}$$

which holds with equality only if constraint (2.16b) is binding. Next, using (2.11), the operational cost is given by

$$\begin{aligned} C(e, s, 0) &= sD_m(e, s, 0) - eD_d(e, s, 0) \\ &= \frac{e((d - p) - (r - d)x - (\delta - 1)(x + 1)) + s((d - p) - (r - d)x)}{(\delta - 1)(x + 1)} + \frac{(e + s)^2}{(\delta - 1)(x + 1)}. \end{aligned}$$

Further, replacing (e, s, z) with $(e, s, 0)$ in (2.8) and (2.12), we obtain $U_d(e, s, 0)$, $U_m(e, s, 0)$ and $W(e, s, 0)$. Then the welfare *increment* is given by

$$W(e,s,0) - W^0 = \left[\int_0^{v(e,s,0)} U_m(e,s,0)dV + \int_{v(e,s,0)}^1 U_d(e,s,0)dV\right] - \left[\int_0^{v^0} U_m^0 dV + \int_v^1 U_d^0 dV\right] \tag{A.13}$$

$$= \frac{e((d-p)-(r-d)x-(\delta-1)(x+1)) + s((d-p)-(r-d)x)}{(\delta-1)(x+1)}$$

$$+ \frac{(e+s)^2}{2(\delta-1)(x+1)}. \tag{A.14}$$

It follows that for all (e, s) such that $D_m(e, s, 0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, it holds that

$$C(e,s,0) = (W(e,s,0) - W^0) + \frac{(e+s)^2}{2(\delta-1)(x+1)}. \tag{A.15}$$

Because constraint (2.16c) implies $W(e, s, 0)-W^0 \geq 0$, it follows from (A.15) that $C(e, s, 0) \geq \frac{(e+s)^2}{2(\delta-1)(x+1)}$, which holds with equality only when constraint (2.16c) is binding. Further, because $x < \bar{x}$, $(d-p)-(r-d)x > 0$. It follows from (A.12) that for all (e, s),

$$[C(e,s,0)]^+ \geq C(e,s,0) \geq \frac{(e+s)^2}{2(\delta-1)(x+1)} \geq \frac{(\beta((d-p)-(r-d)x))^2}{2(\delta-1)(x+1)}, \tag{A.16}$$

Because $\frac{(\beta((d-p)-(r-d)x))^2}{2(\delta-1)(x+1)} \geq 0$, it follows that (A.16) holds with equality if and only if both constraints (2.16b) and (2.16c) are binding. □

Proof of Proposition 1. The proof proceeds in two steps. In Step 1, we find the optimal solution (e^*, s^*). In Step 2, we prove statements (i) and (ii).

Step 1. To solve for (e^*, s^*), we consider two cases based on the definition of $D_m(e, s, 0)$ in (2.10), which implies $D_m(e, s, 0) = \min\left\{1, \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}\right\}$: where (e, s) satisfies $D_m(e, s, 0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, and where (e, s) satisfies $D_m(e, s, 0) = 1$. Then, we show that the optimal cost in is lower in the first case.

Case 1: Suppose $D_m(e,s,0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$. This condition is equivalent to

$$e + s \leq (\delta - 1)(x + 1) - ((d - p) - (r - d)x), \tag{A.17}$$

which we add to problem (2.16) as an additional constraint in this case. Because the objective is to minimize $[C(e,s,z)]^+$, to find an optimal solution it suffices to construct a feasible solution (e_1, s_1) that attains the lower bound in Lemma 1. Let (e_1, s_1) be given by the unique solution to the equations

$$D_m(e,s,0) - D_m^0 = \beta D_m^0,$$
$$W(e,s,0) = W^0,$$

which correspond to constraints (2.16b) and (2.16c). Solving for (e,s) yields the following unique solution:

$$e_1 = \frac{\beta(\beta+2)((d-p)-(r-d)x)^2}{2(\delta-1)(x+1)}$$
$$s_1 = \frac{\beta((d-p)-(r-d)x)(2(\delta-1)(x+1)-(\beta+2)((d-p)-(r-d)x))}{2(\delta-1)(x+1)}.$$

Next, we show that (e_1, s_1) is feasible. By construction, (e_1,s_1) satisfies constraints (2.16b) and (2.16c) with equality. Note that $e_1 + s_1 = \beta((d-p)-(r-d)x)$. Because $0 < \beta < \bar{\tau}(x)$, (e_1, s_1) satisfies constraint (A.17). Next, because $\delta > 1$ and $x > \underline{x}$, we have $e_1 > 0$. In addition, because $0 < \beta < \bar{\tau}(x)$, we also have $s_1 > 0$. Further, for any (e,s), the revenue *increment* of nearby merchants is given by $\Pi_s(e,s,0) - \Pi_s^0 = K\frac{e+s}{(\delta-1)(x+1)}$, which follows from (2.15). Because $e_1 > 0$ and $s_1 > 0$, it follows that $\Pi_s(e,s,0) - \Pi_s^0 > 0$, and thus (e_1, s_1) satisfies constraint (2.16d). Hence, (e_1, s_1) is feasible to (2.16). Because formulation (2.16) minimizes $[C(e,s,0)]^+$, and (e_1, s_1) achieves the lower bound of $[C(e,s,0)]^+$ presented in Lemma 1, it follows that (e_1, s_1) is optimal to (2.16) when restricting (e,s) to satisfy $D_m(e,s,0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$.

Case 2: Suppose $D_m(e,s,0) = 1$. It follows that $D_d(e,s,0) = 0$ and $v(e,s,0) = 1$.

Constraint (2.16c) can then be rewritten as

$$W(e,s,0) - W^0 = \int_0^1 U_m(e,s,0)dV - \left[\int_0^{v^0} U_m^0 dV + \int_{v^0}^1 U_d^0 dV\right] \tag{A.18}$$

$$= s - \frac{((\delta-1)(x+1) - ((d-p) - x(r-d)))^2}{2(\delta-1)(x+1)}. \tag{A.19}$$

Let $\phi = \frac{((\delta-1)(x+1)-((d-p)-x(r-d)))^2}{2(\delta-1)(x+1)}$. Then constraint (2.16c) requires that (e,s) satisfies $s \geq \phi$. Using (2.11) and the fact that $D_m(e,s,0) = 1$, we have $C(e,s,0) = s \geq 0$. Therefore, $[C(e,s,0)]^+ = C(e,s,0) \geq \phi$. Because $0 < \beta < \bar{\tau}(x)$, it is straightforward to verify that $\phi > C(e_1,s_1,0) = [C(e_1,s_1,0)]^+$. Hence, for all (e,s) that satisfies $D_m(e,s,0) = 1$, $[C(e,s,0)]^+ > [C(e_1,s_1,0)]^+$. Therefore, the optimal solution to (2.16), denoted by (e^*,s^*), must satisfy $D_m(e^*,s^*,0) = \frac{(d-p)-(r-d)x+e^*+s^*}{(\delta-1)(x+1)}$. It follows that $(e^*,s^*) = (e_1,s_1)$.

Step 2. Next, we prove statements (i) and (ii). (i). It is straightforward to verify that $\frac{\partial e^*}{\partial \beta} = \frac{(\beta+1)((d-p)-(r-d)x)^2}{(\delta-1)(x+1)} > 0$. Next, by taking the first derivative of e^* with respect to x, we have $\frac{\partial e^*}{\partial x} = \frac{\beta(\beta+2)\left((x+1)^2(r-d)^2-(r-p)^2\right)}{2(\delta-1)(x+1)^2}$. Because $x \in (\underline{x},\bar{x})$ and $r > d > p > 0$, it follows that $(x+1)^2(r-d)^2 < (r-d+d-p)^2 = (r-p)^2$, and thus $\frac{\partial e^*}{\partial x} < 0$. (ii). Note that $\frac{\partial s^*}{\partial \beta} = \frac{((d-p)-(r-d)x)((\delta-1)(x+1)-(\beta+1)((d-p)-(r-d)x))}{(\delta-1)(x+1)}$. Because $\delta > 1$, $\beta < \bar{\tau}(x)$ and $x \in (\underline{x},\bar{x})$, it follows that $\frac{\partial s^*}{\partial \beta} > 0$. To compute $\frac{\partial s^*}{\partial x}$, first note that $\frac{\partial^2 s^*}{\partial x^2} = -\frac{\beta(\beta+2)(r-p)^2}{(\delta-1)(x+1)^3} < 0$, meaning s^* is strictly concave in x and $\frac{\partial s^*}{\partial x}$ decreases in x. We next check the first derivatives at the boundaries, i.e. $\frac{\partial s^*}{\partial x}\big|_{x=\bar{x}}$ and $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}$. Note $\frac{\partial s^*}{\partial x}\big|_{x=\bar{x}} = -\beta(r-d) < 0$. If $\delta < \delta_1 = 1+d-p$, then $\underline{x} = \frac{(d-p)-(\delta-1)}{(\delta-1)+(r-d)}$, and thus $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}} = \frac{1}{2}\beta((\beta+2)(\delta-1)+2(\beta+1)(r-d)) > 0$; if $\delta \geq \delta_1$, then $\underline{x} = 0$, and thus $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}} = -\frac{\beta(2(\delta-1)(r-d)-(\beta+2)(d-p)(2r-d-p))}{2(\delta-1)}$. Similarly, if $\delta < \delta_2 = \frac{(\beta+2)(d-p)(2r-d-p)}{2(r-d)} + 1$, $\frac{\partial s^*}{\partial x}\big|_{x=0} > 0$; if $\delta \geq \delta_2$, $\frac{\partial s^*}{\partial x}\big|_{x=0} \leq 0$. Because $\delta_2 - \delta_1 = \frac{\beta(d-p)(2r-d-p)+2(d-p)(r-p)}{2(r-d)} > 0$, we have $\delta_2 > \delta_1$. Therefore, if $\delta \geq \delta_2$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}} \leq 0$, and if $1 < \delta < \delta_2$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}} > 0$. Therefore, by continuity of $\frac{\partial s^*}{\partial x}$, if $\delta \geq \delta_2$, then s^* strictly decreases in $x \in (\underline{x},\bar{x})$. If $\delta < \delta_2$, then there exists $\tilde{x} \in (\underline{x},\bar{x})$ such that s^* strictly increases in x on $(\underline{x},\tilde{x})$ and strictly decreases in x on $[\tilde{x},\bar{x})$. □

Proof of Corollary 1. Using the expressions for (e^*, s^*) in Proposition 1, we have

$$e^* - s^* = \beta((d-p)-(r-d)x)\left(\frac{(\beta+2)((d-p)-(r-d)x)}{(\delta-1)(x+1)} - 1\right).$$

Because $x \in (\underline{x}, \bar{x})$, $(d-p)-(r-d)x > 0$. Further, because $\delta > 1$, we have $e^* \geq s^*$ if and only if $x \leq \bar{x} - \frac{(\delta-1)(r-p)}{(r-d)((\beta+2)(r-d)+\delta-1)}$. The result follows by setting $\hat{x} = \bar{x} - \frac{(\delta-1)(r-p)}{(r-d)((\beta+2)(r-d)+\delta-1)}$. Next, note that because $\delta > 1$, $r > p$ and $r > d$, it follows that $\hat{x} < \bar{x}$. Define $\bar{\delta} = (\beta+2)(d-p)+1$, which satisfies $\bar{\delta} > 1$. It remains to show that $\underline{x} \leq \hat{x}$ if and only if $\delta \geq \bar{\delta}$. From the definition of $\underline{x}$, if $\delta < 1 + d - p$, then $\underline{x} = \frac{(d-p)-(\delta-1)}{(\delta-1)+(r-d)}$. It is straightforward to check that $\hat{x} \in (\underline{x}, \bar{x})$. If $\delta \geq 1 + d - p$, then $\underline{x} = 0$, and thus $\hat{x} \leq \underline{x}$ only if $\delta \geq \bar{\delta}$, where $\bar{\delta} > 1 + d - p$. Therefore, $s^* > e^*$ holds for all $x \in (\underline{x}, \bar{x})$ if and only if $\hat{x} \leq \underline{x}$, which holds if and only if $\delta \geq \bar{\delta}$.

Proof of Corollary 2. Using the expressions for e^* and s^* from Proposition 1 and the cost function given in (2.11), the total cost under (e^*, s^*) is $C(e^*, s^*, 0) = \frac{\beta^2((d-p)-(r-d)x)^2}{2(\delta-1)(x+1)}$. Because $x < \bar{x}$ and $\delta > 1$, $C(e^*, s^*, 0) > 0$. By noting that $\frac{\partial C(e^*,s^*,0)}{\partial \beta} = \frac{\beta((d-p)-x(r-d))^2}{(\delta-1)(x+1)} > 0$, we conclude that $C(e^*, s^*, 0)$ increases in β. Next, note $\frac{\partial C(e^*,s^*,0)}{\partial x} = \frac{\beta^2\left((x+1)^2(r-d)^2-(r-p)^2\right)}{2(\delta-1)(x+1)^2}$. Because $x \in (\underline{x}, \bar{x})$, we have $0 < (x+1)(r-d) < (r-p)$. It follows that $\frac{\partial C(e^*,s^*,0)}{\partial x} < 0$. □

Proof of Proposition 2. From the definition in (2.11), we have $[C(0,0,z)]^+ = C(0,0,z) = 0$. Thus, under mechanism [I], the operating cost is always 0. It follows that any feasible solution to problem (2.18) is optimal. Therefore, solving for the optimal subsidy z^* is equivalent to finding the optimal solution to the following problem:

$$\max_{z \geq 0} \ \Pi_s(0,0,z) \tag{A.20a}$$

$$\text{s.t. } D_m(0,0,z) - D_m^0 \geq \beta D_m^0 \tag{A.20b}$$

$$W(0,0,z) \geq W^0 \tag{A.20c}$$

$$\Pi_s(0,0,z) \geq \Pi_s^0 \tag{A.20d}$$

The proof is composed of three steps. In Step 1, we prove (A.20) is feasible if and only

if $\beta \leq \tau(x)$, which is also a sufficient and necessary condition for formulation (2.18) to be feasible. Next, in Step 2, we solve for the optimal solution z^* to (A.20), which is also optimal to (2.18). Finally, in Step 3, we conduct the comparative statics analysis of z^*.

Step 1. To construct the sufficient and necessary condition for problem (A.20) to be feasible, we establish the conditions under which the feasible solution set is non-empty. Let us first define the feasible set. Let $\tilde{z} = (\delta-1)(x+1)-((d-p)-(r-d)x)$. To find all the feasible solutions to problem (A.20), we consider two cases based on the definition of $D_m(0,0,z)$ in (2.10), which implies $D_m(0,0,z) = \min\left\{1, \frac{(d-p)-(r-d)x+z}{(\delta-1)(x+1)}\right\}$: $D_m(0,0,z) = \frac{(d-p)-(r-d)x+z}{(\delta-1)(x+1)}$ and $D_m(0,0,z) = 1$. In preparation, note that because $\underline{x} < x < \bar{x}$ and $\delta > 1$, we have $(\delta-1)(x+1) > ((d-p)-x(r-d)) > 0$.

Case 1: Suppose $D_m(0,0,z) = \frac{(d-p)-(r-d)x+z}{(\delta-1)(x+1)}$, which is equivalent to $z \leq \tilde{z}$. In this case, we first show that any $z \geq 0$ satisfies constraint (A.20c). Using the welfare expression given in (2.12), the commuter welfare increment is $W(0,0,z) - W^0 = \frac{z(2(d-p)-2(r-d)x+z)}{2(\delta-1)(x+1)}$. Because $\delta > 1$ and $x \in (\underline{x}, \bar{x})$, we have $W(0,0,z) - W^0 \geq 0$ if $z \geq 0$. Therefore, constraint (A.20c) is satisfied for all $z \geq 0$. Next, we show that when restricting $z \leq \tilde{z}$, the feasible region of problem (A.20) is $[\underline{z}, \bar{z}]$, where $\underline{z} = \beta((d-p)-x(r-d))$ and $\bar{z} = \min\{K, (\delta-1)(x+1)\} - ((d-p)-x(r-d))$. To do so, we prove that constraint (A.20b) is satisfied if and only if $z \geq \underline{z}$, and constraint (A.20d) is satisfied if and only if $0 \leq z \leq K-((d-p)-x(r-d))$. First, using the definition of D_m^0 in (2.3), it is straightforward to verify that constraint (A.20b) holds if and only if $\frac{z}{(\delta-1)(x+1)} \geq \beta\frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}$. Because $\delta > 1$, it follows that constraint (A.20b) holds if and only if $z \geq \underline{z} = \beta((d-p)-x(r-d))$. Because $x \leq \bar{x}$, it follows that $\underline{z} > 0$. Next, using the revenue expression in (2.15), when restricting $z \leq \tilde{z}$, it is straightforward to verify that constraint (A.20d) holds if and only if $0 \leq z \leq K-((d-p)-x(r-d))$. Therefore, when adding the additional constraint $z \leq \tilde{z}$ to problem (A.20), the feasible region is $[\underline{z}, \min\{\tilde{z}, K-((d-p)-x(r-d))\}]$, or equivalently, $[\underline{z}, \bar{z}]$.

Case 2: Suppose $D_m(0,0,z) = 1$, which is equivalent to $z \geq \tilde{z}$. Let $\dot{z} = K(1-\frac{(d-p)-(r-d)x}{(\delta-1)(x+1)})$.

It is straightforward to verify that constraint (A.20d) is satisfied if and only if $z \leq \dot{z}$. Next, we show that constraints (A.20b) and (A.20c) are satisfied if $z \geq \tilde{z}$. First, note that

$$D_m(0,0,z) - D_m^0 = 1 - \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)} \geq \beta \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}$$

holds for $z \geq \tilde{z}$ because $\beta < \bar{\tau}(x)$; thus constraint (A.20b) is satisfied for all $z \geq \tilde{z}$. Next, from the definition in (2.10), if $z \geq \tilde{z}$, then

$$\begin{aligned} W(0,0,z) - W^0 &= \frac{(\delta-1)(x+1)((x+1)(2d-\delta) - 2p - 2rx + 2z + x + 1) - (-d(x+1) + p + rx)^2}{2((\delta-1)(x+1))} \\ &> \frac{(\delta-1)^2(x+1)^2 - ((d-p) - x(r-d))^2}{2((\delta-1)(x+1))} \\ &\geq 0, \end{aligned}$$

which implies constraint (A.20c) is satisfied for all $z \geq \tilde{z}$. Therefore, constraint (A.20d) is satisfied if and only if $z \leq \dot{z}$, and constraints (A.20b) and (A.20c) are satisfied if $z \geq \tilde{z}$.

To summarize Case 1 and Case 2, note that $\bar{z} \leq \tilde{z}$, and thus the feasible region of problem (A.20) is $[\underline{z}, \bar{z}] \cup [\tilde{z}, \dot{z}]$. Now that we have constructed the feasible set, next we prove $[\underline{z}, \bar{z}] \cup [\tilde{z}, \dot{z}]$ is non-empty if and only if $\beta \leq \tau(x)$. The derivation consists of two cases: $K < (\delta-1)(x+1)$ and $K \geq (\delta-1)(x+1)$. If $K < (\delta-1)(x+1)$, it is straightforward to verify that $\dot{z} < \tilde{z}$, in which case the feasible region is $[\underline{z}, \bar{z}]$; note that this interval is non-empty if and only if $\underline{z} \leq \bar{z}$, or equivalently, $\beta \leq \frac{K}{((d-p)-(r-d)x)} - 1$. If $K \geq (\delta-1)(x+1)$, then $\bar{z} = \tilde{z} \leq \dot{z}$, in which case the feasible region is $[\underline{z}, \dot{z}]$; note that this interval is non-empty if and only if $\underline{z} \leq \dot{z}$, or equivalently, $\beta \leq \frac{K}{(\delta-1)(x+1)} \bar{\tau}(x)$. Further, in this case where $K \geq (\delta-1)(x+1)$, it holds that $\bar{\tau}(x) \leq \frac{K}{(\delta-1)(x+1)} \bar{\tau}(x)$. To summarize, because $\beta < \bar{\tau}(x)$, the feasible region $[\underline{z}, \bar{z}] \cup [\tilde{z}, \dot{z}]$ is non-empty if and only if $\beta \leq \min\{\frac{K}{((d-p)-(r-d)x)} - 1, \bar{\tau}(x)\}$. By definition, $\tau(x) = \min\{\frac{K}{((d-p)-(r-d)x)} - 1, \bar{\tau}(x)\}$. Because $K > \underline{K}$, it follows that $\tau(x) > 0$. Therefore, (A.20) is feasible if and only if $\beta \leq \tau(x)$, which is also sufficient and necessary for formulation (2.18) to be feasible.

Step 2. In this step, we solve for the optimal solution z^*. To do so, we first show that problem (A.20) is equivalent to the following problem:

$$\max \Pi_s(0,0,z) \text{ s.t. } z \in [\underline{z}, \bar{z}]. \tag{A.21}$$

To do so, we consider two cases: $K < (\delta-1)(x+1)$ and $K \geq (\delta-1)(x+1)$. If $K < (\delta-1)(x+1)$, we have showed in Step 1 that the optimal solution z^* lies in the feasible region and satisfies $z^* \in [\underline{z}, K-((d-p)-x(r-d))]$. If $K \geq (\delta-1)(x+1)$, it follows from Step 1 that for any β that satisfies $0 < \beta \leq \tau(x)$, the feasible region is $[\underline{z}, \dot{z}]$, where $\dot{z} = K\left(1 - \frac{(d-p)-(r-d)x}{(\delta-1)(x+1)}\right) \geq \tilde{z}$. Note that when $z \in [\tilde{z}, \dot{z}]$, from the definition in (2.15), we have $\Pi_s(0,0,z) = K - z$, which decreases in z. It follows that $\Pi_s(0,0,\tilde{z}) > \Pi_s(0,0,z)$ for all $z > \tilde{z}$. Therefore, when $K \geq (\delta-1)(x+1)$, to maximize $\Pi_s(0,0,z)$, the optimal solution z^* should satisfy $z^* \in [\underline{z}, \tilde{z}]$, where $\tilde{z} = (\delta-1)(x+1) - ((d-p)-(r-d)x)$. To conclude, z^* must satisfy $z^* \in [\underline{z}, \min\{K, (\delta-1)(x+1)\} - ((d-p)-x(r-d))]$, or equivalently, $z^* \in [\underline{z}, \bar{z}]$. Thus formulation (A.20) is equivalent to formulation (A.21). Next, we solve problem (A.21). Using the profit expression given in (2.15), and the fact that $\delta > 1$, we have $\frac{\partial^2 \Pi_s(0,0,z)}{\partial z^2} = -\frac{2}{(\delta-1)(x+1)} < 0$. Hence, $\Pi_s(0,0,z)$ is strictly concave in z. The solution to the first order condition $\frac{\partial \Pi_s(0,0,z)}{\partial z} = 0$ is $\hat{z} = \frac{K}{2} - \frac{1}{2}((d-p)-(r-d)x)$. Therefore, the unique optimal subsidy z^* is given by

$$z^* = \begin{cases} \underline{z} & \text{if } \hat{z} < \underline{z}, \\ \hat{z} & \text{if } \underline{z} \leq \hat{z} \leq \bar{z}, \\ \bar{z} & \text{if } \bar{z} < \hat{z}, \end{cases}$$

or, equivalently, $z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\}$. Note that z^* is the unique optimal solution to problem (A.21), so z^* is the unique feasible, and thus optimal, solution to (2.18).

Step 3. Because $\bar{z} = \min\{K, (\delta-1)(x+1)\} - ((d-p)-x(r-d))$, we can rewrite z^* as $z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}_1, \bar{z}_2\}$, where $\bar{z}_1 = K - ((d-p)-x(r-d))$ and $\bar{z}_2 = (\delta-1)(x+$

1) − ((d − p) − x(r − d)). Note $\frac{\partial \underline{z}}{\partial x} = -\beta(r-d) < 0$, $\frac{\partial \hat{z}}{\partial x} = \frac{r-d}{2} > 0$, $\frac{\partial \bar{z}_1}{\partial x} = r - d > 0$ and $\frac{\partial \bar{z}_2}{\partial x} = (\delta - 1) + (r - d) > 0$. Because $\underline{z}, \hat{z}, \bar{z}_1$, and $\bar{z}_2$ are all continuous in x, z^* is continuous in x. Therefore, if $z^* = \underline{z}$, z^* decreases in x; otherwise, z^* increases in x. It remains to show the conditions under which $z^* = \underline{z}$. In preparation, let $\tilde{x} = \bar{x} - \frac{K}{(2\beta+1)(r-d)}$ and $K_1 = (\bar{x} - \underline{x})(2\beta + 1)(r - d)$, and note that $\tilde{x} < \bar{x}$ by definition of $\tilde{x}$ and because $r > d$. Note that $z^* = \underline{z}$ if and only if $\hat{z} \leq \underline{z}$, which holds if and only if $x \leq \tilde{x}$. Because $x \in (\underline{x}, \bar{x})$, $x \leq \tilde{x}$ can hold only if $\tilde{x} > \underline{x}$. It can be verified that $\tilde{x} > \underline{x}$ if and only if $K < K_1$. Note that $K_1 = (\bar{x} - \underline{x})(2\beta + 1)(r - d) \geq (d - p) - (r - d)\underline{x} > (d - p) - (r - d)x = \underline{K}$. Therefore, (i) if $K \geq K_1$, z^* increases in x on $(\underline{x}, \bar{x})$; (ii) if $\underline{K} < K < K_1$, $z^* = \underline{z}$ decreases in x on $(\underline{x}, \tilde{x}]$ and increases in x on $(\tilde{x}, \bar{x})$. The result follows. □

Proof of Theorem 1. First, we compare the performances of the two mechanisms in metrics I – V. By Corollary 2, we have $C^D = C(e^*, s^*, 0) > 0$, and under mechanism [I], $C^I = C(0, 0, z^*) = 0$. Hence, $C^I = 0 < C^D$. Next, using the expressions for e^* and s^* given in Proposition 1, and the expression for z^* given in Proposition 2, it can be shown that $D_m(0, 0, z^*) - D_m^0 \geq \beta D_m^0 = D_m(e^*, s^*, 0) - D_m^0$, or equivalently, $D_m^I \geq D_m^D$. Because $D_m^I \geq D_m^D$, from (2.13) and (2.14) we immediately obtain $\Pi_p^I \geq \Pi_p^D$ and $\Pi_r^I \geq \Pi_r^D$, respectively. To compare the commuter welfare in mechanism [D] and mechanism [I], it is easy to check that $W^D = W^0$, because constraint (2.16c) is binding in equilibrium. Further, $W^I - W^0 = W(0, 0, z^*) - W^0 = \frac{z^*(2(d-p)-2(r-d)x+z^*)}{2(\delta-1)(x+1)}$, which satisfies $W^I - W^0 > 0$ if and only if $z^* > 0$. To see that $z^* > 0$ holds, let us first define $\gamma_1 = (\delta - 1)(x + 1) > 0$ and $\gamma_2 = (d - p) - x(r - d) > 0$. Then $\underline{z} = \beta\gamma_2 > 0$, $\hat{z} = \frac{K-\gamma_2}{2}$, $\bar{z} = \min\{K - \gamma_2, \gamma_1 - \gamma_2\}$ and $\tau(x) = \min\{\frac{K}{\gamma_2} - 1, \frac{\gamma_1}{\gamma_2} - 1\}$. Because $K > \underline{K} = \gamma_2$, it follows $\hat{z} = \frac{K-\gamma_2}{2} > 0$ and $\bar{z} = \min\{K - \gamma_2, \gamma_1 - \gamma_2\} > 0$. Therefore, $z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\} > 0$, and thus $W^I > W^0 = W^D$.

For the enterprise's profit, using the definition in (2.15), we have $\Pi_s^I - \Pi_s^D = \frac{\hat{z}^2 + K\gamma_2 - (z^* - \hat{z})^2}{\gamma_1} - \frac{K\gamma_2(\beta+1)}{\gamma_1}$. Define $\tilde{\tau}(x) = \frac{\hat{z}^2 - (\hat{z} - z^*)^2}{K\gamma_2}$. Then $\Pi_s^I < \Pi_s^D$ holds if and only if $\beta > \tilde{\tau}(x)$. Note mech-

anism [I] is feasible if and only if $\beta \in (0, \tau(x)]$, by Proposition 2. To see there exists β such that both $\beta > \tilde{\tau}(x)$ holds and mechanism [I] is feasible, it suffices to show $0 \leq \tilde{\tau}(x) \leq \tau(x)$ for all x. Because $z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\} \in \{\underline{z}, \hat{z}, \bar{z}\}$, we shall prove $0 \leq \tilde{\tau}(x) \leq \tau(x)$ for all x by considering three cases: $z^* = \underline{z}$, $z^* = \hat{z}$ and $z^* = \bar{z}$. Because $0 < \beta \leq \tau(x)$, it follows that $\underline{z} \leq \bar{z}$, and thus the three cases correspond to $\hat{z} \leq \underline{z}$, $\underline{z} < \hat{z} \leq \bar{z}$, and $\bar{z} < \hat{z}$.

Case 1: Suppose $\hat{z} \leq \underline{z}$. Then $z^* = \underline{z}$. It follows that $\tilde{\tau}(x) = \frac{\hat{z}^2 - (\hat{z}-\underline{z})^2}{K\gamma_2} = \frac{2\underline{z}\hat{z} - \underline{z}^2}{K\gamma_2} \leq \frac{\underline{z}^2}{K\gamma_2} \leq \frac{\bar{z}^2}{K\gamma_2} = \frac{(\min\{K,\gamma_1\}-\gamma_2)^2}{K\gamma_2} < \frac{(\min\{K,\gamma_1\}-\gamma_2)\min\{K,\gamma_1\}}{K\gamma_2} = \tau(x)\min\{1, \frac{\gamma_1}{K}\} \leq \tau(x)$. Further, because $\beta \leq \tau(x) = \min\left\{\frac{K}{\gamma_2} - 1, \frac{\gamma_1}{\gamma_2} - 1\right\}$, it follows that $\beta\gamma_2 \leq K - \gamma_2$, or equivalently, $z^* = \underline{z} \leq 2\hat{z}$. Thus $\tilde{\tau}(x) = \frac{\hat{z}^2 - (\underline{z}-\hat{z})^2}{\gamma_1\gamma_2} \geq 0$. Therefore, in this case, $0 \leq \tilde{\tau}(x) \leq \tau(x)$ holds for all x.

Case 2: Suppose $\underline{z} < \hat{z} \leq \bar{z}$. Then $z^* = \hat{z}$ and $\tilde{\tau}(x) = \frac{\hat{z}^2}{K\gamma_2} \leq \frac{\bar{z}^2}{K\gamma_2}$. In Case 1 we show that $\tilde{\tau}(x) \leq \frac{\bar{z}^2}{K\gamma_2} < \tau(x)$. Further, it is straightforward to verify that $\tilde{\tau}(x) > 0$. Therefore, $0 \leq \tilde{\tau}(x) \leq \tau(x)$ holds for all x.

Case 3: Suppose $\bar{z} < \hat{z}$. Then $z^* = \bar{z}$. Because $\bar{z} = \min\{\gamma_1 - \gamma_2, K - \gamma_2\}$ and $\hat{z} = \frac{K-\gamma_2}{2} < K - \gamma_2$, $\bar{z} \leq \hat{z}$ implies $\bar{z} = \gamma_1 - \gamma_2 < K - \gamma_2$, or equivalently, $\gamma_1 < K$. It follows that $\tau(x) = \frac{\gamma_1}{\gamma_2} - 1$, and thus $\tilde{\tau}(x) = \frac{2\bar{z}\hat{z} - \bar{z}^2}{K\gamma_2} = \frac{2(\gamma_1-\gamma_2)\frac{K-\gamma_2}{2} - (\gamma_1-\gamma_2)^2}{K\gamma_2} = \tau(x)(1 - \frac{\gamma_1}{K}) < \tau(x)$. It is straightforward to verify that $\tilde{\tau}(x) > 0$. Therefore, $0 \leq \tilde{\tau}(x) \leq \tau(x)$ holds for all x.

To summarize the three cases above, (i) if $\beta \leq \tilde{\tau}(x)$, then $\Pi_s^I \geq \Pi_s^D$; (ii) if $\beta \in (\tilde{\tau}(x), \tau(x)]$, then $\Pi_s^I < \Pi_s^D$. The result follows.

Proof of Proposition 3. Similar to the proof of Proposition 1, the proof proceeds in two steps. In Step 1, we find the optimal solutions (e^*, s^*). In Step 2, we prove the comparative statics results.

Step 1. To solve for (e^*, s^*), we consider two cases: $(D_m(e, s, 0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)})$ and $(D_m(e, s, 0) = 1)$. We then show that the optimal commuter welfare is higher in the latter case.

Case 1: Suppose $D_m(e,s,0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, which is equivalent to

$$e + s \leq (\delta - 1)(x + 1) - ((d - p) - (r - d)x). \tag{A.22}$$

The proof proceeds in two steps. First, we construct an upper bound for the optimal value of the objective function (2.20a) when adding (A.22) as an additional constraint to problem (2.20); second, to show optimality, we construct a feasible solution that attains this bound. First, because $D_m(e,s,0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$, we can apply (A.15) and rewrite constraint (2.20c) as

$$C(e,s,0) = (W(e,s,0) - W^0) + \frac{(e+s)^2}{2(\delta-1)(x+1)} = 0,$$

which implies any feasible (e,s) should satisfy $W(e,s,0) - W^0 = -\frac{(e+s)^2}{2(\delta-1)(x+1)}$. Next, constraint (2.20b) can be rewritten as $e+s \geq \beta((d-p)-(r-d)x)$. Note $\beta((d-p)-(r-d)x) > 0$ because $x \in (\underline{x}, \bar{x})$. Therefore,

$$W(e,s,0) - W^0 = -\frac{(e+s)^2}{2(\delta-1)(x+1)} \leq -\frac{(\beta((d-p)-x(r-d)))^2}{2(\delta-1)(x+1)},$$

which holds with equality if and only if constraint (2.20b) is binding. Next, we prove that the upper bound is attainable. We first solve for (e,s) such that constraint (2.20b) is binding and constraint (2.20c) is satisfied:

$$\begin{aligned} D_m(e,s,0) - D_m^0 &= \beta D_m^0 \\ C(e,s,0) &= 0. \end{aligned} \tag{A.23}$$

Solving the above set of equations yields the unique solution

$$\begin{aligned} e_2 &= \frac{\beta(\beta+1)((d-p)-(r-d)x)^2}{(\delta-1)(x+1)} \\ s_2 &= \frac{\beta((d-p)-(r-d)x)((\delta-1)(x+1)-(\beta+1)((d-p)-(r-d)x))}{(\delta-1)(x+1)}. \end{aligned}$$

Using similar arguments in proof of Proposition 1, it is plain to verify that (e_2, s_2) satisfies the additional constraint (A.22) and it's feasible to problem (2.20), so (e_2, s_2) is the unique optimal solution to formulation (2.20) when restricting $D_m(e,s,0) = \frac{(d-p)-(r-d)x+e+s}{(\delta-1)(x+1)}$. The corresponding optimal commuter welfare increment is $\Delta W(e_2, s_2, 0) = -\frac{(\beta((d-p)-x(r-d)))^2}{2(\delta-1)(x+1)}$.

Case 2: Suppose $D_m(e,s,0)=1$, which is equivalent to $e+s\geq(\delta-1)(x+1)-((d-p)-(r-d)x)$. It follows that $D_d(e,s,0)=0$. Then the budget neutrality constraint (2.20c) can be rewritten as

$$C(e,s,0)=sD_m(e,s,0)-eD_d(e,s,0)=s=0. \tag{A.24}$$

From $D_m(e,s,0)=1$, it follows that $e\geq(\delta-1)(x+1)-((d-p)-(r-d)x)$. From the definition in (2.12), the commuter welfare increment is

$$\Delta W(e,s,0)=-\frac{((\delta-1)(x+1)-((d-p)-x(r-d)))^2}{2(\delta-1)(x+1)}. \tag{A.25}$$

Because $0<\beta<\bar{\tau}(x)$, it can be verified that for all (e,s) that satisfies $D_m(e,s,0)=1$, $\Delta W(e,s,0)<\Delta W\left(e_2,s_2,0\right)$. Therefore, the unique optimal solution to (2.20), denoted by (e^*,s^*), must satisfy $D_m\left(e^*,s^*,0\right)=\frac{(d-p)-(r-d)x+e^*+s^*}{(\delta-1)(x+1)}$ and $(e^*,s^*)=(e_2,s_2)$.

Step 2. Next, we prove statements (i)-(iii) in order. (i). We first take the first derivative of e^* with respect to x and β. Because $\delta>1$, $0<x<\bar{x}$ and $r>d>p>0$, it follows that $\frac{\partial e^*}{\partial\beta}=\frac{(2\beta+1)((d-p)-(r-d)x)^2}{(\delta-1)(x+1)}>0$. Next, $\frac{\partial e^*}{\partial x}=\frac{\beta(\beta+1)\left((x+1)^2(r-d)^2-(r-p)^2\right)}{(\delta-1)(x+1)^2}$. Because $r-p=(d-p)+(r-d)>(r-d)(x+1)$, it follows that $\frac{\partial e^*}{\partial x}<0$. Therefore, e^* strictly increases in in β and strictly decreases in x. (ii). First, note that s^* is concave in x: $\frac{\partial^2 s^*}{\partial x^2}=-\frac{2\beta(\beta+1)(p-r)^2}{(\delta-1)(x+1)^3}<0$. Then, to show the result, it suffices to check the first derivatives at the boundaries: $\frac{\partial s^*}{\partial x}\big|_{x=\bar{x}}$ and $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}$. If $\delta<\delta_1=1+d-p$, then $\underline{x}=\frac{(d-p)-(\delta-1)}{(\delta-1)+(r-d)}$, and $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}=\beta((\beta+1)(\delta-1)+(2\beta+1)(r-d))>0$; if $\delta\geq\delta_1$, then $\underline{x}=0$, and $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}=\frac{\beta((\beta+1)(d-p)(2r-d-p)-(\delta-1)(r-d))}{\delta-1}$. If $\delta<\delta_3=\frac{(\beta+1)(d-p)(2r-d-p)}{r-d}+1$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}>0$; if $\delta\geq\delta_3$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}\leq 0$. Notice $\delta_3-\delta_1=\frac{(d-p)(\beta(2r-d-p)+r-p)}{r-d}>0$, or $\delta_3>\delta_1$. We conclude that if $\delta\geq\delta_3$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}\leq 0$, and if $1<\delta<\delta_3$, $\frac{\partial s^*}{\partial x}\big|_{x=\underline{x}}>0$. Therefore, by continuity of $\frac{\partial s^*}{\partial x}$ in x, it follows that if $\delta\geq\delta_3$, s^* decreases in $x\in(\underline{x},\bar{x})$, and if $\delta<\delta_3$, there exists $\tilde{x}\in(\underline{x},\bar{x})$ such that s^* strictly increases in x on $(x,\tilde{x})$ and strictly decreases in x on $[\tilde{x},\bar{x})$. (iii). Note $\frac{\partial s^\star}{\partial\beta}=\frac{1}{(\delta-1)(x+1)}(((d-p)-(r-d)x)((\delta-1)(x+1)-(2\beta+1)((d-p)-(r-d)x)))$. Then if $\beta<\frac{(\delta-1)(x+1)}{2((d-p)-(r-d)x)}-\frac{1}{2}=\frac{\bar{\tau}(x)}{2}$, $\frac{\partial s^*}{\partial\beta}>0$, and if $\beta\geq\frac{\bar{\tau}(x)}{2}$, $\frac{\partial s^*}{\partial\beta}\leq 0$. Therefore, s^* increases in

$\beta \in (0, \frac{\bar{\tau}(x)}{2})$ and decreases in $\beta \in [\frac{\bar{\tau}(x)}{2}, \bar{\tau}(x))$. $\square$

Proof of Corollary 3. Note $e^* - s^* = \beta((d-p)-(r-d)x)\left(\frac{2(\beta+1)((d-p)-(r-d)x)}{(\delta-1)(x+1)} - 1\right)$. Because $x \in (\underline{x}, \bar{x})$, $(d-p)-(r-d)x > 0$. Therefore, $e^* \geq s^*$ if and only if $\frac{2(\beta+1)((d-p)-(r-d)x)}{(\delta-1)(x+1)} - 1 \geq 0$. It follows that $e^* \geq s^*$ if and only if $x \leq \bar{x} - \frac{(\delta-1)(r-p)}{(r-d)(2(\beta+1)(r-d)+\delta-1)}$. Defining $\hat{x} = \bar{x} - \frac{(\delta-1)(r-p)}{(r-d)(2(\beta+1)(r-d)+\delta-1)}$ yields the main result. Following a parallel argument to the proof of Corollary 1, define $\bar{\delta} = 2(\beta+1)(d-p)+1$, which satisfies $\bar{\delta} > 1$. It can then be shown that $s^* > e^*$ for all $x \in (\underline{x}, \bar{x})$ if and only if $\hat{x} \leq \underline{x}$, which holds if and only if $\delta \geq \bar{\delta}$. $\square$

Proof of Corollary 4. By plugging in the optimal solutions (e^*, s^*) in Proposition 3, the change in commuter welfare is given by

$$\Delta W\left(e^*, s^*, 0\right) = W\left(e^*, s^*, 0\right) - W^0 = -\frac{\beta^2((d-p)-(r-d)x)^2}{2(\delta-1)(x+1)} < 0.$$

Note $\frac{\partial \Delta W(e^*,s^*,0)}{\partial \beta} = -\frac{\beta((d-p)-x(r-d))^2}{(\delta-1)(x+1)} < 0$, and thus $\Delta W\left(e^*, s^*, 0\right)$ strictly decreases in β. Next, note $\frac{\partial \Delta W(e^*,s^*,0)}{\partial x} = \frac{\beta^2\left((r-p)^2-(x+1)^2(r-d)^2\right)}{2(\delta-1)(x+1)^2}$. Because $x \in (\underline{x}, \bar{x})$, $0 < (x+1)(r-d) < (r-p)$. It follows that $\frac{\partial \Delta W(e^*,s^*,0)}{\partial x} > 0$, and thus $\Delta W\left(e^*, s^*, 0\right)$ strictly increases in x. $\square$

Proof of Proposition 4. Based on the proof of Proposition 2, constraints (2.18b), (2.18d) and (2.18e), which are equivalent to constraints (2.21b), (2.21d) and (2.21e), uniquely define a feasible and thus optimal solution $z^* = \min\{\max\{\underline{z}, \hat{z}\}, \bar{z}\}$. Because $C(0,0,z^*) = 0$, it follows that z^* is also the unique feasible and thus optimal solution to formulation (2.21). To show that $W(0,0,z) > W^0$, note that because $z^* > 0$, using the definition in (2.12), we have $W(0,0,z) - W^0 = \frac{z(2(d-p)-2(r-d)x+z)}{2(\delta-1)(x+1)} > 0$. $\square$

Proof of Theorem 2. First, we compare the performance of mechanisms [D-A] and [I-A] in metrics I – V. By constraints (2.20c) and (2.21c), we have $C^I = C^D = 0$. Next, using the expressions for e^* and s^* given in Proposition 3, and the expression for z^* given in Proposi-

tion 4, it can be shown that $D_m(0,0,z^*) - D_m^0 \geq \beta D_m^0 = D_m(e^*,s^*,0) - D_m^0$, or equivalently, $D_m^I \geq D_m^D$. Because $D_m^I \geq D_m^D$, from (2.13) and (2.14) we immediately obtain $\Pi_p^I \geq \Pi_p^D$ and $\Pi_r^I \geq \Pi_r^D$, respectively. Recall from Corollary 4 and Proposition 4 that, $W(e^*,s^*,0) - W^0 < 0$ and $W(0,0,z^*) - W^0 > 0$, and thus $W^D < W^0 < W^I$. For metric VI, using definition in (2.15), we have $\Pi_s^I - \Pi_s^D = \frac{\hat{z}^2 + K((d-p)-x(r-d)) - (z^*-\tilde{z})^2}{(\delta-1)(x+1)} - \frac{K((d-p)-x(r-d))(\beta+1)}{(\delta-1)(x+1)}$, which is equal to the enterprise's profit difference between mechanisms [D] and [I]. Following the proof of Theorem 1, if $\beta \leq \tilde{\tau}(x)$, then $\Pi_s^I \geq \Pi_s^D$, and if $\beta \in (\tilde{\tau}(x), \tau(x)]$, then $\Pi_s^I < \Pi_s^D$. The result follows. $\square$

Before proving Proposition 7, we introduce two supporting results: Lemma 2 and Lemma 3. For ease of exposition, let $\gamma_1 = (\delta-1)(x+1)$ and $\gamma_2 = (d-p) - x(r-d)$, and note $\gamma_1 > 0$ and $\gamma_2 > 0$.

Lemma 2. *Suppose the constraint* $D_m(e,s,z) = \frac{\gamma_2+e+s+z}{\gamma_1}$, *or equivalently,* $e+s+z \leq \gamma_1 - \gamma_2$, *is added to formulation* (A.1). *Then formulation* (A.1) *is equivalent to*

$$\min_{e,s,z \geq 0} \ [C(e,s,z)]^+ \tag{A.26a}$$

$$s.t. \ z = \min\{\max\{z_1(e,s), z_2(e,s), z_3(e,s), 0\}, z_4(e,s), z_5(e,s)\} \tag{A.26b}$$

$$e \leq \frac{\gamma_1^2 - \gamma_2^2}{2\gamma_1} \tag{A.26c}$$

$$e + s \geq \frac{{\gamma_2}^2 + 2\gamma_1 e + \gamma_2 K}{\sqrt{\gamma_2^2 + 2\gamma_1 e}} - (\gamma_2 + K) \tag{A.26d}$$

$$e + s \leq \gamma_1 - \gamma_2 \tag{A.26e}$$

$$e + s \geq \frac{\beta((\beta+1)\gamma_2 - K)}{\beta+1}, \tag{A.26f}$$

where $z_1(e,s) = \frac{1}{2}(K - (\gamma_2 + (e+s)))$, $z_2(e,s) = \beta\gamma_2 - (e+s)$, $z_3(e,s) = \sqrt{\gamma_2^2 + 2\gamma_1 e} - (\gamma_2 + e + s)$, $z_4(e,s) = \gamma_1 - \gamma_2 - (e+s)$, *and* $z_5(e,s) = \sqrt{\left(\frac{K}{2} - \frac{1}{2}(\gamma_2 + (e+s))\right)^2 + K(e+s)} + \frac{K}{2} - \frac{1}{2}(\gamma_2 + (e+s))$.

Proof. Note that formulations (A.1) and (A.26) have the same objective functions. There-

fore, to establish the equivalence between the two formulations, it suffices to prove that the feasible regions are equivalent. First, we show that when adding an additional constraint $e + s + z \leq \gamma_1 - \gamma_2$, which is equivalent to $z \leq z_4(e, s)$, the feasible region of problem (A.1) is $\mathcal{P}_1 = \{(e, s, z) | \max\{z_2(e, s), z_3(e, s), 0\} \leq z \leq \min\{z_4(e, s), z_5(e, s)\}, z = \min\{\max\{z_1(e, s), z_2(e, s), \ z_3(e, s), 0\}, \ z_4(e, s), z_5(e, s)\}, e, s, z \geq 0\}$. Because we assume (e, s, z) satisfies $D_m(e, s, z) = \frac{\gamma_2+e+s+z}{\gamma_1}$, it follows that $D_m(e, s, z) - D_m^0 = \frac{e+s+z}{\gamma_1}$, and thus constraint (A.1b) is equivalent to $z \geq z_2(e, s)$. Using the definition of $W(e, s, z)$ in (2.12), we obtain that $W(e, s, z) - W^0 = \frac{(\gamma_2+e+s+z)^2-\gamma_2^2}{2\gamma_1} - e$, and thus constraint (A.1c) is equivalent to $z \geq z_3(e, s)$. Using the definition of $\Pi_s(e, s, z)$ in (2.15), we obtain that $\Pi_s(e, s, z) - \Pi_s^0 = \frac{(K-z)(e+s+z)}{\gamma_1} - \frac{\gamma_2 z}{\gamma_1}$, and thus constraint (A.1d) is equivalent to $z_5'(e, s) \leq z \leq z_5(e, s)$, where $z_5'(e, s) = -\sqrt{\left(\frac{K}{2} - \frac{1}{2}(\gamma_2 + (e + s))\right)^2 + K(e + s)} + \frac{K}{2} - \frac{1}{2}(\gamma_2 + (e + s)) \leq 0$. Thus, constraints (A.1b), (A.1c) and (A.1d) along with constraints $z \geq 0$ and $e + s + z \leq \gamma_1 - \gamma_2$ are equivalent to

$$z \in [\max\{z_2(e, s), z_3(e, s), 0\}, \min\{z_4(e, s), z_5(e, s)\}]. \tag{A.27}$$

For constraint (A.1e), note that when fixing e and s, $\Pi_s(e, s, z) = (K - z)\frac{e+s+z+\gamma_2}{\gamma_1}$ is concave in z. By solving the first order condition $\frac{\partial \Pi_s(e,s,z)}{\partial z} = 0$, and denoting the solution as $z_1(e, s)$, we obtain $z_1(e, s) = \frac{1}{2}(K - (\gamma_2 + (e + s)))$. It is straightforward to show that $z_1(e, s) \leq z_5(e, s)$. Then for any (e, s, z) that satisfies (A.27), constraint (A.1e) is equivalent to $z = \min\{\max\{z_1(e, s), z_2(e, s), z_3(e, s), 0\}, z_4(e, s), z_5(e, s)\}$. Therefore, if $e+s+z \leq \gamma_1-\gamma_2$, the feasible region of (A.1) is $\mathcal{P}_1 = \{(e, s, z) | \max\{z_2(e, s), z_3(e, s), 0\} \leq z \leq \min\{z_4(e, s), z_5(e, s)\}, z = \min\{\max\{z_1(e, s), z_2(e, s), z_3(e, s), 0\}, z_4(e, s), z_5(e, s)\}, e, s, z \geq 0\}$.

Next, from the definitions of $z_i(e, s), i \in \{1, 2, \ldots, 5\}$ in Lemma 2, it is straightforward to show that the feasible region of problem (A.26) is $\mathcal{P}_2 = \{(e, s, z) | \max\{z_2(e, s), z_3(e, s), 0\} \leq \min\{z_4(e, s), z_5(e, s)\}, z = \min\{\max\{z_1(e, s), z_2(e, s), z_3(e, s), 0\}, z_4(e, s), z_5(e, s)\}, e, s, z \geq 0\}$.

It's plain to show that for any $(e', s', z') \in \mathcal{P}_1$, it holds that $(e', s', z') \in \mathcal{P}_2$, and for

any $(e'', s'', z'') \in \mathcal{P}_2$, we have $(e'', s'', z'') \in \mathcal{P}_1$. It follows that $\mathcal{P}_1$ is equivalent to $\mathcal{P}_2$. The equivalence between the two problems then follows. $\square$

For use in Lemma 3 below, let $\breve{x} = \frac{(d-p)-(\delta-1)+2K}{(r-d)+(\delta-1)}$, $\bar{\beta}(x) = \frac{2K}{\gamma_2}$, and $\beta_1(x) = \frac{2\gamma_1 - 3\gamma_2 - \sqrt{(2\gamma_1-\gamma_2)^2 - 8\gamma_1 K}}{2\gamma_2}$.

Lemma 3. *Suppose $K < \gamma_1$ and $\beta \in (\tau(x), \bar{\tau}(x))$. Further, suppose the constraint $D_m(e,s,z) = \frac{\gamma_2+e+s+z}{\gamma_1}$, or equivalently, $e+s+z \leq \gamma_1-\gamma_2$, is added to formulation* (A.1)*. Then the following statements hold:*

(i) If $x \leq \breve{x}$, then $e^* > 0$, $s^* \geq 0$, and $z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = 0$.

(ii) If $x > \breve{x}$, and

(a) if $\tau(x) < \beta \leq \beta_1(x)$, then $e^* > 0, s^* \geq 0$ and $z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = 0$.

(b) if $\beta_1(x) < \beta \leq \bar{\beta}(x)$, then $e^* > 0, s^* > 0$, and $z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = 0$.

(c) if $\bar{\beta}(x) < \beta < \bar{\tau}(x)$, then $e^* > 0, s^* > 0$, and $z^* > 0$ and $[C(e^*, s^*, z^*)]^+ > 0$.

Further, the optimal objective value is $[C(e^*, s^*, z^*)]^+ = \max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1}\right\}$.

Proof. Note that $\bar{\tau}(x) = \frac{\gamma_1}{\gamma_2} - 1$ and $\tau(x) = \min\left\{\frac{K}{\gamma_2} - 1, \frac{\gamma_1}{\gamma_2} - 1\right\}$. Therefore, $K < \gamma_1$ implies that $(\tau(x), \bar{\tau}(x))$ is non-empty, and that $\tau(x) = \frac{K}{\gamma_2} - 1$. Note that in Lemma 2, we showed that when imposing the additional constraint $e + s + z \leq \gamma_1 - \gamma_2$, solving problem (A.1) is equivalent to solving problem (A.26); therefore, we will identify the optimal solution (e^*, s^*, z^*) to (A.1) by solving (A.26). The proof proceeds in three steps. In Step 1, we show that any feasible z to problem (A.26) must satisfy one of the following five equations: $z = z_1(e,s), z = z_2(e,s), z = 0, z = z_3(e,s), z = z_4(e,s)$. In Step 2, we solve for the optimal solutions to problem (A.26) by imposing each of these five equations as an additional constraint. In Step 3, we identify the optimal solution (e^*, s^*, z^*) in formulation (A.26) by comparing the optimal objective function values for each of these five subproblems.

Step 1. To solve formulation (A.26), let us first consider constraint (A.26b), which requires $z = \breve{z}(e,s)$, where $\breve{z}(e,s) = \min\{\max\{z_1(e,s), z_2(e,s), z_3(e,s), 0\}, z_4(e,s), z_5(e,s)\}$ for simplicity. Note that $\breve{z}(e,s) \neq z_5(e,s)$. To see why this holds, note that from the proof of Lemma 2, $z_1(e,s) < z_5(e,s)$; in addition, constraints (A.26c) – (A.26f) are equivalent to $\max\{z_2(e,s), z_3(e,s), 0\} \leq \min\{z_4(e,s), z_5(e,s)\}$. Therefore, $\breve{z}(e,s)$ must satisfy one of the following five conditions: $\breve{z}(e,s) = z_1(e,s)$, $\breve{z}(e,s) = z_2(e,s)$, $\breve{z}(e,s) = 0$, $\breve{z}(e,s) = z_3(e,s)$, or $\breve{z}(e,s) = z_4(e,s)$.

Step 2. We include each one of the five conditions listed above as additional constraint in problem (A.26), and solve the corresponding subproblem.

Case 1: Suppose $\breve{z}(e,s) = z_1(e,s)$. We show this case is infeasible. Specifically, if $\breve{z}(e,s) = z_1(e,s)$, then

$$z_1(e,s) \geq z_2(e,s) \text{ if and only if } e+s \geq 2(\beta+1)\gamma_2 - (\gamma_2 + K), \tag{A.28a}$$

$$z_1(e,s) \geq z_3(e,s) \text{ if and only if } e+s \geq 2\sqrt{\gamma_2^2 + 2\gamma_1 e} - (\gamma_2 + K), \tag{A.28b}$$

$$z_1(e,s) \geq 0 \text{ if and only if } e+s \leq K - \gamma_2, \tag{A.28c}$$

$$z_1(e,s) \leq z_4(e,s) \text{ if and only if } e+s \leq 2\gamma_1 - \gamma_2 - K, \tag{A.28d}$$

and it is trivial to prove $z_1(e,s) \leq z_5(e,s)$. Define $y_1 \equiv \frac{\gamma_2^2 + 2\gamma_1 e + \gamma_2 K}{\sqrt{{\gamma_2}^2 + 2\gamma_1 e}} - (\gamma_2 + K), y_2 \equiv 2(\beta+1)\gamma_2 - (\gamma_2 + K), y_3 \equiv 2\sqrt{{\gamma_2}^2 + 2\gamma_1 e} - (\gamma_2 + K)$ and $y_4 \equiv \frac{\beta((\beta+1)\gamma_2 - K)}{\beta+1}$. Combining constraints (A.28) with problem (A.26), then a feasible (e,s) must satisfy

$$e \leq \frac{\gamma_1^2 - \gamma_2^2}{2\gamma_1}, \max\{e, y_1, y_2, y_3\} \leq e+s \leq K - \gamma_2. \tag{A.29}$$

Because $\beta > \frac{K}{\gamma_2} - 1$, it follows that $\max\{e, y_1, y_2, y_3\} \geq y_2 = 2(\beta+1)\gamma_2 - (\gamma_2 + K) > K - \gamma_2$, which contradicts (A.29). Therefore, a feasible solution(e,s,z) to problem (A.1) never satisfies $z = \breve{z}(e,s) = z_1(e,s)$.

Case 2: Suppose $\breve{z}(e,s) = z_2(e,s)$. Similar to Case 1, the condition implies $s+e \leq 2\beta\gamma_2 + \gamma_2 - K$, $e \leq \frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1}$ and $s+e \leq \beta\gamma_2$. Note that $e \leq \frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1}$ implies $\max\{y_1, y_4\} = y_4 < \beta\gamma_2$,

where y_1 and y_4 are defined in Case 1. By plugging $z = z_2(e,s)$ into the objective, we obtain the cost function as $C(e,s,z_2(e,s)) = \frac{(\beta+1)\gamma_2}{\gamma_1}(e+s)-e$. Then given the additional constraint $\breve{z}(e,s) = z_2(e,s)$, problem (A.26) is equivalent to the following linear program:

$$\min_{y\geq e\geq 0} \left[\frac{(\beta+1)\gamma_2}{\gamma_1}(e+s)-e\right]^+ \tag{A.30a}$$

$$\text{s.t. } e \leq \frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1} \tag{A.30b}$$

$$\max\{y_4, e\} \leq e+s \leq \beta\gamma_2, \tag{A.30c}$$

By solving the above linear program, it follows that if $\frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1} \geq y_4$, the optimal solution (e^*,s^*,z^*) must satisfy $e^*>0$, $s^*\geq 0$, $z^*>0$ and $\left[C\left(e^*,s^*,z^*\right)\right]^+ = 0$; if $\frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1} < y_4$ and $\beta \leq \frac{2K}{\gamma_2}$, then $e^*>0$, $s^*>0$, $z^*>0$ and $\left[C\left(e^*,s^*,z^*\right)\right]^+ = 0$; if $\frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1} < y_4$ and $\beta > \frac{2K}{\gamma_2}$, then $e^*>0$,$s^*>0$,$z^*>0$ and $\left[C\left(e^*,s^*,z^*\right)\right]^+ = \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1} > 0$. Note that $\frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1} < y_4$ if and only if $(\gamma_1,\beta) \in \left\{\gamma_1,\beta \mid \frac{\gamma_2}{2}+K+\sqrt{K\left(\gamma_2+K\right)} < \gamma_1 \leq 2K+\gamma_2, \beta\in(\beta_1,\beta_2)\right\} \cup \{\gamma_1,\beta \mid \gamma_1 > 2K+\gamma_2, \beta\in(\beta_1,\bar{\tau}(x))\}$, where $\beta_2 = \frac{2\gamma_1-3\gamma_2+\sqrt{(2\gamma_1-\gamma_2)^2-8\gamma_1 K}}{2\gamma_2}$.

Case 3: Suppose $\breve{z}(e,s) = z_3(e,s)$. Using a similar argument to Case 2, by letting $Y = (e+s)\sqrt{\gamma_2^2+2\gamma_1 e}$ and $t = \sqrt{\gamma_2^2+2\gamma_1 e}$, and rewriting the cost function as as $C_3(Y,t) = C(e,s,z_3(e,s)) = \frac{2Y-t^2+\gamma_2^2}{2\gamma_1}$, we can show that with the additional constraint $\breve{z}(e,s) = z_3(e,s)$, problem (A.26) is equivalent to the following:

$$\min_{Y,t\geq 0} \left[C_3(Y,t)\right]^+ \tag{A.31a}$$

$$\text{s.t. } (\beta+1)\gamma_2 \leq t \leq \gamma_1 \tag{A.31b}$$

$$\max\left\{t^2-(\gamma_2+K)\,t+K\gamma_2, \frac{t\left(t^2-\gamma_2^2\right)}{2\gamma_1}\right\} \leq Y \leq t^2-\gamma_2 t \tag{A.31c}$$

Let $Y_1(t) = t^2-(\gamma_2+K)\,t+K\gamma_2$ and $Y_2(t) = \frac{t\left(t^2-\gamma_2^2\right)}{2\gamma_1}$ for simplicity. Then constraint (A.31c) becomes $\{Y_1(t),Y_2(t)\} \leq Y \leq t^2-\gamma_2 t$. Next, we consider two cases according to constraint (A.31c): $Y_1(t) \leq Y_2(t)$ and $Y_1(t) > Y_2(t)$. In preparation, let $\mathcal{B}_1 = \{\gamma_1,\beta \mid \gamma_1 \leq 2K+\gamma_2\} \cup \left\{\gamma_1,\beta \mid \gamma_1 > 2K+\gamma_2, \beta\in\left(\frac{K}{\gamma_2}-1,\beta_1\right]\right\}$ and $\mathcal{B}_2 = \{\gamma_1,\beta \mid \gamma_1 > 2K+\gamma_2, \beta\in(\beta_1,\bar{\tau}(x))\}$; then

we have if $(\gamma_1, \beta) \in \mathcal{B}_1$, there exist a t satisfying constraint (A.31b) such that $Y_1(t) \leq Y_2(t)$; if $(\gamma_1, \beta) \in \mathcal{B}_2$, for all t satisfying constraint (A.31b), it holds that $Y_1(t) > Y_2(t)$.

First, suppose $Y_1(t) \leq Y_2(t)$. We only focus on the regime where $(\gamma_1, \beta) \in \mathcal{B}_1$. It is straightforward to verify that in this case, $e^* \geq 0$, $s^* \geq 0$, $z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = 0$. Next, when restricting $Y_1(t) > Y_2(t)$, note that we have already proved that when $(\gamma_1, \beta) \in \mathcal{B}_1$, (e^*, s^*, z^*) satisfy $e^* > 0$, $s^* \geq 0$ and $z^* > 0$, so we only need to consider the regime where $(\gamma_1, \beta) \in \mathcal{B}_2$. In this case, it is straightforward to show that if $2K + \gamma_2 \geq (\beta + 1)\gamma_2$, or equivalently $\beta \leq \frac{2K}{\gamma_2}$, then $e^*, s^*, z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = 0$; if $2K + \gamma_2 < (\beta + 1)\gamma_2$, or equivalently $\beta > \frac{2K}{\gamma_2}$, then $e^*, s^*, z^* > 0$ and $[C(e^*, s^*, z^*)]^+ = \frac{\beta\gamma_2(\beta\gamma_2 - 2K)}{2\gamma_1} > 0$.

Case 4: Suppose $\breve{z}(e, s) = 0$. We show that this case is dominated by Case 3. Specifically, when restricting $\breve{z}(e, s) = 0$, it is easy to show that problem (A.26) reduces to (2.16). Thus the minimal cost in this case must be no less than that of mechanism [D], namely, $C(e^*, s^*, z^*) \geq \frac{\beta^2\gamma_2^2}{2\gamma_1} > 0$. Comparing the minimal objective value in this case with that in Case 3, which is given by $\max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2 - 2K)}{2\gamma_1}\right\}$, we can show that $[C(e^*, s^*, z^*)]^+ = C(e^*, s^*, z^*) \geq \frac{\beta^2\gamma_2^2}{2\gamma_1} > \max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2 - 2K)}{2\gamma_1}\right\}$, which implies this case is dominated by Case 3.

Case 5: Suppose $\breve{z}(e, s) = z_4(e, s)$. We show this case is infeasible. Specifically, assuming $\breve{z}(e, s) = z_4(e, s)$, this is equivalent to

$$z_1(e, s) \geq z_2(e, s) \text{ if and only if } e + s \geq y_2 = 2(\beta + 1)\gamma_2 - (\gamma_2 + K), \tag{A.32a}$$

$$z_1(e, s) \geq z_3(e, s) \text{ if and only if } e + s \geq y_3 = 2\sqrt{\gamma_2^2 + 2\gamma_1 e} - (\gamma_2 + K), \tag{A.32b}$$

$$z_1(e, s) \geq 0 \text{ if and only if } e + s \leq K - \gamma_2, \tag{A.32c}$$

$$z_1(e, s) \geq z_4(e, s) \text{ if and only if } e + s \geq 2\gamma_1 - \gamma_2 - K, \tag{A.32d}$$

and $\max\{z_2(e, s), z_3(e, s), 0\} \leq \min\{z_4(e, s), z_5(e, s)\}$. Combining with constraints (A.26b)–(A.26f), because $\gamma_1 > K$ and thus $2\gamma_1 - \gamma_2 - K > \gamma_1 - \gamma_2$, inequality (A.32d) contradicts constraint (A.26e), which requires $e + s \leq \gamma_1 - \gamma_2$. Therefore, $z = \breve{z}(e, s) = z_4(e, s)$ cannot hold.

Step 3. We have established in Step 2 that the optimal private subsidy z must be equal to either $z_2(e,s)$ or $z_3(e,s)$. Combining Case 2 and Case 3, we conclude that if $(\gamma_1,\beta) \in \{\gamma_1,\beta \mid \gamma_1 \leq 2K+\gamma_2\} \cup \left\{\gamma_1,\beta \mid \gamma_1 > 2K+\gamma_2, \beta \in \left(\frac{K}{\gamma_2}-1, \beta_1\right]\right\}$, the optimal solution (e^*,s^*,z^*) must satisfy $e^*>0$, $s^* \geq 0$ and $z^*>0$; if $(\gamma_1,\beta) \in \{\gamma_1,\beta \mid \gamma_1 > 2K+\gamma_2, \ \ \beta \in (\beta_1, \bar{\tau}(x))\}$, the optimal solution (e^*,s^*,z^*) must satisfy $e^*,s^*,z^*>0$. Moreover, $[C(e^*,s^*,z^*)]^+ = \max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1}\right\}$. Note that by plugging in $\gamma_1=(\delta-1)(x+1)$ and $\gamma_2=(d-p)-x(r-d)$, $\gamma_1 \leq 2K+\gamma_2$ is equivalent to $x \leq \breve{x}$, where $\breve{x}=\frac{(d-p)-(\delta-1)+2K}{(r-d)+(\delta-1)}$. The result follows. $\square$

Proof of Proposition 7. Similar to the proofs of Propositions 1 and 2, the proof is composed of two cases: $D_m(e,s,z)=\frac{\gamma_2+e+s+z}{\gamma_1}$, or equivalently, $e+s+z \leq \gamma_1-\gamma_2$; and $D_m(e,s,z)=1$, or equivalently, $e+s+z \geq \gamma_1-\gamma_2$.

Case 1: Suppose (e,s,z) is restricted to the regime where $D_m(e,s,z)=\frac{\gamma_2+e+s+z}{\gamma_1}$, which is equivalent to $e+s+z \leq \gamma_1-\gamma_2$. We divide the analysis into two further cases with respect to β: $\beta \in (0,\tau(x)]$ and $\beta \in (\tau(x),\bar{\tau}(x))$.

0.6cm Case 1.1: Suppose $\beta \in (0,\tau(x)]$. Because $[C(e,s,z)]^+ \geq 0$, if a feasible solution (e,s,z) satisfies $C(e,s,z) \leq 0$, it must be optimal. Note that the minimal cost of mechanism [I] is 0, so the optimal solution to problem (2.18) under mechanism [I] is also optimal to (A.1) in the case when $\beta \in (0,\tau(x)]$. Namely, there exists an optimal solution to (A.1) $(e^*,s^*,z^*)=(0,0,\min\{\max\{\underline{z},\hat{z}\},\bar{z}\})$, where $\underline{z},\hat{z},\bar{z}$ are as defined in Proposition 2. It can also be shown that $(e^*,s^*,z^*) \in \mathcal{Q}$ is optimal, where $\mathcal{Q}$ is defined as $\mathcal{Q}=\left\{(e^*,s^*,z^*)|e^*=\frac{\beta(\beta+2)\gamma_2^2}{2\gamma_1}, \ \ s^* \in \left[0, \frac{\beta(\beta+2)\gamma_2(\gamma_1-(\beta+1)\gamma_2)}{2(\beta+1)\gamma_1}\right], z^*=z_2(e^*,s^*)\right\}$, and the objective value is $[C(e^*,s^*,z^*)]^+=0$. Note that the optimal private subsidy must be strictly positive, $z^*>0$; otherwise, if $z^*=0$, the optimal cost should be no less than that of mechanism [D] as presented in Corollary 2 and is thus strictly positive. To summarize, when $\beta \in (0,\tau(x)]$, the optimal solutions (e^*,s^*,z^*) must satisfy $e^*,s^* \geq 0$ and $z^*>0$ and $[C(e^*,s^*,z^*)]^+=0$.

Case 1.2: When $\beta \in (\tau(x),\bar{\tau}(x))$, the optimal solutions have been characterized in

Lemma 3.

To summarize, when requiring $D_m(e,s,z) = \frac{\gamma_2+e+s+z}{\gamma_1}$, the optimal objective value is $[C(e^*,s^*,z^*)]^+ = \max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1}\right\}$ and the optimal solution (e^*,s^*,z^*) is characterized by the following:

(i) Suppose $x \le \breve{x}$. Then $[C(e^*,s^*,z^*)]^+ = 0$. Further,

(a) if $0 \le \beta \le \tau(x)$, then $e^* \ge 0, s^* \ge 0$ and $z^* > 0$;

(b) if $\tau(x) < \beta < \bar{\tau}(x)$, then $e^* > 0, s^* \ge 0$, and $z^* > 0$.

(ii) Suppose $x > \breve{x}$. Then

(a) if $0 \le \beta \le \bar{\beta}(x)$, $[C(e^*,s^*,z^*)]^+ = 0$. Further,

i. if $0 \le \beta \le \tau(x)$, then $e^* \ge 0, s^* \ge 0$, and $z^* > 0$;

ii. if $\tau(x) < \beta \le \beta_1(x)$, then $e^* > 0, s^* \ge 0$ and $z^* > 0$;

iii. if $\beta_1(x) < \beta \le \bar{\beta}(x)$, then $e^* > 0, s^* > 0$, and $z^* > 0$

(b) if $\bar{\beta}(x) < \beta < \bar{\tau}(x)$, $[C(e^*,s^*,z^*)]^+ = \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1} > 0$. Further, $e^* > 0, s^* > 0$, and $z^* > 0$.

Case 2: Suppose $D_m(e,s,z) = 1$, which is equivalent to $e+s+z \ge \gamma_1 - \gamma_2$. We show this case is never optimal. Using the definition in (2.15), in this case, $\Pi_s(e,s,z) = (K-z)\cdot D_m(e,s,z) = K - z$. Then from constraint (A.1e), a feasible private subsidy z must satisfy $z = \arg\max_z \Pi_s(e,s,z) = \arg\max_z (K-z) = 0$. Using the definition in (2.12), constraint (A.1c) can then be rewritten as

$$W(e,s,z) - W^0 = s + z - \frac{(\gamma_1-\gamma_2)^2}{2\gamma_1} = s - \frac{(\gamma_1-\gamma_2)^2}{2\gamma_1} \ge 0, \tag{A.33}$$

which implies feasible s must satisfy $s \ge \frac{(\gamma_1-\gamma_2)^2}{2\gamma_1}$. Recall from the proof of Proposition 1, when $D_m(e,s,z) = 1$, $C(e,s,z) = s \ge \frac{(\gamma_1-\gamma_2)^2}{2\gamma_1} > \frac{\beta^2\gamma_2^2}{2\gamma_1}$, where $\frac{\beta^2\gamma_2^2}{2\gamma_1}$ is the minimal cost under mechanism [D]. Further, because the optimal objective value in Case 1 is $\max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1}\right\}$ and it is straightforward to verify that the $\frac{\beta^2\gamma_2^2}{2\gamma_1} > \max\left\{0, \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1}\right\}$,

it follows the optimal solution (e^*, s^*, z^*) must satisfy $D_m(e, s, z) = \frac{\gamma_2+e+s+z}{\gamma_1}$. Hence $D_m(e^*, s^*, z^*) = 1$ cannot hold at optimality. $\square$

Proof of Corollary 5. Using the optimal solutions to mechanism [G] presented in the proof of Proposition 7, it follows the optimal objective value under mechanism [G] is

$$[C(e^*, s^*, z^*)]^+ = \begin{cases} 0 & \text{if } \beta \leq \bar{\beta}(x), \\ \frac{\beta\gamma_2(\beta\gamma_2-2K)}{2\gamma_1} & \text{if } \beta > \bar{\beta}(x), \end{cases}$$

where $\bar{\beta}(x) = \frac{2K}{\gamma_2}$ and γ_1, γ_2 are defined as in Proposition 7. Using the definition above, it is straightforward to show that $[C(e^*, s^*, z^*)]^+$ increases in β, and is weakly less than the minimal costs of both mechanisms [D] and [I]. $\square$

Proof of Proposition 8. The proof consists of two steps. In Step 1, we establish an equivalent formulation to problem (A.2). In Step 2, we solve for the optimal solutions to the equivalent problem and problem (A.2).

Step 1. We first prove problem (A.2) is equivalent to the following problem:

$$\max_{e,s\geq 0} \ \min\{e + s, \gamma_1 - \gamma_2\} \tag{A.34a}$$

$$\text{s.t. } C(e, s, 0) \leq B \tag{A.34b}$$

$$W(e, s, 0) \geq W^0. \tag{A.34c}$$

To show the equivalence between the two problems, first note that from the definition of $D_m(e, s, z)$ in (2.10), we have $D_m(e, s, 0) - D_m^0 = \frac{\min\{e+s,\gamma_1-\gamma_2\}}{\gamma_1}$, where $\gamma_1 = (\delta - 1)(x + 1)$ and $\gamma_2 = (d - p) - x(r - d)$. Because $\delta > 1$ and $x > 0$, maximizing $D_m(e, s, 0) - D_m^0$ is equivalent to maximizing $\min\{e + s, \gamma_1 - \gamma_2\}$. Further, from the definition of $\Pi_s(e, s, z)$ in (2.15), $\Pi_s(e, s, z) - \Pi_s^0 = (K - z)\frac{\min\{e+s,\gamma_1-\gamma_2\}}{\gamma_1}$. Because $\underline{x} < x < \bar{x}$, then $\gamma_1 - \gamma_2 > 0$.

Therefore, $e, s \geq 0$ implies $\Pi_s(e, s, z) \geq \Pi_s^0$, which allows us to omit the corresponding constraint in (A.2) without changing the feasible region. Therefore, problem (A.2) is equivalent to problem (A.34).

Step 2. To solve problem (A.34), we change variables by letting $y = e + s$. Then using the definitions of $C(e, s, 0)$ in (2.11) and $W(e, s, 0)$ in (2.12), formulation (A.34) can be rewritten as

$$\max_{0 \leq s \leq y} \ \min\{y, \gamma_1 - \gamma_2\} \tag{A.35a}$$

$$\text{s.t. } s \leq B + \frac{\left(\frac{\gamma_1 - \gamma_2}{2}\right)^2 - \left(\min\{y, \gamma_1 - \gamma_2\} - \frac{\gamma_1 - \gamma_2}{2}\right)^2}{\gamma_1} \tag{A.35b}$$

$$s \geq \frac{(\gamma_1 - \gamma_2)^2 - (\min\{y, \gamma_1 - \gamma_2\} - (\gamma_1 - \gamma_2))^2}{2\gamma_1}. \tag{A.35c}$$

Let us relax the non-negativity constraint $0 \leq s \leq y$ for now. Note that a feasible y must satisfy

$$\frac{(\gamma_1 - \gamma_2)^2 - (\min\{y, \gamma_1 - \gamma_2\} - (\gamma_1 - \gamma_2))^2}{2\gamma_1} \leq B + \frac{\left(\frac{\gamma_1 - \gamma_2}{2}\right)^2 - \left(\min\{y, \gamma_1 - \gamma_2\} - \frac{\gamma_1 - \gamma_2}{2}\right)^2}{\gamma_1}, \tag{A.36}$$

or equivalently, $B \geq \frac{(\min\{y, \gamma_1 - \gamma_2\})^2}{2\gamma_1}$. Define $\tilde{B} = \frac{(\gamma_1 - \gamma_2)^2}{2\gamma_1}$. We then decompose the analysis in two cases: (i) when $B \leq \tilde{B}$ and (ii) when $B > \tilde{B}$, which corresponds to statements (i) and (ii) in the proposition.

1. If $B \leq \tilde{B}$, a feasible y must satisfy $\frac{(\min\{y, \gamma_1 - \gamma_2\})^2}{2\gamma_1} \leq \tilde{B} = \frac{(\gamma_1 - \gamma_2)^2}{2\gamma_1}$, which implies $y \leq \gamma_1 - \gamma_2$, and in this case, (A.36) is equivalent to $y^2 \leq 2\gamma_1 B$. Because $\min\{y, \gamma_1 - \gamma_2\} = y$, then the unique optimal solution is $y^* = \sqrt{2\gamma_1 B} \leq \gamma_1 - \gamma_2$ and $s^* = B\left(\frac{\sqrt{2}(\gamma_1 - \gamma_2)}{\sqrt{\gamma_1 B}} - 1\right)$. It follows that $e^* = y^* - s^* = \frac{\sqrt{2B}\gamma_2}{\sqrt{\gamma_1}} + B$ and $D_m(e^*, s^*, 0) = \frac{\sqrt{2B\gamma_1} + \gamma_2}{\gamma_1}$. Because $y^* \geq s^* \geq 0$, then (y^*, s^*) is optimal to problem (A.35) and thus (e^*, s^*) is optimal to both problems (A.2) and (A.34).

2. Suppose $B > \tilde{B}$. We consider two subproblems, that follow from imposing the additional constraints $y \geq \gamma_1 - \gamma_2$ and $y < \gamma_1 - \gamma_2$. When $y \geq \gamma_1 - \gamma_2$, then the optimal solution (s^*, y^*) is such that $s^* \in [\tilde{B}, B]$ and $y^* \in [\gamma_1 - \gamma_2, \infty)$. The optimal objective value is $\gamma_1 - \gamma_2$. In this case, it must hold that $e^* \geq \gamma_1 - \gamma_2 - s^*$ and $D(e^*, s^*, 0) = 1$. However, when $y < \gamma_1 - \gamma_2$, the optimal objective value is strictly less than $\gamma_1 - \gamma_2$, which is dominated by the case when $y \geq \gamma_1 - \gamma_2$. Because $0 \leq s^* \leq y^*$, (y^*, s^*) is optimal to problem (A.35).

The result follows by plugging in $\gamma_1 = (\delta - 1)(x + 1)$ and $\gamma_2 = (d - p) - x(r - d)$. $\square$

Proof of Proposition 9. Similar to the proof of Proposition 4, it is straightforward to show that constraints (A.4d) and (A.4e) uniquely determines a solution $z^* = \min\{\hat{z}, (\delta - 1)(x + 1) - ((d - p) - x$ where $\hat{z} = \frac{1}{2}(K - ((d-p) - x(r-d)))$. Using (2.11) and (2.12), we find z^* satisfies constraints (A.4b) and (A.4c). Therefore, z^* is the unique feasible and thus optimal solution to problem (A.4). $\square$

Proof of Proposition 10. Similar to the proof of Proposition 1, to solve for the optimal solution (e^*, s^*), we consider two cases using the definition of $D_m(e, s, 0)$ in (A.6), which implies $D_m(e, s, 0) = \min\left\{\sigma, \frac{(d-p)-(\delta-1)V+e+s}{(r-d)+(\delta-1)V}\right\}$: $D_m(e, s, 0) = \frac{(d-p)-(\delta-1)V+e+s}{(r-d)+(\delta-1)V}$ and $D_m(e, s, 0) = \sigma$.

Case 1: Suppose we add $D_m(e, s, 0) = \frac{(d-p)-(\delta-1)V+e+s}{(r-d)+(\delta-1)V}$ as an additional constraint. Using a similar argument to the proof of Lemma 1, it can be shown that

$$[C(e, s, 0)]^+ \geq C(e, s, 0) \geq \frac{\beta^2((d - p) - (\delta - 1)V)^2}{2((r - d) + (\delta - 1)V)},$$

which holds with equality if and only if both constraints (A.8b) and (A.8c) are binding. To find the optimal solution, it suffices to construct a feasible solution (e_1, s_1) that attains the

lower bound, where (e_1, s_1) is the unique solution to the equations

$$D_m(e, s, 0) - D_m^0 = \beta D_m^0$$
$$W(e, s, 0) = W^0,$$

which correspond to constraints (A.8b) and (A.8c). Then (e_1, s_1) is given by:

$$\begin{aligned} e_1 &= \frac{\beta(\beta+2)((d-p)-(\delta-1)V)^2}{2\sigma((r-d)+(\delta-1)V)}, \\ s_1 &= \beta((d-p)-(\delta-1)V) - \frac{\beta(\beta+2)((d-p)-(\delta-1)V)^2}{2\sigma((r-d)+(\delta-1)V)}. \end{aligned}$$

Using a similar argument to the proof of Proposition 1, it can be verified that (e_1, s_1) satisfies $D_m(e, s, 0) = \frac{(d-p)-(\delta-1)V+e+s}{(r-d)+(\delta-1)V}$ and is feasible to (A.8). Therefore, (e_1, s_1) is the unique optimal solution to problem (A.8) when requiring $D_m(e, s, 0) = \frac{(d-p)-(\delta-1)V+e+s}{(r-d)+(\delta-1)V}$.

Case 2: Suppose $D_m(e, s, 0) = \sigma$. It follows that $D_d(e, s, 0) = 0$ and the operational cost is $C(e, s, 0) = s\sigma$. In this case, constraint (A.8c) can be rewritten as

$$W(e, s, 0) - W^0 = s\sigma - \frac{((d-p) - \sigma((r-d)+(\delta-1)V) - (\delta-1)V)^2}{2((r-d)+(\delta-1)V)} \geq 0. \qquad \text{(A.37)}$$

Let $\phi \equiv \frac{((d-p)-\sigma((r-d)+(\delta-1)V)-(\delta-1)V)^2}{2((r-d)+(\delta-1)V)}$. Then constraint (A.8c) implies $s\sigma \geq \phi \geq 0$ and thus that $[C(e, s, 0)]^+ = C(e, s, 0) = s\sigma \geq \phi$. It is straightforward to verify that $\phi > C(e_1, s_1, 0) = [C(e_1, s_1, 0)]^+$, where (e_1, s_1) is the optimal solution in Case 1, and it follows that $[C(e, s, 0)]^+ \geq \phi > [C(e_1, s_1, 0)]^+$. Therefore, the optimal solution to (A.8) is $(e^*, s^*) = (e_1, s_1)$, and the resulting minimal cost is strictly positive, $C(e_1, s_1, 0) = \frac{\beta^2((d-p)-(\delta-1)V)^2}{2((r-d)+(\delta-1)V)}$. □

Proof of Proposition 11. Using a parallel argument to the proof of Proposition 2, we first note that $[C(0, 0, z)]^+ = C(0, 0, z) = 0$. As such, any feasible solution to problem (A.8) is optimal. Therefore, to find the feasible and thus optimal solution to problem (A.8), it

suffices to find the optimal solution z^* to the following problem:

$$\max_{z\geq 0} \Pi_s(0,0,z) \tag{A.38a}$$

$$\text{s.t. } D_m(0,0,z) - D_m^0 \geq \beta D_m^0, \tag{A.38b}$$

$$W(0,0,z) \geq W^0, \tag{A.38c}$$

$$\Pi_s(0,0,z) \geq \Pi_s^0. \tag{A.38d}$$

Using a similar argument in the proof of Proposition 2, it is straightforward to show that the feasible region of (A.38) is $[\underline{z},\bar{z}]\cup(\tilde{z},\dot{z}]$, where $\underline{z} = \beta((d-p)-(\delta-1)V)$, $\bar{z} = \min\{K, \sigma((r-d)+(\delta-1)V)\} - ((d-p)-(\delta-1)V)$, $\tilde{z} = \sigma((r-d)+(\delta-1)V) - ((d-p)-(\delta-1)V)$ and $\dot{z} = K\left(1-\frac{(d-p)-(\delta-1)V}{\sigma((r-d)+(\delta-1)V)}\right)$. Note (A.38) and (A.10) are feasible if and only if $[\underline{z},\bar{z}]\cup(\tilde{z},\dot{z}]$ is non-empty, or equivalently, $\beta \leq \tau(x)$, where $\tau(x) = \min\left\{\frac{K}{(d-p)-(\delta-1)V} - 1, \bar{\tau}(x)\right\}$ and $\bar{\tau}(x) = \frac{\sigma((r-d)+(\delta-1)V)}{(d-p)-(\delta-1)V} - 1$. We can now solve (A.38) assuming $\beta \leq \tau(x)$. Similar to the proof of Proposition 2, we can show it suffices to solve the following problem

$$\max \Pi_s(0,0,z) \text{ s.t. } z \in [\underline{z},\bar{z}].$$

Because $\frac{\partial^2 \Pi_s(0,0,z)}{\partial z^2} = -\frac{2}{(r-d)+(\delta-1)V} < 0$, it follows that $\Pi_s(0,0,z)$ is strictly concave in z. The solution to the first order condition $\frac{\partial \Pi_s(0,0,z)}{\partial z} = 0$ is $\hat{z} = \frac{K}{2} - \frac{1}{2}((r-d)+(\delta-1)V)$. Therefore, the unique optimal subsidy z^* is given by

$$z^* = \begin{cases} \underline{z} & \text{if } \hat{z} < \underline{z}, \\ \hat{z} & \text{if } \underline{z} \leq \hat{z} \leq \bar{z}, \\ \bar{z} & \text{if } \bar{z} < \hat{z}, \end{cases}$$

or, equivalently, $z^* = \min\{\max\{\underline{z},\hat{z}\},\bar{z}\}$. It follows that z^* is the unique feasible and thus optimal solution to (A.10). The result follows. □

Proof of Theorem 5. Similar to the proof of Theorem 1, using the optimal incentives under mechanisms [D-H] and [D-I] presented in Propositions 10 and 11, it is straightforward to show

that $C^I = 0 < C^D$ and $W^I > W^0 = W^D$. Also, it can be shown that $D_m(0, 0, z^*) - D_m^0 \geq \beta D_m^0 = D_m(e^*, s^*, 0) - D_m^0$, or equivalently, $D_m^I \geq D_m^D$. Using the definitions in (2.13) and (2.14), it then follows that $\Pi_p^I \geq \Pi_p^D$ and $\Pi_r^I \geq \Pi_r^D$, respectively. With respect to the enterprise's profit, from (2.15) we have $\Pi^I - \Pi^D = \frac{(K-z^*)((d-p)-(\delta-1)V+z^*)}{(r-d)+(\delta-1)V} - \frac{(\beta+1)K((d-p)-(\delta-1)V)}{(r-d)+(\delta-1)V}$. Define $\tilde{\tau}(x) = \frac{\hat{z}^2-(\hat{z}-z^*)^2}{K((d-p)-(\delta-1)V)}$. Then $\Pi_s^I < \Pi_s^D$ holds if and only if $\beta > \tilde{\tau}(x)$. The results follow. $\square$

APPENDIX B

Planning Bike Lanes with Data: Ridership, Congestion, and Path Selection

B.1 Congestion Parameter Estimation

This section presents additional details on the data used to estimate and validate the congestion parameter (B.1.1), the estimation procedure described in in §4.4 (B.1.2), and tuning the penalty parameter λ (B.1.3).

B.1.1 Data and Summary Statistics

To make full use of the available data, we estimate $\boldsymbol{\theta}$ using data from the entire City of Chicago (larger area in Figure 4.1). The size of the network is given in Table 4.1 in §4.3. Table B.1 presents summary statistics for the following variables: The annual average daily traffic (AADT) during morning rush hours, which serve as the observed traffic flows z_s; the driving demand originating from each census tract, $d_i^{(o)}$; total road segment width $n_s \cdot w_s$; segment length, ℓ_s; and path free-flow travel time $\sum_{s \in \mathcal{S}_p^D} T_s$. We note that although we cannot observe the free-flow driving time at the segment-level, path-level free-flow times are sufficient for the purpose of estimating $\boldsymbol{\theta}$. To validate model fit, we use detailed data on 1,649,214 taxi trips taken during weekday morning rush hours (6am to 10am) in 2018. We match each taxi trip record to the road network to obtain average taxi travel times for 4,400 OD pairs. We drop records with trip times in the upper and lower 2.5% quantiles to remove extreme observations. A summary of taxi trip times is presented in Table B.2.

Table B.1: Summary of data for estimating congestion parameter $\boldsymbol{\theta}$.

	AADT z_s	Origin demand $d_i^{(o)}$	Road width $n_s \cdot w_s$	Road length ℓ_s	Free-flow time $\sum_{s \in \mathcal{S}_p^D} T_s$
Units	1	1	meters	meters	minutes
Mean	1,730	441	10.4	141	20
SD	2,698	272	4.3	127	9
Min	54	15	3.0	0.4	1
25%	837	236	7.2	99	13
50%	1,222	386	9.0	102	19
75%	1,671	575	13.6	201	25
Max	29,220	2,919	43.2	5,006	71

Table B.2: Summary of OD-level taxi trip times.

		Units	Mean	SD	Min	25%	50%	75%	Max
Taxi travel time	taxi_w	minutes	17.1	8.1	2.0	11.2	16.0	21.5	48.5

B.1.2 Estimation Procedure

For ease of solving WLS-A, throughout this section we eliminate the variable ϵ, and focus on the equivalent problem

$$\begin{aligned}&\underset{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b}}{\text{minimize}} \;\; \|\mathbf{v}-\mathbf{z}\|_2^2+\lambda\cdot\left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^{\top}\boldsymbol{\phi}-\left(\mathbf{d}^D\right)^{\top}\mathbf{b}\right)\\&\text{subject to (3.16c)}-\text{(3.16f)}.\end{aligned}$$

Let $d_i^{(o)}$ be the driving demand originating from each census tract $i\in\mathcal{I}$. Then any vehicle traffic pattern $\left(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}\right)$ should satisfy

$$\sum_{\substack{j\in\mathcal{I}\\ j\neq i}}\;\sum_{(i,j)\in\mathcal{W}} d_w^D=d_i^{(o)},\quad i\in\mathcal{I}. \tag{B.2}$$

Let $\Lambda=\{\left(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}\right)\mid\left(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}\right)$ satisfies (B.2) and $(\boldsymbol{\phi},\mathbf{v})\in\Omega(\mathbf{d}^D)\}$. We can then jointly impute the driving demands $\mathbf{d}^D$ by solving a variation of WLS-A that also includes the constraint $\left(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}\right)\in\Lambda$.

To ensure that driving times are non-decreasing in segment flows, we also impose the constraint $\boldsymbol{\theta}^{\top}\mathbf{q}_s(\tilde{\mathbf{x}})\geq\delta$, where $\delta>0$ is a small constant. Then the modified form of the estimator WLS-A for jointly estimating the driving demands $\mathbf{d}^D$ and the congestion parameter $\boldsymbol{\theta}$ is

$$\underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\theta},\mathbf{b}}{\text{minimize}} \;\; \|\mathbf{v}-\mathbf{z}\|_2^2+\lambda\cdot\left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^{\top}\boldsymbol{\phi}-\left(\mathbf{d}^D\right)^{\top}\mathbf{b}\right) \tag{B.3a}$$

$$\text{subject to}\sum_{\{w\in\mathcal{W}\mid p\in\mathcal{P}_w^D\}} b_w\leq\nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right),\quad p\in\mathcal{P}, \tag{B.3b}$$

$$\left(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}\right)\in\Lambda, \tag{B.3c}$$

$$\boldsymbol{\theta}^{\top}\mathbf{q}_s(\tilde{\mathbf{x}})\geq\delta,\quad s\in\mathcal{S}, \tag{B.3d}$$

$$\boldsymbol{\theta}\in\Theta. \tag{B.3e}$$

Formulation (B.3) is multi-convex in the sense that fixing $\left(\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v}\right)$ yields a convex subproblem in $(\boldsymbol{\theta}, \mathbf{b})$ and vice versa. We obtain estimates using a block coordinate descent (BCD) method that alternately minimizes the loss function over each block while fixing the other block at their last updated values (see Xu and Yin (2013)). Specifically, let $\left(\left(\mathbf{d}^D\right)^k, \left(\boldsymbol{\phi}\right)^k, \left(\mathbf{v}\right)^k\right)$ and $\left(\boldsymbol{\theta}^k, \mathbf{b}^k\right)$ be the variable values after their k^{th} update. The BCD method then consists of iteratively solving the following two subproblems:

$$\begin{aligned}\begin{bmatrix} \left(\mathbf{d}^D\right)^k \\ \left(\boldsymbol{\phi}\right)^k \\ \left(\mathbf{v}\right)^k \end{bmatrix} = \underset{\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v}}{\operatorname{argmin}} \ & \|\mathbf{v} - \mathbf{z}\|_2^2 + \lambda \cdot \left(\left(\nabla_{\phi} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}^{k-1}\right)\right)^{\top} \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^{\top} \mathbf{b}^{k-1}\right) \\ & + \frac{1}{2\mu_1^k} \left\| \begin{bmatrix} \mathbf{d}^D \\ \boldsymbol{\phi} \\ \mathbf{v} \end{bmatrix} - \begin{bmatrix} \left(\mathbf{d}^D\right)^{k-1} \\ \left(\boldsymbol{\phi}\right)^{k-1} \\ \left(\mathbf{v}\right)^{k-1} \end{bmatrix} \right\|^2 \end{aligned} \tag{B.4a}$$

$$\text{subject to} \sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w^D\}} b_w^{k-1} \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}^{k-1}\right), \quad p \in \mathcal{P} \tag{B.4b}$$

which is a quadratic program (QP), and

$$\begin{aligned}\begin{bmatrix} \boldsymbol{\theta}^k \\ \boldsymbol{b}^k \end{bmatrix} = \underset{\boldsymbol{\theta}, \mathbf{b}}{\operatorname{argmin}} \ & \|(\mathbf{v})^k - \mathbf{z}\|_2^2 + \lambda \cdot \left(\left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)\big|_{\boldsymbol{\phi}=(\boldsymbol{\phi})^k}\right)^{\top} (\boldsymbol{\phi})^k - \mathbf{b}^{\top} \left(\mathbf{d}^D\right)^k\right) \\ & + \frac{1}{2\mu_2^k} \left\| \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\theta}^{k-1} \\ \mathbf{b}^{k-1} \end{bmatrix} \right\|^2 \end{aligned} \tag{B.5a}$$

$$\text{subject to} \sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)\big|_{\boldsymbol{\phi}=(\boldsymbol{\phi})^k}, \quad p \in \mathcal{P} \tag{B.5b}$$

$$\boldsymbol{\theta}^{\top} \mathbf{q}_s(\tilde{\mathbf{x}}) \geq \delta, \quad s \in \mathcal{S}, \tag{B.5c}$$

$$\boldsymbol{\theta} \in \Theta, \tag{B.5d}$$

which is a linear program (LP). The sequences $\{\mu_1^k\}$ and $\{\mu_2^k\}$ are constants that are uniformly lower bounded from zero and uniformly upper bounded (Xu and Yin 2013). In our implementation, we simply set $\mu_1^k = \mu_2^k = 100$ for all k. Let $G_k = \|\mathbf{v} - \mathbf{z}\|_2^2 + \lambda \cdot$

$(\nabla_{\boldsymbol{\phi}} g(\boldsymbol{\phi}, \boldsymbol{\theta})^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b})$ be the loss after solving (B.5) in the k^{th} iteration. We terminate the BCD algorithm after $\frac{G_k - G_{k+1}}{G_k} < 10^{-4}$ holds for three consecutive iterations or when $k = 50$.

Because the BCD algorithm does not converge to a globally optimal solution, we aim to improve the model fit by repeating the estimation procedure multiple times with random initializations. We use 10 random starts, where each start consists of randomly drawing the initial estimate $\boldsymbol{\theta}^0$ from the uniform distribution, and then initialize the remaining parameters by solving (B.3) when $\lambda = 0$ and $\boldsymbol{\theta}^0$ is held fixed (in which case (B.3) becomes a quadratic program that is efficiently solved by commercial solvers). We observed that increasing the number of random starts beyond 10 only minimally improves the loss function at termination. For a wide range of λ values, we observed that the algorithm terminated after approximately 35 iterations, with an average convergence time of 5 hours. The estimation procedure was implemented using Python 3 and the optimization solver Gurobi, and run on a desktop computer with an Intel Core i9-10900K 3.70GHz processor and 128 GB of RAM.

B.1.3 Selecting Wardrop Error Penalty λ

To identify a value of the Wardrop error penalty parameter λ that produces a reasonable model fit, we repeat the estimation procedure using the BCD algorithm for each $\lambda \in \{1000, 1500, 2000, 2500, 3000\}$. For each value of λ, we measure model fit according to three error metrics: (1) flow error, (2) Wardrop error,[1] and (3) out-of-sample error[2]. The reason for evaluating model fit according to these three criteria is to balance the fit of the

[1] An alternative metric of evaluating how data comply with the Wardrop equilibrium is via the value of ϵ in WLS-A. Intuitively, the closer ϵ is to 0, the better the data satisfy Wardrop equilibrium. However, for ease of interpretation, here we measure the extent to which Wardrop equilibrium is violated as the relative difference between the maximum and minimum driving times over the chosen paths, averaged over all OD pairs.

[2] We use taxi trip number weighted average discrepancy between observed taxi time and predicted OD driving time as a measure of out-of-sample error as we expect more used OD pairs are more representative when measuring prediction errors.

data to the underlying traffic equilibrium model (metrics 1 and 2) with how well the model predicts driving travel times (metric 3). The three errors are computed as follows:

$$\text{Flow error } = \frac{\|\hat{\mathbf{v}} - \mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2}, \tag{B.6a}$$

$$\text{Wardrop error} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \frac{\max_{\{p \in \mathcal{P}_w | \hat{\phi}_p > 0\}} t_p\left(\tilde{\mathbf{x}}, \hat{\mathbf{v}}\right) - \min_{p \in \mathcal{P}_w} t_p\left(\tilde{\mathbf{x}}, \hat{\mathbf{v}}\right)}{\min_{p \in \mathcal{P}_w} t_p\left(\tilde{\mathbf{x}}, \hat{\mathbf{v}}\right)}, \tag{B.6b}$$

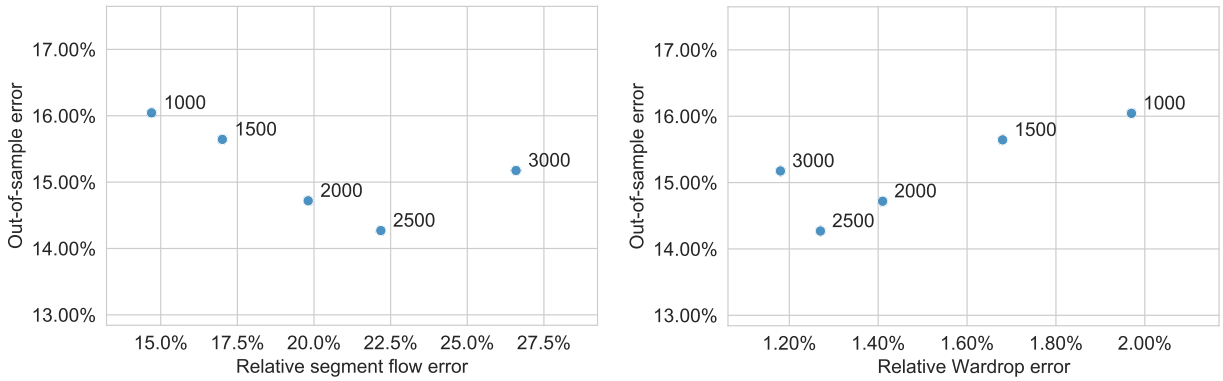
$$\text{Out-of-sample error} = \frac{1}{\sum_{w \in \mathcal{W}^{\text{taxi}}} |N_w^{\text{taxi}}|} \sum_{w \in \mathcal{W}^{\text{taxi}}} \left|N_w^{\text{taxi}}\right| \frac{\left|t_w\left(\tilde{\mathbf{x}}, \hat{\mathbf{v}}\right) - \text{taxi}_w\right|}{\text{taxi}_w}. \tag{B.6c}$$

In the expressions above, $(\hat{\mathbf{v}}, \hat{\boldsymbol{\phi}})$ represents the estimates of the equilibrium flows at termination of the estimation procedure, and $\tilde{\mathbf{x}}$ is the status quo bike lane network. In the computation of the out-of-sample error, N_w^{taxi} is the total number of taxi trips for OD pair w, $\mathcal{W}^{\text{taxi}} \subset \mathcal{W}$ is the set of OD pairs for which $N_w^{\text{taxi}} > 0$, and taxi_w is the average duration of all taxi trips on OD pair $w \in \mathcal{W}^{\text{taxi}}$. Note that the out-of-sample error is computed only using the taxi trip data (described in §4.3 and §B.1.1), which is not used during estimation.

Figure B.1 visualizes the performance of each value of λ according to the three error metrics. Although there is no Pareto optimal value, we adopt $\lambda = 2500$ for our §4 empirical study, which is the only value of λ that underperforms the other values on only one error metric, and also has the lowest out-of-sample error. Under $\lambda = 2500$ the flow error is 22%, the Wardrop error is 1.3%, and the out-of-sample error is 14.3%.

For completeness, Table B.3 presents the simple average error across OD pairs as an alternative measure of out-of-sample error, which is given by setting $|N_w^{\text{taxi}}| = 1$ in (B.6c).

The errors in Table B.3 are competitive with previously proposed methods for predicting travel times in a network using taxi trip data (cf. Table 2 in Zhan et al. (2013)). Table B.3 also reports the errors if the free-flow travel time from Google Maps is naively used to predict taxi trip times. We observe that the error from using free-flow times alone is significantly higher than our method regardless of which value of λ is selected; in the case where $\lambda = 2500$, we obtain a weighted out-of-sample error of 14%, whereas the free-flow travel times yield an error of 27%. This improvement in prediction errors can be interpreted



(a) Taxi error vs. flow error.

(b) Taxi error vs. Wardrop error.

Figure B.1: Estimation errors for varying values of penalty parameter λ.

as the value added by using a model of traffic congestion in addition to free-flow travel times when making predictions.

Table B.3: Weighted and simple out-of-sample errors for varying λ and free-flow time.

λ	1000	1500	2000	2500	3000	Free-flow time
Out-of-sample error (Weighted)	16.0%	15.6%	14.7%	14.3%	15.2%	27.3%
Out of sample error (Simple)	19.7%	19.6%	19.1%	18.9%	19.1%	31.6%

B.2 Map of Divvy Bike Share Stations for Mode Choice Estimation

We use Divvy bike share trip data as a proxy for total cycling demand. Because not all census tracts have Divvy stations, in estimating the mode choice parameters $\boldsymbol{\beta}$ we focus on a contiguous region within Chicago, which we identify by selecting a subset of census tracts for which a Divvy station exists in every adjacent census tract. Figure B.2 depicts locations of all Divvy stations in Chicago, and the study region for estimating the mode choice parameters, which approximately corresponds to downtown Chicago. The road network corresponding to the study region area is summarized in Table B.4.

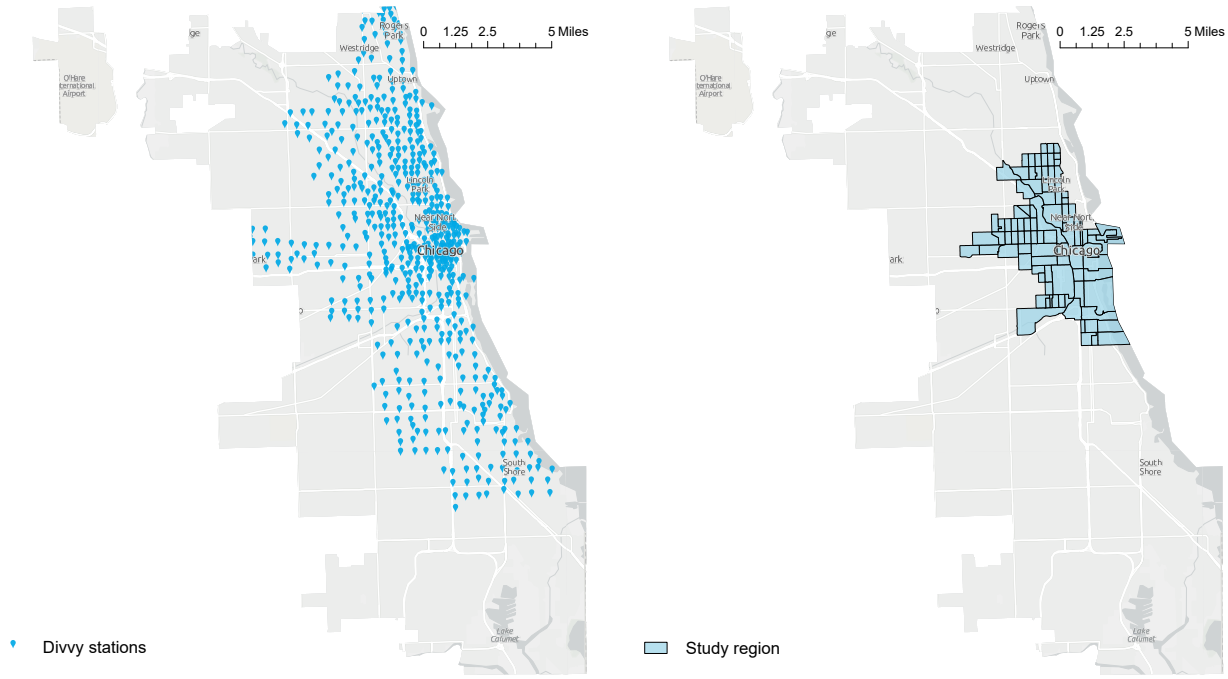


Figure B.2: Locations of Divvy stations and study region for mode choice estimation.

Table B.4: Network Size for Mode Choice Estimation.

	Road segments $\lvert\mathcal{S}\rvert$	Driving paths $\left\lvert\mathcal{P}^D\right\rvert$	OD pairs $\lvert\mathcal{W}\rvert$
Size	6,101	14,632	5,272

B.3 MILP Reformulation of Path Selection Model BLP-A

This section develops the MILP reformulation of the linearized bike lane path selection model BLP-A given in (3.33). In preparation, we re-write the constraints (3.31d) and (3.31e) as

$$\xi_s \geq \rho_s^r\left(x_s\right) \cdot v_s + \epsilon_s^r\left(x_s\right), \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{B.7a}$$

$$\psi_w^m \geq a_w^r \cdot d_w^m + b_w^r, \quad r \in \mathcal{R}_\psi, w \in \mathcal{W}, m \in \mathcal{M}, \tag{B.7b}$$

where

$$\rho_s^r\left(x_s\right) = \alpha_s\left(x_s\right) \cdot v_s^r \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{B.8a}$$

$$\epsilon_s^r\left(x_s\right) = -\frac{1}{2} \cdot \alpha_s\left(x_s\right) \cdot \left(v_s^r\right)^2, \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{B.8b}$$

$$a_w^r = \left[1 + \log\left(d^r\right)\right], \quad r \in \mathcal{R}_\psi, w \in \mathcal{W}, m \in \mathcal{M}, \tag{B.8c}$$

$$b_w^r = -d^r, \quad r \in \mathcal{R}_\psi, w \in \mathcal{W}, m \in \mathcal{M}. \tag{B.8d}$$

The lower-level (i.e., user equilibrium) problem in in (3.33) is a linear program. As a result, by leveraging linear programming duality, the bi-level optimization problem (3.33) can be reformulated exactly as the following single-level mixed-integer program, where $\boldsymbol{\sigma}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\gamma}$

are dual variables corresponding to each of the primal feasibility constraints:

$$\underset{\substack{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\xi},\\ \boldsymbol{\psi},\boldsymbol{\mu},\boldsymbol{\gamma},\boldsymbol{\lambda},\boldsymbol{\pi},\boldsymbol{\sigma}}}{\text{maximize}} \quad \sum_{w\in\mathcal{W}} d_w^C$$

$$\text{subject to} \quad (3.24\text{b}) - (3.24\text{d}), \qquad \text{upper level constraints}$$

$$S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) = \sum_{w\in\mathcal{W}} \left(\mu_w d_w + \sum_{r\in\mathcal{R}_\psi} \sum_{m\in\mathcal{M}} \pi_w^{mr} b_w^r \right)$$

$$+ \sum_{r\in\mathcal{R}_\xi} \sum_{s\in\mathcal{S}} \sigma_s^r \epsilon_s^r\left(x_s\right), \qquad \text{strong duality}$$

$$\left.\begin{array}{ll}
\xi_s \geq \rho_s^r\left(x_s\right)\cdot v_s + \epsilon_s^r\left(x_s\right), & r\in\mathcal{R}_\xi, s\in\mathcal{S}, \\
\psi_w^m \geq a_w^r \cdot d_w^m + b_w^r, & r\in\mathcal{R}_\psi, w\in\mathcal{W}, m\in\mathcal{M}, \\
d_w^O + d_w^C + d_w^D = \bar{d}_w, & w\in\mathcal{W}, \\
\sum_{p\in\mathcal{P}_w^D} \phi_p = d_w^D, & w\in\mathcal{W}, \\
v_s = \sum_{\{p\in\mathcal{P}^D | s\in\mathcal{S}_p^D\}} \phi_p, & s\in\mathcal{S}, \\
\boldsymbol{\phi} \geq \mathbf{0}, &
\end{array}\right\} \text{primal feasibility}$$

$$\left.\begin{array}{ll}
\mu_w - \sum_{r\in\mathcal{R}_\psi} a_w^r \cdot \pi_w^{mr} \leq u_w^m, & w\in\mathcal{W}, m\in\{C,O\}, \\
\mu_w + \gamma_w - \sum_{r\in\mathcal{R}_\psi} a_w^r \cdot \pi_w^{Dr} \leq \beta_0^D, & w\in\mathcal{W}, \\
-\sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w^D\}} \gamma_w - \sum_{s\in\mathcal{S}_p^D} \lambda_s \leq 0, & p\in\mathcal{P}^D, \\
\lambda_s - \sum_{r\in\mathcal{R}_\xi} \rho_s^r\left(x_s\right)\cdot\sigma_s^r \leq \beta_1^D \cdot T_s, & s\in\mathcal{S}, \\
\sum_{r\in\mathcal{R}_\psi} \sigma_s^r = \beta_1^D, & s\in\mathcal{S}, \\
\sum_{n\in\mathcal{N}} \pi_w^{mn} = 1, & w\in\mathcal{W}, m\in\mathcal{M}, \\
\boldsymbol{\pi},\boldsymbol{\sigma} \geq \mathbf{0}. &
\end{array}\right\} \text{dual feasibility}$$

The advantage of the formulation above is that it only contains linear and bi-linear terms, and can therefore be transformed into a MILP via introduction of "big-M" constants, which permits solution by commercial optimization software packages. Specifically, because $\rho_s^r(x_s)$ and $\epsilon_s^r(x_s)$ are functions of the bike lane decision $\mathbf{x}$, we rewrite the bi-linear constraints $\xi_s \geq \rho_s^r(x_s) \cdot v_s + \epsilon_s^r(x_s)$ and $\lambda_s - \sum_{r \in \mathcal{R}_\xi} \rho_s^r(x_s) \cdot \sigma_s^r \leq \beta_1^D \cdot T_s$ equivalently as

$$\xi_s \geq \rho_s^r(1) \cdot v_s + \epsilon_s^r(1) - M_1 \cdot (1 - x_s), \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{B.10a}$$

$$\xi_s \geq \rho_s^r(0) \cdot v_s + \epsilon_s^r(0) - M_1 \cdot x_s, \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{B.10b}$$

and

$$\lambda_s \leq \beta_1^D \cdot T_s + \sum_{r \in \mathcal{R}_\xi} \rho_s^r(1) \cdot \sigma_s^r + M_2 \cdot (1 - x_s), \quad s \in S, \tag{B.11a}$$

$$\lambda_s \leq \beta_1^D \cdot T_s + \sum_{r \in \mathcal{R}_\xi} \rho_s^r(0) \cdot \sigma_s^r + M_2 \cdot x_s, \quad s \in S, \tag{B.11b}$$

where M_1 and M_2 are large constants. For computational efficiency in solving the bike lane path selection model BLP-A, we omit the OD pairs representing the smallest 20% of pairs by demand within the study region (shaded area of Figure B.2). This yields a sub-network that is still representative of commute patterns in the study region, but is significantly more tractable. The resulting network is summarized in Table B.5.

Table B.5: Network Size in Bike Lane Planning

	Road segment set $\lvert\mathcal{S}\rvert$	Driving path set $\lvert\mathcal{P}^D\rvert$	OD pair set $\lvert\mathcal{W}\rvert$
Size	4,127	2,458	887

B.3.1 Sensitivity of Approximation Error to Number of Linear Segments

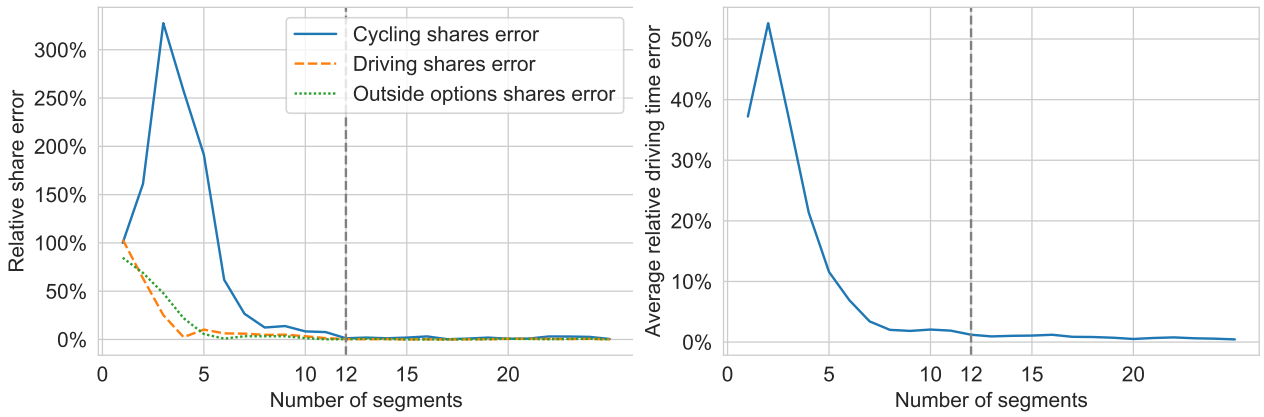
In our §4 empirical study, we set the number of piecewise linear segments in the approximation BLP-A as $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 12$. In this section, we investigate how the number of

segments used in the linearization technique affects the approximation error between the exact formulation BLP and the approximation BLP-A. To do so, we fix $|\mathcal{R}_\xi| = |\mathcal{R}_\psi|$, vary both from 1 to 50, and compare the mode shares ($d_w^m/\bar{d}_w$) and driving times (t_p) computed by the linearized lower-level problem (3.31) to the true mode shares and driving times from the exact lower-level problem (3.25). We assume the status quo bike lane network (i.e., $\mathbf{x} = \tilde{\mathbf{x}}$). Note that the exact lower-level problem is a convex (quadratic) optimization problem and can thus be solved to optimality.

Figure B.3 shows the relative errors in mode shares and driving times between the exact and linearized user equilibrium problems. When $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 12$, the relative errors for cycling, driving and outside options shares are 1.25%, 0.42% and 0.27%, respectively, and the average relative driving time error across all OD pairs is 1.21%. Note that because the objective function of BLP is simply the total cycling demand $\sum_{w \in \mathcal{W}} d_w^c$, the relative error in cycling mode shares of 1.25% suggests the suboptimality bound in Theorem 4 is tight when $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 12$. In general, these results suggest that a modest number of linear segments is sufficient for obtaining a good-quality MILP approximation of the exact path selection model BLP.

B.3.2 Optimality Gaps at Termination of MILP Formulation

In solving the MILP reformulation of BLP-A with commercial solvers, we terminate solution when either the optimality gap (MIP gap) is within 0.1% or the solution time reaches 6 hours. In Table B.6, we present the relative MIP gaps at termination for each of the 12 combinations of τ and B values used in the §4 study. The gaps are largest when both the budget constraint B and the tolerance τ are small, with a maximum MIP gap of 23.4% for the instance $(B, \tau) = (10, 5\%)$. This magnitude of MIP gap is commonly observed for computationally challenging bi-level programs (Dan and Marcotte 2019). Because BLP-A is not solved to optimality, the estimated cycling ridership under the recommended bike lane network presented in §4.5 and §4.6 are conservative estimates of the ridership under an



(a) Mode share errors.

(b) Path driving time errors.

Figure B.3: Approximation errors due to linearization of user equilibrium problem (3.25).

optimal solution to BLP-A.

Table B.6: Relative optimality gaps at termination when solving BLP-A as a MILP.

	$B = 10$	$B = 25$	$B = 50$	$B = 75$
$\tau = 5\%$	23.59%	19.33%	12.83%	12.42%
$\tau = 10\%$	12.78%	5.02%	2.58%	0.90%
$\tau = 15\%$	12.72%	4.82%	1.76%	0.31%

B.4 Comparison with Alternative Bike Lane Planning Methods

This section provides additional technical details and numerical results for the benchmark methods discussed in §4.

B.4.1 Fixed-Time Model

The fixed-time model assumes that all driving times are unaffected by the addition of bike lanes. The model is a variation of BLP-A, where the segment travel times t_s are fixed to their values under the status quo bike lane network $\tilde{\mathbf{x}}$. Under this assumption, the total driving disutility is

$$\sum_{w\in\mathcal{W}}\left(\int_0^{d_w^D(\mathbf{x})} u_w^D(\mathbf{x})\cdot dd\right) = \sum_{w\in\mathcal{W}}\beta_0^D\cdot d_w^D(\mathbf{x}) + \beta_1^D\sum_{s\in\mathcal{S}} t_s(\tilde{\mathbf{x}})\cdot v_s. \tag{B.12}$$

We then obtain the following fixed-time analog of $S(\mathbf{x},\mathbf{d},\mathbf{v})$:

$$S^{\mathrm{FT}}(\mathbf{x},\mathbf{d},\mathbf{v}) = \sum_{w\in\mathcal{W}}\left(u_w^C(\mathbf{x})\cdot d_w^C + \beta_0^D\cdot d_w^D\right) + \beta_1^D\sum_{s\in\mathcal{S}} t_s(\tilde{\mathbf{x}})\cdot v_s + \sum_{w\in\mathcal{W}}\sum_{m\in\mathcal{M}} d_w^m\cdot\log\left(d_w^m\right). \tag{B.13}$$

Note that the dependence on the bike lane design variable $\mathbf{x}$ is now only via the cycling utility u_w^C. The full fixed-time model is then given by

$$\underset{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v}}{\text{maximize}} \quad \sum_{w\in\mathcal{W}} d_w^C \tag{B.14a}$$

$$\text{subject to} \quad (3.24\text{b}) - (3.24\text{d}) \tag{B.14b}$$

$$(\mathbf{d},\boldsymbol{\phi},\mathbf{v}) = \underset{\mathbf{d},(\phi,\mathbf{v})\in\Omega(\mathbf{d}^D)}{\text{argmin}} \quad S^{\mathrm{FT}}(\mathbf{x},\mathbf{d},\mathbf{v}). \tag{B.14c}$$

As a result of the fixed-time assumption, congestion tolerance constraints such as (3.35) are irrelevant in this model, and (B.14) does not permit control over driving times. The model can be reformulated and linearized in the same manner as described in §3.4.2, which leads to a MILP that can be solved using commercial optimization software. The termination

criteria are the same as in §B.3.2: the solver is terminated when the optimality gap (MIP gap) is within 0.1% or the solution time reaches 6 hours. Table B.7 reports the optimality gaps at termination for different values of B. We observe that the optimality gaps for the fixed-time model (B.14) are comparable to the gaps for BLP-A presented in Table B.6.

Table B.7: Relative optimality gaps at termination when solving fixed-time model as a MILP.

B	10	25	50	75
Gap	15.74%	3.78%	0.83%	0.17%

B.4.2 Greedy Heuristic

Here we present detailed steps for the greedy heuristic discussed in §4.6. With a slight abuse of notation, let $\tilde{d}_w^C$ be the cycling demand on OD pair w under the status quo bike lane network $\tilde{\mathbf{x}}$, and let $d_{w'}^{C,w}$ be the cycling demand on OD pair w' *if* bike lanes are added to path p_w^C. Similarly, let $\tilde{t}_p$ be the driving time on path p under $\tilde{\mathbf{x}}$, and let t_p^w be the driving time on path p if a bike lane is added to path p_w^C. Define

$$\tau_w = \max_{p \in \mathcal{P}^D} \left\{ \frac{t_p^w - \tilde{t}_p}{\tilde{t}_p} \right\} \tag{B.15}$$

to be the worst-case relative increase in driving times over all paths if a bike lane is added to path p_w^C. Note that $d_{w'}^{C,w}$ and t_p^w can be computed by modifying the network variable $\mathbf{x}$ to contain the desired bike lanes and then solving the convex user equilibrium problem (3.25). Next, define

$$B_w = \sum_{s \in S_w^C} \ell_s (1 - \tilde{x}_s), \tag{B.16}$$

which represents the consumption of the budget B due to the addition of bike lanes to path p_w^C. Lastly, define

$$\delta_w = \frac{1}{B_w} \sum_{w' \in \mathcal{W}} \left(d_{w'}^{C,w} - \tilde{d}_{w'}^C \right). \tag{B.17}$$

Intuitively, the parameter δ_w represents the average increase in system-wide cycling demand *per unit* of bike lane added to path p_w^C. In other words, δ_w reflects the sensitivity (marginal increase) of cycling ridership to the addition of bike lanes on path p_w^C. The pseudocode for the greedy heuristic is presented below.

Greedy heuristic.

Input: Budget B, existing bike lane network $\tilde{\mathbf{x}}$ and $\tilde{S}$, congestion tolerance τ, step parameter $\gamma \in [0, 1]$, maximum of driving time increase $\sigma = \max_{w \in \mathcal{W}} \tau_w$.

Output: New bike lane network $\mathbf{x}^G$.

1. Let $\mathcal{W}^k \subseteq \mathcal{W}$ index the k largest values of δ_w. Let $\mathcal{W}_\sigma^k = \{w \in \mathcal{W}^k | \tau_w \leq \sigma\}$ and $\mathcal{S}_\sigma^k = \bigcup_{w \in \mathcal{W}_\sigma^k} \mathcal{S}_w^C$.
2. Set $\bar{\mathbf{x}} = \tilde{\mathbf{x}}$. Let G be the largest index in $\mathcal{W}_\sigma^k$ such that $\sum_{s \in \mathcal{S}_\sigma^G} \ell_s (1 - \tilde{x}_s) \leq B$. Set $\bar{x}_s = 1$ for all $s \in \mathcal{S}_\sigma^G \bigcup \tilde{S}$.
3. If $\max_{p \in \mathcal{P}^D} \{t_p(\bar{\mathbf{x}})/\tilde{t}_p\} > 1 + \tau$, update $\sigma \leftarrow \sigma - \gamma$ and go to Step 1; else, set $\mathbf{x}^G = \bar{\mathbf{x}}$ and terminate.

B.4.3 Demand Heuristic

The demand heuristic greedily selects bike lane paths according the OD pairs with the highest status-quo cycling demand. It differs from the greedy heuristic in that it ignores both the congestion constraints and the network structure, but is still restricted to select from the same set of candidate paths. The pseudocode is presented below.

Demand heuristic.

Input: Budget B, existing bike lane network $\tilde{\mathbf{x}}$ and $\tilde{S}$.

Output: New bike lane network $\mathbf{x}^H$.

1. Initialize $\mathbf{x}^H = \tilde{\mathbf{x}}$. Let $\mathcal{W}^k \subseteq \mathcal{W}$ index the k largest demands d_w^C and let $\mathcal{S}^k = \bigcup_{w \in \mathcal{W}^k} \mathcal{S}_w^C$.
2. Let H be the largest index such that $\sum_{s \in \mathcal{S}^H} \ell_s \left(1 - \tilde{x}_s\right) \leq B$.
3. Set $x_s^H = 1$ for all $s \in \mathcal{S}^H \bigcup \tilde{S}$.

B.4.4 Results

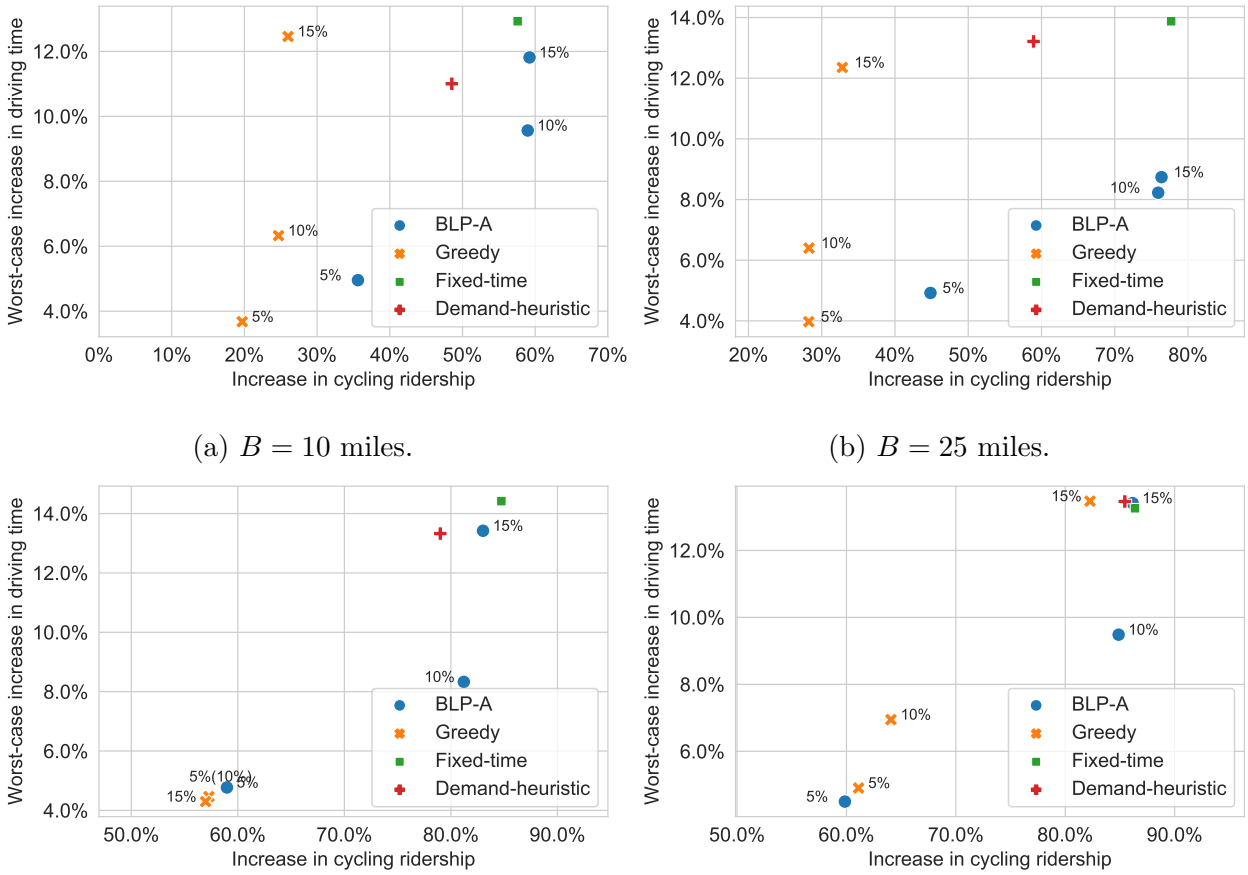
Figure B.4 depicts the performance of BLP-A, the fixed-time model, the greedy heuristic, and the demand heuristic for the full set of cases where $B = \{10, 25, 50, 75\}$. The results are consistent with the findings discussed in §4.6 for the $B = 25$ case. First, the greedy heuristic attains the lowest increase in cycling ridership. Second, although the fixed-time model and demand heuristic both lead to substantial increases in cycling ridership, the increase in congestion under those methods are comparable or worse than the prescriptions from BLP-A.

In addition to the worst-case increase in driving time over all routes, as an alternative measure of congestion we also evaluate the total system-wide driving time, which is computed as

$$\frac{\sum_{w \in \mathcal{W}} d_w^D\left(\mathbf{x}^*\right) \cdot t_w(\mathbf{x}^*) - \sum_{w \in \mathcal{W}} d_w^D\left(\tilde{\mathbf{x}}\right) \cdot t_w(\tilde{\mathbf{x}})}{\sum_{w \in \mathcal{W}} d_w^D\left(\tilde{\mathbf{x}}\right) \cdot t_w(\tilde{\mathbf{x}})}. \tag{B.18}$$

Figure B.5 presents a comparison of all four methods according to the increase in cycling ridership and total system-wide driving time. We note that the congestion-reducing *demand effect* and congestion-increasing *road effect* discussed in §4.5 remain in effect here. Notably, under most bike lane plans, the total system-wide driving time *decreases* compared to the status quo. This is the result of the demand effect dominating the road effect, which also explains why driving time is often decreasing in τ. Further, when the bike lane budget B is small ($B \in \{10, 25, 50\}$), the recommended bike lane plan from BLP-A strictly dominates the

greedy heuristic, and outperforms the fixed-time model and demand heuristic in at least one criterion; when the budget is large ($B = 75$), all methods tend to perform similarly. This is unsurprising because the bike lane plans produced by all methods less likely to differ under a generous budget. Overall, these results suggest that the bike lane plans recommended by our model continue to outperform the benchmark methods using the alternative congestion metric of total system-wide driving time, and that our approach is particularly valuable when the bike lane mileage budget is modest.



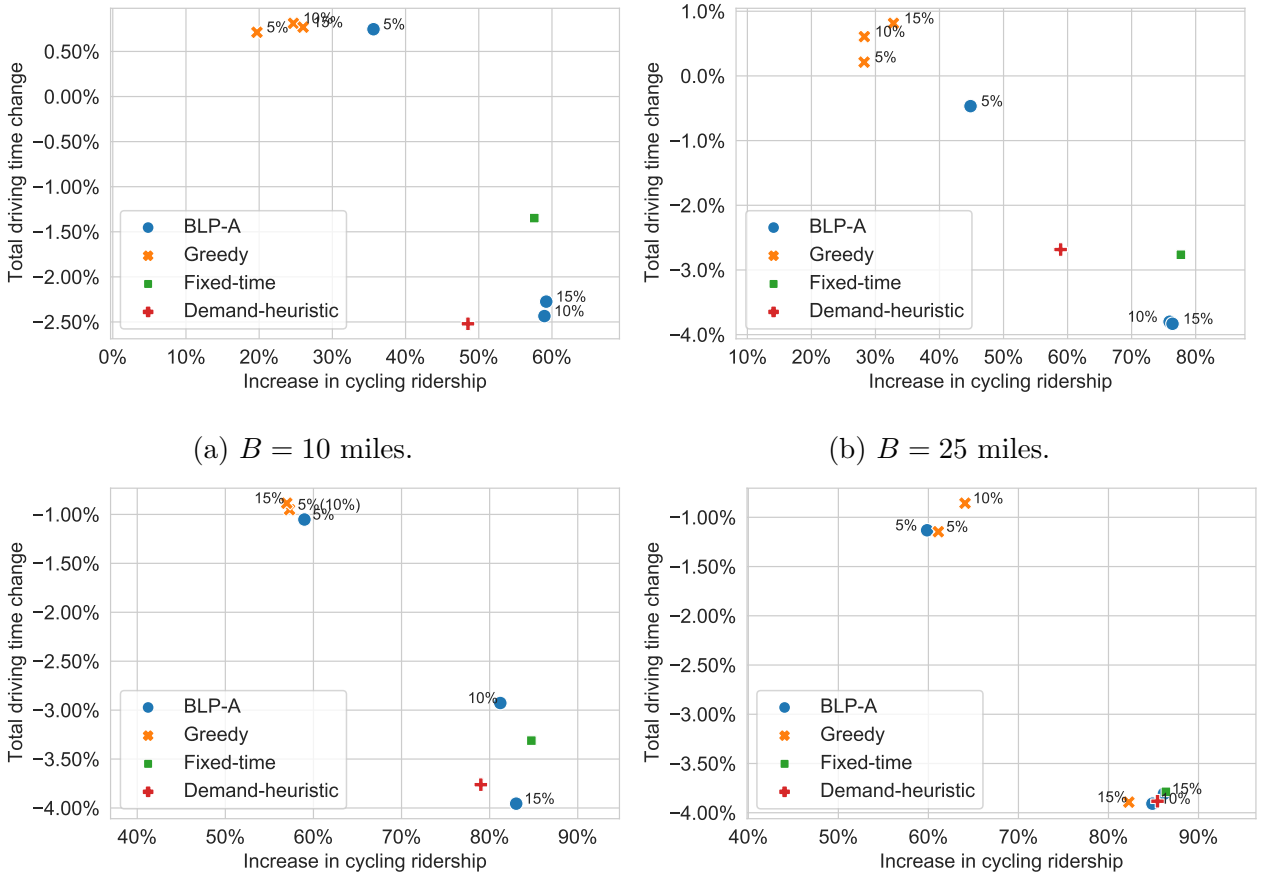
(a) $B = 10$ miles.

(b) $B = 25$ miles.

(c) $B = 50$ miles.

(d) $B = 75$ miles.

Figure B.4: Comparison of BLP-A and benchmark methods on worst-case driving time and cycling ridership.



(a) $B = 10$ miles.

(b) $B = 25$ miles.

(c) $B = 50$ miles.

(d) $B = 75$ miles.

Figure B.5: Comparison of BLP-A and benchmark methods on system-wide driving time and cycling ridership.

B.5 Validating WLS-A Estimator with Synthetic Data

We provide additional numerical results for validating our estimation procedure for $\boldsymbol{\theta}$. The purpose of these experiments is two-fold: (1) to test the estimator's ability to learn the true parameter $\boldsymbol{\theta}$ and the demand $\mathbf{d}^D$ when demand data is incomplete (i.e., only observed at the origins), and (2) to empirically verify the statistical consistency of the estimator (i.e., that estimation error decreases in network size.)

B.5.1 Data and Setup

We consider four different sizes of networks, where $|\mathcal{I}| \in \{250, 500, 750, 1000\}$. For each network size, the underlying data-generation process is as follows. We first randomly generate an undirected connected road network with $|\mathcal{I}|$ nodes, where the probability of any edge (i.e., road segment) existing is 0.1. The length of each road segment ℓ_s is generated from the uniform distribution with support $[0, 100]$. The free flow time T_s assumed to be $0.4 \cdot \ell_s$. Each road segment has a total width $n_s \cdot w_s$ uniformly distributed over the set $\{3, 3.3, 3.6\}$, which are standard lane width values according to the Illinois Department of Transportation's Bureau of Local Roads and Streets Manual (Illinois DOT 2018).

Next, we randomly generate $|\mathcal{W}| = |\mathcal{I}|$ different OD pairs from all possible OD combinations. For each OD pair $w \in \mathcal{W}$, the ground truth demand d_w^D is a random draw from the uniform distribution with support $[1, 101]$, and the path set $\mathcal{P}_w^D$ is constructed to be the three shortest paths (in length) connecting the OD pair. We let $d_w^{(o)}$ be the true total driving demand at the origin for the pair w, and assume the observed demand at the origin of w is given by $d_w^{(o)} + \varepsilon_w^o$, where $\varepsilon_w^{(o)}$'s are drawn i.i.d. from the standard normal distribution.

We assume the congestion function $\alpha_s(\boldsymbol{\theta})$ is as given in (4.2). We generate a four-dimensional ground-truth parameter vector $\boldsymbol{\theta}_0$ by randomly generating each component from the uniform distribution with support $[0, 10^{-4}]$. The (unobserved) true equilibrium traffic flow $\mathbf{v}^*(\boldsymbol{\theta}_0)$ is then obtained by solving problem (B.20). The observed traffic flow is then

assumed to be $z_s = v_s^*(\boldsymbol{\theta}_0) + \varepsilon_s^v$ for $s \in \mathcal{S}$, where ε_s^v's are drawn i.i.d. from the standard normal distribution.

For each randomly generated network, we repeat the estimation procedure for 50 different randomly generated values of $\boldsymbol{\theta}$. To evaluate the performance of the estimator with and without knowledge of demand destination data, we repeat the experiment assuming complete knowledge of the demand vector $\mathbf{d}^D$ for the case where $|\mathcal{I}| = |\mathcal{W}| = 1000$.

Let $\hat{\mathbf{d}}^D, \hat{\boldsymbol{\theta}}$ be the estimated OD demand and θ vectors. A summary of the experimental setup is provided in Table B.8.

Table B.8: Summary of synthetic experiments.

Experiment	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{W}\rvert$	$\lvert\mathcal{S}\rvert$	Observed Data	Estimates	Trials
O-250	250	250	201	$\mathbf{d}^{(o)} + \boldsymbol{\varepsilon}^{(o)}$	$\hat{\mathbf{d}}^D, \hat{\boldsymbol{\theta}}$	50
O-500	500	500	483	$\mathbf{d}^{(o)} + \boldsymbol{\varepsilon}^{(o)}$	$\hat{\mathbf{d}}^D, \hat{\boldsymbol{\theta}}$	50
O-750	750	750	1,107	$\mathbf{d}^{(o)} + \boldsymbol{\varepsilon}^{(o)}$	$\hat{\mathbf{d}}^D, \hat{\boldsymbol{\theta}}$	50
O-1000	1,000	1,000	1,630	$\mathbf{d}^{(o)} + \boldsymbol{\varepsilon}^{(o)}$	$\hat{\mathbf{d}}^D, \hat{\boldsymbol{\theta}}$	50
OD-1000	1,000	1,000	2,235	$\mathbf{d}^D$	$\hat{\boldsymbol{\theta}}$	50

B.5.2 Estimation and Results

For each of the 50 trials, we apply the estimation procedure described in §B.1.2. In each trial, we apply the BCD algorithm using 10 random starts of $\boldsymbol{\theta}$ and varying $\lambda \in \{50, 100, 150, 200\}$. The algorithm is terminated whenever the relative reduction in objective value is no more than 10^{-4} from the previous iteration for three iterations in a row or the number of iterations exceeds 50. For each value of λ, we identify the random start value for $\hat{\boldsymbol{\theta}}$ that attains the lowest loss according to the objective function of WLS-A. We then select λ as the value that

minimizes the sum of three error terms:

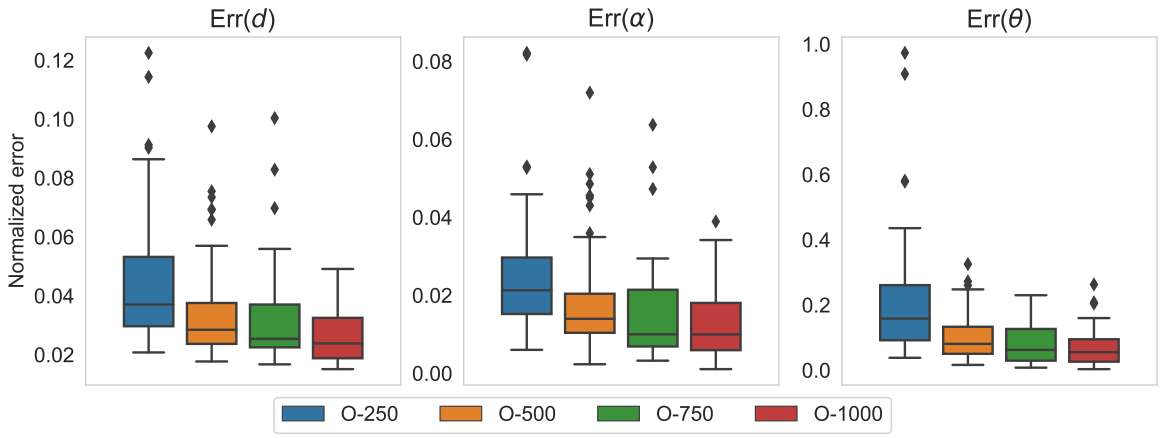
$$\underbrace{\frac{\|\hat{\mathbf{d}}^D - \mathbf{d}^D\|_2}{\|\mathbf{d}^D\|_2}}_{\text{Err}(d)} + \underbrace{\frac{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2}{\|\boldsymbol{\theta}_0\|_2}}_{\text{Err}(\theta)} + \underbrace{\frac{\|\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(\boldsymbol{\theta}_0)\|_2}{\|\boldsymbol{\alpha}(\boldsymbol{\theta}_0)\|_2}}_{\text{Err}(\alpha)}, \tag{B.19}$$

The first two terms are simply the normalized estimation errors for the true parameters $\mathbf{d}^D$ and $\boldsymbol{\theta}_0$. Note that because the congestion functions $\alpha_s(\boldsymbol{\theta})$, $s \in \mathcal{S}$ depend on the unknown parameter $\boldsymbol{\theta}_0$, we can also measure how well the estimator recovers the true congestion functions, i.e., $\alpha_s(\boldsymbol{\theta}_0)$. This error is captured by the third term in (B.19), where $\hat{\boldsymbol{\alpha}}$ is the estimate of the function value.

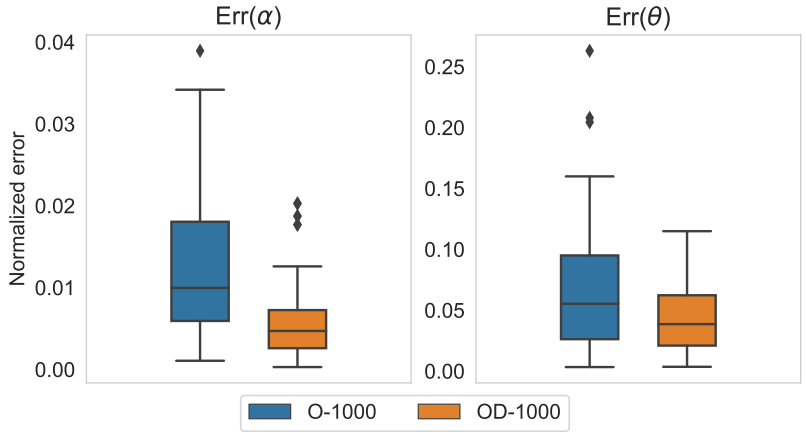
Figure B.6 visualizes the errors for all experiments, and Table B.9 reports average errors over 50 trials. We offer a few remarks on the results. First, in the main setting where demand data is incomplete (O-250 to O-1000), the estimation error for $\boldsymbol{\theta}$ decreases from 0.22 to 0.066 as the size of the network grows from 250 to 1000 nodes (and the number of segments grows from 201 to 2,235). This decrease in error aligns with the consistency result provided in Theorem 3. However, we note that our experimental setting differs from what is assumed by Theorem 3 in the following ways: 1) for tractability we solve the approximate estimator WLS-A instead of the exact estimator WLS, 2) the BCD algorithm is not guaranteed to return a globally optimal solution to either WLS-A or WLS, and 3) we assume the demand data is incomplete in that only aggregated origin-level demands are observed. Nevertheless, our numerical results suggest that the proposed estimation procedure is reasonably accurate at recovering the ground-truth congestion parameter, even when the assumptions of Theorem 3 are not strictly satisfied. Second, as expected, when demand data is incomplete (O-1000), the estimation errors for $\boldsymbol{\theta}$ is higher (0.066 vs. 0.044) than when complete demand data is available (OD-1000). Taken together, these results provide empirical evidence in support of Theorem 3 and suggest that solving the approximate estimator WLS-A using the block coordinate descent algorithm described in §B.1.2 produces sound estimates on realistically sized networks.

Table B.9: Normalized estimation errors over 50 trials.

Experiment	Err(d)	Err(α)	Err(θ)
O-250	0.0450	0.0263	0.2211
O-500	0.0351	0.0189	0.1043
O-750	0.0319	0.0153	0.0799
O-1000	0.0262	0.0130	0.0663
OD-1000	-	0.0057	0.0440



(a) Errors with origin demands only.



(b) Comparison of errors with and without destination demands.

Figure B.6: Normalized estimation errors of $\mathbf{d}^D$, $\boldsymbol{\theta}$, and $\boldsymbol{\alpha}$ for varying network sizes and demand data availability.

B.6 Proofs

Proof of Proposition 5. First, for any fixed bike lane design $\mathbf{x}$, driving demand $\mathbf{d}^D$, and congestion parameter $\boldsymbol{\theta}$, it can be shown that $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$ if and only if $(\boldsymbol{\phi}, \mathbf{v})$ is an optimal solution to the following convex program:

$$\underset{\boldsymbol{\phi}, \mathbf{v}}{\text{minimize}} \quad \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot v_s^2 + T_s \cdot v_s \right) \tag{B.20a}$$

$$\text{subject to} \quad \sum_{p \in \mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{B.20b}$$

$$v_s = \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p, \quad s \in \mathcal{S}, \tag{B.20c}$$

$$\boldsymbol{\phi}, \ \mathbf{v} \geq \mathbf{0}. \tag{B.20d}$$

This equivalence follows from a well-known result shown in Beckmann et al. (1956) [add more here]. Next, we re-write formulation (B.20) by using a variable change based on the equality in (B.20c). Substituting $\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p$ for v_s allows us to take the constraint (B.20c) into the objective (B.20a), which yields

$$g(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\tilde{\mathbf{x}}) \cdot \left[\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p \right]^2 + T_s \cdot \left[\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p \right] \right) \tag{B.21}$$

It follows that $(\boldsymbol{\phi}, \ \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$ if and only if $\boldsymbol{\phi}$ is an optimal solution to

$$\underset{\boldsymbol{\phi}}{\text{minimize}} \quad g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right) \tag{B.22a}$$

$$\text{subject to} \quad \sum_{p \in \mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{B.22b}$$

$$\boldsymbol{\phi} \geq \mathbf{0}. \tag{B.22c}$$

Next, expressing optimality conditions as a variational inequality, a candidate solution $\boldsymbol{\phi}$ is optimal to (B.22) if and only if the following inequalities hold:

$$\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)^{\top} \left(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \geq 0, \tag{B.23a}$$

$$\sum_{p \in \mathcal{P}_w^D} \tilde{\phi}_p = d_w^D, \quad w \in \mathcal{W}, \tag{B.23b}$$

$$\tilde{\boldsymbol{\phi}} \geq \mathbf{0}, \tag{B.23c}$$

where ∇ is the gradient operator and the pth (for all $p \in \mathcal{P}$) element in $\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)$ is $\frac{\partial g(\boldsymbol{\phi}, \boldsymbol{\theta})}{\partial \phi_p} = \sum_{s \in \mathcal{S}_p^D} \alpha_s(\tilde{\mathbf{x}}) v_s + \sum_{s \in \mathcal{S}_p^D} T_s$. Following Theorem 1 of Aghassi et al. (2006), it can be shown using linear programming duality that a candidate solution $\boldsymbol{\phi}$ solves (B.23) if and only if there exists a vector $\mathbf{b} \in \mathbb{R}^{|\mathcal{W}|}$ such that

$$\sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w^D\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right), \quad p \in \mathcal{P}, \tag{B.24a}$$

$$\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)^{\top} \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^{\top} \mathbf{b} = 0. \tag{B.24b}$$

Therefore, $(\boldsymbol{\phi},\ \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$ if and only if $(\boldsymbol{\phi},\ \mathbf{v},\ \mathbf{b})$ satisfies (B.20b), (B.20c), and (B.24). The result follows because WLS also minimizes $\|\ \mathbf{v} - \ \mathbf{z}\|_2^2$ over Θ in addition to satisfying $(\boldsymbol{\phi},\ \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$. $\square$

Proof of Proposition 6. We assume $\mathbf{x}$ and $\boldsymbol{\theta}$ are fixed and suppress dependence on them throughout the proof. It is straightforward to verify that $S(\mathbf{x}, \mathbf{d},\ \mathbf{v})$ is strictly convex in $(\mathbf{d},\ \mathbf{v})$. Uniqueness of $(\mathbf{d},\ \mathbf{v})$ them follows from standard arguments using strict convexity of the optimization problem (3.25). In the remainder of the proof, we show that $(\mathbf{d}, \boldsymbol{\phi},\ \mathbf{v}) \in \Gamma(\mathbf{x})$ if and only if $(\mathbf{d}, \boldsymbol{\phi},\ \mathbf{v})$ solves (3.25). Using $v_s = \sum_{\left\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\right\}} \phi_p,\ s \in \mathcal{S}$ from (3.1),

the user equilibrium problem (3.25) can be rewritten as

$$\underset{\boldsymbol{\phi}\geq 0,\mathbf{d}}{\text{minimize}} \;\; S'(\mathbf{d},\boldsymbol{\phi}) \tag{B.25a}$$

$$\text{subject to} \;\; \tilde{d}_w = \sum_{m\in\mathcal{M}} d_w^m, \quad w\in\mathcal{W}, \tag{B.25b}$$

$$d_w^D - \sum_{p\in\mathcal{P}_w} \phi_p^D = 0, \quad w\in\mathcal{W}, \tag{B.25c}$$

where

$$S'(\mathbf{d},\boldsymbol{\phi}) = \sum_{w\in\mathcal{W}} \left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D\right) + \beta_1^D \sum_{s\in\mathcal{S}} \left(\int_0^{\sum\left\{p\in\mathcal{P}^D | s\in\mathcal{S}_p^D\right\}\phi_p} t_s(v)\cdot dv \right) \tag{B.26a}$$

$$+ \sum_{w\in\mathcal{W}} \left(d_w^D \log\left(d_w^D\right) + d_w^C \log\left(d_w^C\right) + d_w^O \log\left(d_w^O\right)\right). \tag{B.26b}$$

Let λ_p, μ_w and γ_w be the Lagrangian multipliers for constraints $\phi_p \geq 0, p\in\mathcal{P}^D$, (B.25b) and (B.25c), respectively. We can then write the Lagrangian as

$$\mathcal{L}(\mathbf{d},\boldsymbol{\phi}) = S'(\mathbf{d},\boldsymbol{\phi}) - \sum_{p\in\mathcal{P}^D} \lambda_p\phi_p + \sum_{w\in\mathcal{W}} \left(\mu_w \left(d_w^O + d_w^C + d_w^D - \tilde{d}_w \right) + \gamma_w \left(d_w^D - \sum_{p\in\mathcal{P}_w^D} \phi_p^D \right) \right). \tag{B.27}$$

It is straightforward to verify that $S'(\mathbf{d},\boldsymbol{\phi})$ is convex in $(\mathbf{d},\boldsymbol{\phi})$ and that $d_w^m > 0$ for all $m\in\mathcal{M}$. It follows that a solution $(\mathbf{d},\boldsymbol{\phi},\mathbf{v})$ is optimal to problem (3.25) if and only if there exist $\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\gamma}$ that satisfy the Karush-Kuhn Tucker conditions of (3.25), i.e., the primal feasibility constraints $\phi_p\geq 0$, $p\in\mathcal{P}^D$, (B.25b) and (B.25c), the dual feasibility constraints $\lambda_p\geq 0$, $p\in\mathcal{P}^D$, complementary slackness conditions $\lambda_p\phi_p = 0$, $p\in\mathcal{P}^D$, and the stationarity conditions

$$\frac{\partial\mathcal{L}}{\partial d_w^O} = u_w^O + \log d_w^O + \mu_w = 0, \quad w\in\mathcal{W} \tag{B.28a}$$

$$\frac{\partial\mathcal{L}}{\partial d_w^C} = u_w^C + \log d_w^C + \mu_w = 0, \quad w\in\mathcal{W}, \tag{B.28b}$$

$$\frac{\partial\mathcal{L}}{\partial d_w^D} = \beta_0^D + \log d_w^D + \mu_w + \gamma_w = 0, \quad w\in\mathcal{W}, \tag{B.28c}$$

$$\frac{\partial\mathcal{L}}{\partial \phi_p} = \beta_1^D \sum_{s\in\mathcal{S}_p^D} t_s(v_s) - \sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_W^D\}} \gamma_w = \lambda_p, \quad p\in\mathcal{P}^D. \tag{B.28d}$$

Using the definition given in (3.5) that $t_p(\mathbf{v}) = \sum_{s \in \mathcal{S}_p^D} t_s(v_s)$, and using the dual feasibility and complementary slackness conditions, it is straightforward to show that (B.28d) is equivalent to

$$\frac{\partial \mathcal{L}}{\partial \phi_p} = \beta_1^D \cdot t_p(\mathbf{v}) - \sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_W^D\}} \gamma_w \begin{cases} = 0, & \text{if } \phi_p > 0, \\ \geq 0, & \text{if } \phi_p = 0, \end{cases} \quad p \in \mathcal{P}^D. \tag{B.29}$$

Next, using (3.6), it can be shown that $\gamma_w = \beta_1 \cdot t_w$ for all $w \in \mathcal{W}$ satisfies (B.29). Then by the definition of driving disutility u_w^D in (3.8), it follows that $u_w^D = \beta_0 + \gamma_w$. Next, note that solving (B.28a)–(B.28c) and (B.25b) yields the unique solution

$$d_w^m = \tilde{d}_w \frac{e^{-u_w^m}}{e^{-u_w^C} + e^{-(\beta_0+\gamma_w)} + e^{-u_w^O}}, \quad m \in \{C, O\}, w \in \mathcal{W}, \tag{B.30a}$$

$$d_w^D = \tilde{d}_w \frac{e^{-(\beta_0+\gamma_w)}}{e^{-u_w^C} + e^{-(\beta_0+\gamma_w)} + e^{-u_w^O}}, \quad w \in \mathcal{W}, \tag{B.30b}$$

which implies

$$d_w^m = \tilde{d}_w \frac{e^{-u_w^m}}{e^{-u_w^C} + e^{-u_w^D} + e^{-u_w^O}}, \quad m \in \mathcal{M}, w \in \mathcal{W}. \tag{B.31}$$

Note that $(\mathbf{d}^*, \boldsymbol{\phi}^*)$ is optimal to (B.25) if and only if $(\mathbf{d}^*, \boldsymbol{\phi}^*, \mathbf{v}^*)$ is optimal to (3.25), where $v_s^* = \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p^*$, for all $s \in \mathcal{S}$ from (3.1b). Then observing the equivalence between (B.29) and (3.6), and between (B.31) and (3.12), we conclude that $(\mathbf{d}^*, \boldsymbol{\phi}^*, \ \mathbf{v}^*)$ is an optimal solution to problem (3.25) if and only if $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x})$. □

Proof of Theorem 3. The proof proceeds in two steps. First, we show that for any $\mathbf{x}$, $L_n(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$. Second, we prove the main result. For conciseness, we assume $\mathbf{x}$ to be fixed and suppress dependence on it throughout. Step 1. By definition, $L_n(\boldsymbol{\theta}) = \| \mathbf{v} - \mathbf{z}\|_2^2$ where $\mathbf{v}$ is an equilibrium segment flow under $\boldsymbol{\theta}$. Therefore, to show continuity of $L_n(\boldsymbol{\theta})$, it suffices to show there is a unique equilibrium segment flow $\mathbf{v}^*(\boldsymbol{\theta})$ that is also continuous in $\boldsymbol{\theta}$. Uniqueness of $\mathbf{v}^*(\boldsymbol{\theta})$ is a straightforward extension of a classical result regarding Wardrop equilibrium (Hall 1978). Next, we prove continuity by following the argument for Proposition 3.1 in Cominetti et al. (2021). First, define $\tau_s(\boldsymbol{\theta}) := t_s(\ \mathbf{v}^*(\boldsymbol{\theta}), \boldsymbol{\theta})$

to be the equilibrium travel time on segment s, and let $\boldsymbol{\tau}(\boldsymbol{\theta})$ and $\boldsymbol{t}(\boldsymbol{\theta})$ be the vectors of τ_s and t_s, respectively. Next, from the proof of Proposition 5, note that $\mathbf{v}^*(\boldsymbol{\theta})$ is attained at an optimal solution to the convex optimization problem (B.20). Let the objective function of (B.20) be defined as:

$$f(\cdot) := \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot v_s^2 + T_s \cdot v_s \right) \tag{B.32}$$

Next, following Fukushima (1984), for any fixed $\boldsymbol{\theta}$, $\boldsymbol{\tau}(\boldsymbol{\theta})$ is the unique solution to the following strictly convex dual program:

$$\min_{\mathbf{t}} \; f^*(\mathbf{t}, \boldsymbol{\theta}) - \sum_{w \in \mathcal{W}} \left(d_w^D \cdot \min_{p \in \mathcal{P}_w} \sum_{s \in \mathcal{S}_p^D} t_s(\boldsymbol{\theta}) \right) \tag{B.33}$$

where $f^*(\cdot)$ is the conjugate of $f(\cdot)$ and is strictly convex. By Berge's maximum theorem, because (B.33) is jointly continuous in $(\mathbf{t}, \boldsymbol{\theta})$, it follows that $\boldsymbol{\tau}(\boldsymbol{\theta})$ is upper hemicontinuous in $\boldsymbol{\theta}$ (Charalambos and Aliprantis 2013). Because $\boldsymbol{\tau}(\boldsymbol{\theta})$ is unique, it is then continuous in $\boldsymbol{\theta}$. Next, because $t_s(\mathbf{v})$ is strictly increasing and continuous in v_s for all $s \in \mathcal{S}$, the inverse $\mathbf{v}^*(\boldsymbol{\theta}) = \mathbf{t}^{-1}(\boldsymbol{\tau}(\boldsymbol{\theta}))$ is also continuous in $\boldsymbol{\theta}$. It follows that $L_n(\boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$. Step 2. Because $L_n(\boldsymbol{\theta})$, By Lemma 1 of Wu (1981), it suffices to show that for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, $L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0) \longrightarrow \infty$ as $n \longrightarrow \infty$. Note that we can write the loss function more simply as $L_n(\boldsymbol{\theta}) = \| \mathbf{v}^*(\boldsymbol{\theta}) - \mathbf{z} \|_2^2$. Further, note that for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, we have

$$\begin{aligned}
L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0) &= \| \mathbf{v}(\boldsymbol{\theta}) - \mathbf{z} \|_2^2 - \| \mathbf{v}(\boldsymbol{\theta}_0) - \mathbf{z} \|_2^2 && \text{(B.34a)} \\
&= \sum_{s=1}^{n} (v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0) - \varepsilon_s)^2 - \sum_{s=1}^{n} (v_s(\boldsymbol{\theta}_0) - v_s(\boldsymbol{\theta}_0) - \varepsilon_s)^2 && \text{(B.34b)} \\
&= \sum_{s=1}^{n} (v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0))^2 - 2 \sum_{s=1}^{n} \varepsilon_s (v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0)), && \text{(B.34c)}
\end{aligned}$$

where the final inequality follows from algebraic simplification. Because the errors ε_s are drawn i.i.d. from a distribution with $\mathbb{E}[\varepsilon] = 0$, and $v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0)$ is bounded from above and below for all $s = 1, \ldots, n$, it follows from the weak law of large numbers that $\sum_{s=1}^{n} \varepsilon_s (v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0)) \longrightarrow 0$ as $n \longrightarrow \infty$. Next, because $\sum_{s=1}^{n} (v_s(\boldsymbol{\theta}) - v_s(\boldsymbol{\theta}_0))^2 \longrightarrow \infty$ as $n \longrightarrow \infty$ by

condition (ii), it follows that $L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0) \longrightarrow \infty$. The result follows. $\square$

Proof of Theorem 4. We first introduce some additional notation. Let

$$\boldsymbol{\gamma} = \begin{bmatrix} \mathbf{d} \\ \mathbf{v} \end{bmatrix} \tag{B.35}$$

be the stacked vector of $\mathbf{d}$ and $\mathbf{v}$. Further, for any $\mathbf{x}$, let

$$\boldsymbol{\gamma}^*(\mathbf{x}) = \begin{bmatrix} \mathbf{d}^*(\mathbf{x}) \\ \mathbf{v}^*(\mathbf{x}) \end{bmatrix}, \quad \tilde{\boldsymbol{\gamma}}(\mathbf{x}) = \begin{bmatrix} \tilde{\mathbf{d}}(\mathbf{x}) \\ \tilde{\mathbf{v}}(\mathbf{x}) \end{bmatrix}, \tag{B.36}$$

where $(\mathbf{d}^*(\mathbf{x}), \mathbf{v}^*(\mathbf{x}))$ and $(\tilde{\mathbf{d}}(\mathbf{x}), \tilde{\mathbf{v}}(\mathbf{x}))$ are the optimal solutions of the exact user equilibrium problem (3.25) and the linearized user equilibrium problem (3.31), respectively. Let $\hat{S}(\mathbf{x}, \boldsymbol{\gamma}) = S(\mathbf{x}, \mathbf{d},\ \mathbf{v})$ and $\hat{C}(\mathbf{x}, \boldsymbol{\gamma}) = C(\mathbf{x}, \mathbf{d},\ \mathbf{v})$. Finally, let

$$e = \mu_1 \sqrt{|\mathcal{W}|} \left(\mu_2 \frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi| - 1} + \mu_3 \frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi| - 1} \right), \tag{B.37}$$

where $\mu_1 = \max\left\{\bar{d}, 1/(\beta_1^D \cdot \underline{\alpha})\right\}$, $\mu_2 = \beta_1^D \cdot \alpha_s(\mathbf{x}) \cdot (\bar{v} - \underline{v})$, and $\mu_3 = \log\left(\bar{d}\right) - \log\left(\underline{d}\right)$. The proof proceeds in three steps. First, we show that for any $\mathbf{x}$, $R\left(\mathbf{x}, \boldsymbol{\gamma}\right)$ is strongly monotone of modulus $\mu_1 = \max\left\{\bar{d}, \frac{1}{\beta_1^D \cdot \underline{\alpha}}\right\}$ in $\boldsymbol{\gamma}$. Second, we show that for any $\mathbf{x}$, $|\hat{C}\left(\mathbf{x}, \boldsymbol{\gamma}^*(\mathbf{x})\right) - \hat{C}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}(\mathbf{x}))| \leq e$. Lastly, we use the results from Step 1 and Step 2 to prove the main result.

Step 1. By definition of $R\left(\mathbf{x}, \boldsymbol{\gamma}\right)$, we have

$$R\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \nabla_{\boldsymbol{\gamma}} S\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \operatorname{diag}\left(\begin{bmatrix} \log\left(d_w^D\right) + \beta_0^D + 1 : w \in \mathcal{W} \\ \log\left(d_w^C\right) + u_w^C + 1 : w \in \mathcal{W} \\ \log\left(d_w^O\right) + u_w^O + 1 : w \in \mathcal{W} \\ \beta_1^D \left(\alpha_s(\mathbf{x}) \cdot v_s + T_s\right) : s \in \mathcal{S} \end{bmatrix} \right), \tag{B.38}$$

where $\operatorname{diag}\left(\mathbf{k}\right)$ is a diagonal matrix whose diagonal elements are vector $\mathbf{k}$. It then follows that

$$\nabla_{\boldsymbol{\gamma}} R\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \operatorname{diag}\left(\begin{bmatrix} 1/d_w^m : w \in \mathcal{W}, m \in \mathcal{M} \\ \beta_1^D \alpha_s(\mathbf{x}) : s \in \mathcal{S} \end{bmatrix} \right). \tag{B.39}$$

Note that for any fixed $\mathbf{x}$, because $d_w^m \leq \bar{d}\, \nabla_{\boldsymbol{\gamma}} R(\mathbf{x}, \boldsymbol{\gamma})$ is a positive definite diagonal matrix, with smallest possible eigenvalues $\min\{1/\bar{d}, \beta_1^D \cdot \underline{\alpha}\}$. Therefore, $R(\mathbf{x}, \boldsymbol{\gamma})$ is strongly monotone of modulus $\max\left\{\bar{d}, 1/\left(\beta_1^D \cdot \underline{\alpha}\right)\right\}$ in $\boldsymbol{\gamma}$ (cf. Proposition 8, Dan and Marcotte (2019)).

Step 2. From (3.28), $\hat{S}(\mathbf{x}, \boldsymbol{\gamma})$ can be written as

$$\hat{S}(\mathbf{x}, \boldsymbol{\gamma}) = \underbrace{\beta_1^D \sum_{s \in \mathcal{S}} (T_s \cdot v_s) + \beta_1^D \sum_{s \in \mathcal{S}} \xi_s(v_s)}_{S_1(\mathbf{x}, \boldsymbol{\gamma})} + \underbrace{\sum_{w \in \mathcal{W}} \left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D\right) + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \psi(d_w^m)}_{S_2(\mathbf{x}, \boldsymbol{\gamma})}. \tag{B.40}$$

Let $R_1(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} S_1(\mathbf{x}, \boldsymbol{\gamma})$ and $R_2(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} S_2(\mathbf{x}, \boldsymbol{\gamma})$. Let $\tilde{S}_1$, $\tilde{S}_2$, and $\tilde{S}$ be the piecewise linear approximations of S_1, S_2 and $\hat{S}$, respectively, and let $\tilde{R}_1(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} \tilde{S}_1(\mathbf{x}, \boldsymbol{\gamma})$, $\tilde{R}_2(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} \tilde{S}_2(\mathbf{x}, \boldsymbol{\gamma})$, and $\tilde{R}(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} \tilde{S}(\mathbf{x}, \boldsymbol{\gamma})$. Because $\hat{S}(\mathbf{x}, \boldsymbol{\gamma})$ and $\tilde{S}(\mathbf{x}, \boldsymbol{\gamma})$ are both convex in $\boldsymbol{\gamma}$, it follows that

$$\langle -R(\mathbf{x}, \boldsymbol{\gamma}^*), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \rangle \geq 0, \tag{B.41a}$$

$$\langle \tilde{R}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \rangle \geq 0, \tag{B.41b}$$

where $\langle \cdot \rangle$ is the Euclidean inner product operator. By summing the two inequalities above, we obtain

$$\langle \tilde{R}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}) - R(\mathbf{x}, \boldsymbol{\gamma}^*), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \rangle \geq 0. \tag{B.42}$$

Next, by Step 1 and the definition of strong monotonicity, it follows that

$$\langle R(\mathbf{x}, \boldsymbol{\gamma}^*) - R(\mathbf{x}, \tilde{\boldsymbol{\gamma}}), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \rangle \geq \frac{1}{\mu_1} \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|^2. \tag{B.43}$$

Combining (B.42) and (B.43) yields

$$\langle \tilde{R}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}) - R(\mathbf{x}, \tilde{\boldsymbol{\gamma}}), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \rangle \geq \frac{1}{\mu_1} \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|^2. \tag{B.44}$$

Applying the Cauchy-Schwarz inequality, we have

$$\mu_1 \|\tilde{R}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}) - R(\mathbf{x}, \tilde{\boldsymbol{\gamma}})\| \geq \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|. \tag{B.45}$$

Next, using the definition of $\hat{C}(\mathbf{x}, \boldsymbol{\gamma})$ and the Cauchy–Schwarz inequality, we have

$$\left|\hat{C}(\mathbf{x}, \boldsymbol{\gamma}^*) - \hat{C}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right)\right| = \left|\sum_{w \in \mathcal{W}} \left(d_w^{C*}(\mathbf{x}) - \tilde{d}_w^C(\mathbf{x})\right)\right| \tag{B.46a}$$

$$\leq \sqrt{|\mathcal{W}|} \cdot \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|. \tag{B.46b}$$

It follows from (B.45) and (B.46b) and that

$$\left|C\left(\mathbf{x}, \boldsymbol{\gamma}^*\right) - C\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right)\right| \leq \sqrt{|\mathcal{W}|} \cdot \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\| \tag{B.47a}$$

$$\leq \mu_1 \sqrt{|\mathcal{W}|} \cdot \|\tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right)\|. \tag{B.47b}$$

Next, we bound the right hand side of (B.46b). We do so by first decomposing the term $\|\tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right)\|$ using R_1, R_2, $\tilde{R}_1$, and $\tilde{R}_1$. Let $\tilde{\xi}$ and $\tilde{\psi}$ be the piecewise linear approximations of ξ and ψ, respectively. Then

$$\tilde{S}_1(\mathbf{x}, \boldsymbol{\gamma}) = \beta_1^D \sum_{s \in \mathcal{S}} \left(T_s \cdot v_s\right) + \beta_1^D \sum_{s \in \mathcal{S}} \tilde{\xi}_s\left(v_s\right), \tag{B.48a}$$

$$\tilde{S}_2(\mathbf{x}, \boldsymbol{\gamma}) = \sum_{w \in \mathcal{W}} \left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D\right) + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \tilde{\psi}\left(d_w^m\right). \tag{B.48b}$$

By differentiating, we obtain

$$\tilde{R}_1(\mathbf{x}, \boldsymbol{\gamma}) - R_1(\mathbf{x}, \boldsymbol{\gamma}) = \operatorname{diag}\left(\begin{bmatrix} \mathbf{0}_{3|\mathcal{W}|} \\ \beta_1^D \left(\frac{\partial \tilde{\xi}_s(v_s)}{\partial v_s} - \alpha_s(\mathbf{x}) \cdot v_s\right) : s \in \mathcal{S} \end{bmatrix}\right), \tag{B.49a}$$

$$\tilde{R}_2(\mathbf{x}, \boldsymbol{\gamma}) - R_2(\mathbf{x}, \boldsymbol{\gamma}) = \operatorname{diag}\left(\begin{bmatrix} \frac{\partial \tilde{\psi}(d_w^m)}{\partial d_w^m} - \left(\log\left(d_w^m\right) + 1\right) : w \in \mathcal{W}, m \in \mathcal{M} \\ \mathbf{0}_{|\mathcal{S}|} \end{bmatrix}\right), \tag{B.49b}$$

where $\mathbf{0}_a$ is the $a \times 1$ vector of zeros. By construction, the linear segments in the piecewise linear function $\tilde{\xi}_s\left(\cdot\right)$ have slopes that are equidistant in the interval $\alpha_s\left(\mathbf{x}\right) \cdot [\underline{v}, \bar{v}]$, and the difference between two consecutive slope values is $\left(\beta_1^D \alpha_s(\mathbf{x}) \cdot \left(\bar{v} - \underline{v}\right) / \left(|\mathcal{R}_\xi| - 1\right)\right.$. It follows that

$$\left|\beta_1^D \left(\frac{\partial \tilde{\xi}_s\left(v_s\right)}{\partial v_s} - \alpha_s(\mathbf{x}) \cdot v_s\right)\right| \leq \mu_2 \cdot \frac{1}{|\mathcal{R}_\xi| - 1} \tag{B.50}$$

holds for all $s \in \mathcal{S}$, which implies

$$\|\tilde{R}_1(\mathbf{x},\boldsymbol{\gamma}) - R_1(\mathbf{x},\boldsymbol{\gamma})\| = \sqrt{\sum_{s\in\mathcal{S}} \left(\beta_1^D \left(\frac{\partial \tilde{\xi}_s\left(v_s\right)}{\partial v_s} - \alpha_s(\mathbf{x})\cdot v_s\right)\right)^2} \tag{B.51a}$$

$$\leq \mu_2 \cdot \frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi| - 1}. \tag{B.51b}$$

Similarly, for any $w \in \mathcal{W}$ and $m \in \mathcal{M}$, because the linear segments in the piecewise linear function $\tilde{\psi}\left(\cdot\right)$ have slopes that are equidistant in the interval $\left[1+\log(\underline{d}), 1+\log(\bar{d})\right]$, the difference between two consecutive slope values is $(\log\left(\bar{d}\right) - \log\left(\underline{d}\right))/(|\mathcal{R}_\psi| - 1)$. It follows that

$$\frac{\partial \tilde{\psi}\left(d_w^m\right)}{\partial d_w^m} - \left(\log\left(d_w^m\right) + 1\right) \leq \mu_3 \cdot \frac{1}{|\mathcal{R}_\psi| - 1} \tag{B.52}$$

holds for every $w \in \mathcal{W}$ and $m \in \mathcal{M}$, which implies

$$\left\|\tilde{R}_2(\mathbf{x},\boldsymbol{\gamma}) - R_2(\mathbf{x},\boldsymbol{\gamma})\right\| = \sqrt{\sum_{m\in\mathcal{M}}\sum_{w\in\mathcal{W}} \left(\frac{\partial \tilde{\psi}\left(d_w^m\right)}{\partial d_w^m} - \left(\log\left(d_w^m\right) + 1\right)\right)^2} \tag{B.53a}$$

$$\leq \mu_3 \cdot \frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi| - 1}. \tag{B.53b}$$

Combining inequalities (B.47b), (B.51b) and (B.53b), we obtain

$$\left|C\left(\mathbf{x},\boldsymbol{\gamma}^*\right) - C\left(\mathbf{x},\tilde{\boldsymbol{\gamma}}\right)\right| \leq \mu_1\sqrt{|\mathcal{W}|}\left\|\tilde{R}\left(\mathbf{x},\tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x},\tilde{\boldsymbol{\gamma}}\right)\right\| \tag{B.54a}$$

$$\leq \mu_1\sqrt{|\mathcal{W}|}\left(\left\|\tilde{R}_1(\mathbf{x},\tilde{\boldsymbol{\gamma}}) - R_1(\mathbf{x},\tilde{\boldsymbol{\gamma}})\right\| + \left\|\tilde{R}_2(\mathbf{x},\tilde{\boldsymbol{\gamma}}) - R_2(\mathbf{x},\tilde{\boldsymbol{\gamma}})\right\|\right) \tag{B.54b}$$

$$\leq \mu_1\sqrt{|\mathcal{W}|}\left(\mu_2 \cdot \frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi| - 1} + \mu_3 \cdot \frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi| - 1}\right) \tag{B.54c}$$

$$= e, \tag{B.54d}$$

as desired. Step 2. We now show $0 \leq C\left(\mathbf{x}^*, \mathbf{d}^*\left(\mathbf{x}^*\right), \mathbf{v}^*\left(\mathbf{x}^*\right)\right) - C\left(\tilde{\mathbf{x}}, \mathbf{d}^*\left(\tilde{\mathbf{x}}\right), \mathbf{v}^*\left(\tilde{\mathbf{x}}\right)\right) \leq 2e$. Note that from (B.54d) and the definitions of $\boldsymbol{\gamma}^*$ and $\tilde{\boldsymbol{\gamma}}$, we have

$$C(\tilde{\mathbf{x}}, \mathbf{d}^*\left(\tilde{\mathbf{x}}\right), \mathbf{v}^*(\tilde{\mathbf{x}}) + e \geq C(\tilde{\mathbf{x}}, \tilde{\mathbf{d}}\left(\tilde{\mathbf{x}}\right), \tilde{\mathbf{v}}\left(\tilde{\mathbf{x}}\right)), \tag{B.55a}$$

$$C(\mathbf{x}^*, \mathbf{d}^*\left(\mathbf{x}^*\right), \mathbf{v}^*\left(\mathbf{x}^*\right)) - e \leq C(\mathbf{x}^*, \tilde{\mathbf{d}}\left(\mathbf{x}^*\right), \tilde{\mathbf{v}}\left(\mathbf{x}^*\right)). \tag{B.55b}$$

Next, because $\mathbf{x}^*$ and $\tilde{\mathbf{x}}$ are the optimal solutions to BLP and BLP-A and $C(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is the objective function of both BLP and BLP-A, it follows that

$$C(\mathbf{x}^*, \tilde{\mathbf{d}}(\mathbf{x}^*), \tilde{\mathbf{v}}(\mathbf{x}^*)) \leq C(\tilde{\mathbf{x}}, \tilde{\mathbf{d}}(\tilde{\mathbf{x}}), \tilde{\mathbf{v}}(\tilde{\mathbf{x}})), \tag{B.56a}$$

$$C(\tilde{\mathbf{x}}, \mathbf{d}^*(\tilde{\mathbf{x}}), \mathbf{v}^*(\tilde{\mathbf{x}})) \leq C(\mathbf{x}^*, \mathbf{d}^*(\mathbf{x}^*), \mathbf{v}^*(\mathbf{x}^*)). \tag{B.56b}$$

Combining (B.55) and (B.56), we have

$$0 \leq C(\mathbf{x}^*, \mathbf{d}^*(\mathbf{x}^*), \mathbf{v}^*(\mathbf{x}^*)) - C(\tilde{\mathbf{x}}, \mathbf{d}^*(\tilde{\mathbf{x}}), \mathbf{v}^*(\tilde{\mathbf{x}})) \leq 2e, \tag{B.57}$$

which concludes the proof. □

Bibliography

117th Congress. 2021. Infrastructure Investment and Jobs Act, H.R. 3684. URL https://www.epw.senate.gov/public/_cache/files/e/a/ea1eb2e4-56bd-45f1-a260-9d6ee951bc96/F8A7C77D69BE09151F210EB4DFE872CD.edw21a09.pdf.

Abdulaal, Mustafa, Larry J LeBlanc. 1979. Methods for combining modal split and equilibrium assignment models. Transportation Science 13(4) 292–314.

Aghassi, Michele, Dimitris Bertsimas, Georgia Perakis. 2006. Solving asymmetric variational inequalities via convex optimization. Operations Research Letters 34(5) 481–490.

Ahuja, Ravindra K, James B Orlin. 2001. Inverse optimization. Operations Research 49(5) 771–783.

Akkaya, Duygu, Kostas Bimpikis, Hau Lee. 2019. Government interventions in promoting sustainable practices in agriculture. Manufacturing & Service Operations Management .

Alizamir, Saed, Francis de Véricourt, Peng Sun. 2016. Efficient feed-in-tariff policies for renewable energy technologies. Operations Research 64(1) 52–66.

Alizamir, Saed, Foad Iravani, Hamed Mamani. 2019. An analysis of price vs. revenue protection: Government subsidies in the agriculture industry. Management Science 65(1) 32–49.

Allen, Stephanie, John P Dickerson, Steven A Gabriel. 2021. Using inverse optimization to learn cost functions in generalized Nash games. arXiv preprint arXiv:2102.12415 .

American Automobile Association. 2018. Your driving costs: how much are you really paying to drive.

Anandalingam, G. 1987. Asymmetric players and bargaining for profit shares in natural resource development. Management Science 33(8) 1048–1057.

Anderson, Michael L. 2014. Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. American Economic Review 104(9) 2763–96.

APTA. 2019a. Apta public transportation fact book. URL https://www.apta.com/wp-content/uploads/APTA_Fact-Book-2019_FINAL.pdf.

APTA. 2019b. First last/mile solutions. URL https://www.apta.com/research-technical-resources/mobility-innovation-hub/first-last-mile-solutions/.

APTA. 2019c. Public transportation facts. URL https://www.apta.com/news-publications/public-transportation-facts/.

APTA. 2019d. Transit Agency and Urbanized Area Operating Statistics.

APTA. 2019e. Transit and TNC partnerships. URL https://www.apta.com/research-technical-resources/mobility-innovation-hub/transit-and-tnc-partnerships/.

Arifoglu, Kenan, Christopher S Tang. 2019. A two-sided budget-neutral incentive program for coordinating an influenza vaccine supply chain with endogenous supply and demand under uncertainty. Forthcoming, Manufacturing & Service Operations Management .

Aswani, Anil, Zuo-Jun Shen, Auyon Siddiq. 2018. Inverse optimization with noisy data. Operations Research 66(3) 870–892.

Aswani, Anil, Zuo-Jun Max Shen, Auyon Siddiq. 2019. Data-driven incentive design in the medicare shared savings program. Operations Research 67(4) 1002–1026.

Badstuber, Nicole. 2018. London's congestion charge is showing its age. URL https://www.citylab.com/transportation/2018/04/londons-congestion-charge-needs-updating/557699/.

Bakshi, Nitin, Noah Gans. 2010. Securing the containerized supply chain: analysis of government incentives for private investment. Management Science 56(2) 219–233.

Bansal, Sangeeta, Shubhashis Gangopadhyay. 2003. Tax/subsidy policies in the presence of environmentally aware consumers. Journal of Environmental Economics and Management 45(2) 333–355.

Bao, Jie, Tianfu He, Sijie Ruan, Yanhua Li, Yu Zheng. 2017. Planning bike lanes based on sharing-bikes' trajectories. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1377–1386.

BBC. 2021. London congestion: Cycle lanes blamed as city named most congested. URL https://www.bbc.com/news/uk-england-london-59559863.

Beckmann, Martin, Charles B McGuire, Christopher B Winsten. 1956. Studies in the economics of transportation. Tech. rep.

Beltran, Borja, Stefano Carrese, Ernesto Cipriani, Marco Petrelli. 2009. Transit network design

with allocation of green vehicles: A genetic algorithm approach. Transportation Research Part C: Emerging Technologies 17(5) 475–483.

Ben-Akiva, Moshe, Michel Bierlaire. 1999. Discrete choice methods and their applications to short term travel decisions. Handbook of transportation science. Springer, 5–33.

Bertinelli, Luisito, Duncan Black. 2004. Urbanization and growth. Journal of Urban Economics 56(1) 80–96.

Bertsekas, Dimitri P. 1997. Nonlinear programming. Journal of the Operational Research Society 48(3) 267.

Bertsimas, Dimitris, Vishal Gupta, Ioannis Paschalidis. 2015. Data-driven estimation in equilibrium using inverse optimization. Mathematical Programming 153(2) 595–633.

Bertsimas, Dimitris, Yee Sian Ng, Julia Yan. 2020. Joint frequency-setting and pricing optimization on multimodal transit networks at scale. Transportation Science 54(3) 839–853.

Bickel, Peter J, Kjell A Doksum. 2015. Mathematical statistics: Basic ideas and selected topics. CRC Press.

Broach, Joseph, Jennifer Dill, John Gliebe. 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transportation Research Part A: Policy and Practice 46(10) 1730–1740.

Brown, Eliot. 2020. The ride-hail utopia that got stuck in traffic. URL `https://www.wsj.com/articles/the-ride-hail-utopia-that-got-stuck-in-traffic-11581742802`.

Bureau of Transportation Statistics. 2018. Transportation statistics annual report. URL `https://www.bts.dot.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-and-data/transportation-statistics-annual-reports/Preliminary-TSAR-Full-2018-a.pdf`.

Burkhauser, Richard V, Timothy M Smeeding. 1994. Social security reform: A budget neutral approach to reducing older women's disproportional risk of poverty. Available at SSRN 1824592 .

Burton, Didier, Ph L Toint. 1992. On an instance of the inverse shortest paths problem. Mathematical programming 53(1) 45–61.

Chan, Timothy CY, Maria Eberg, Katharina Forster, Claire Holloway, Luciano Ieraci, Yusuf Shalaby, Nasrin Yousefi. 2021a. An inverse optimization approach to measuring clinical pathway concordance. Management Science .

Chan, Timothy CY, Taewoo Lee, Daria Terekhov. 2019. Inverse optimization: Closed-form solutions, geometry, and goodness of fit. Management Science 65(3) 1115–1135.

Chan, Timothy CY, Rafid Mahmood, Ian Yihang Zhu. 2021b. Inverse optimization: Theory and applications. arXiv preprint arXiv:2109.03920 .

Charalambos, D, Border Aliprantis. 2013. Infinite Dimensional Analysis: A Hitchhiker's Guide. Springer.

Chemama, Jonathan, Maxime C Cohen, Ruben Lobel, Georgia Perakis. 2019. Consumer subsidies with a strategic supplier: Commitment vs. flexibility. Management Science 65(2) 681–713.

Chen, Linxi, Hai Yang. 2012. Managing congestion and emissions in road networks with tolls and rebates. Transportation Research Part B: Methodological 46(8) 933–948.

Chicago Data Portal. 2010. Boundaries - Census tracts - 2010. URL https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik.

Chicago Data Portal. 2018. Bike Routes (Deprecated November 2018). URL https://data.cityofchicago.org/Transportation/Bike-Routes-Deprecated-November-2018-/pznz-m9ui.

Chicago Data Portal. 2021. Taxi trips. URL https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew.

Chicago DOT. 2020. Chicago streets for cycling plan 2020. URL https://www.chicago.gov/content/dam/city/depts/cdot/bike/general/ChicagoStreetsforCycling2020.pdf.

Chicago DOT. 2021. Chicago community cycling network update. URL https://www.chicago.gov/content/dam/city/depts/cdot/bike/2021/Chicago%20Community%20Cycling_2021-09-21.pdf.

Cho, Soo-Haeng, Xin Fang, Sridhar Tayur, Ying Xu. 2019. Combating child labor: Incentives

and information disclosure in global supply chains. Manufacturing & Service Operations Management 21(3) 692–711.

Chow, Joseph YJ, Stephen G Ritchie, Kyungsoo Jeong. 2014. Nonlinear inverse optimization for parameter estimation of commodity-vehicle-decoupled freight assignment. Transportation Research Part E: Logistics and Transportation Review 67 71–91.

Cipriani, Ernesto, Marco Petrelli, Gaetano Fusco. 2006. A multimodal transit network design procedure for urban areas. Advances in Transportation Studies 10.

CMS. 2018. Budget neutrality policies for section 1115(a) medicaid demonstration projects. URL https://www.medicaid.gov/sites/default/files/federal-policy-guidance/downloads/smd18009.pdf.

Cohen, Morris A, Shiliang Cui, Ricardo Ernst, Arnd Huchzermeier, Panos Kouvelis, Hau L Lee, Hirofumi Matsuo, Marc Steuber, Andy A Tsay. 2018. Om forum—benchmarking global production sourcing decisions: Where and why firms offshore and reshore. Manufacturing and Service Operations Management 20(3) 389–402.

Cohen, Morris A, Hau L Lee. 2020. Designing the right global supply chain network. Manufacturing & Service Operations Management .

Çolak, Serdar, Antonio Lima, Marta C González. 2016. Understanding congested travel in urban areas. Nature Communications 7(1) 1–8.

Cole, Matt. 2017. Mobility as a service: Putting transit front and center of the conversation. Cubic White Paper .

Cominetti, Roberto, Valerio Dose, Marco Scarsini. 2021. The price of anarchy in routing games as a function of the demand. Mathematical Programming 1–28.

Condor, Bob. 2020. Get me to the game on time. URL https://www.nhl.com/kraken/news/nhl-seattle-announces-monorail-partnership-and-free-public-transit/c-315511164.

Correia, Isabel, Emmanuel Farhi, Juan Pablo Nicolini, Pedro Teles. 2013. Unconventional fiscal policy at the zero bound. American Economic Review 103(4) 1172–1211.

Corselli, Andrew. 2020. Cubic reveals public transit loyalty program. URL https://www.railwayage.com/passenger/cubic-reveals-public-transit-loyalty-program/.

Croci, Edoardo. 2016. Urban road pricing: a comparative study on the experiences of london, stockholm and milan. Transportation Research Procedia 14 253–262.

Currie, Graham. 2018. Lies, damned lies, avs, shared mobility, and urban transit futures. Journal of Public Transportation 21(1) 3.

Dafermos, Stella. 1980. Traffic equilibrium and variational inequalities. Transportation Science 14(1) 42–54.

Dafermos, Stella. 1982. The general multimodal network equilibrium problem with elastic demand. Networks 12(1) 57–72.

Dan, Teodora, Patrice Marcotte. 2019. Competitive facility location with selfish users and queues. Operations Research 67(2) 479–497.

Dantzig, George B, Roy P Harvey, Zachary F Lansdowne, David W Robinson, Steven F Maier. 1979. Formulating and solving the network design problem by decomposition. Transportation Research Part B: Methodological 13(1) 5–17.

Dempe, Stephan. 2002. Foundations of bilevel programming. Springer Science & Business Media.

Development Asia. 2018. The case for electronic road pricing. URL `https://development.asia/case-study/case-electronic-road-pricing`.

Divvy. 2021. Download divvy trip history data. URL `https://divvy-tripdata.s3.amazonaws.com/index.html`.

Domencich, Thomas A, Daniel McFadden. 1975. Urban travel demand-a behavioral analysis. Tech. rep.

DOT, NYC. 2019. Green wave: A plan for cycling in New York. URL `https://www1.nyc.gov/html/dot/downloads/pdf/bike-safety-plan.pdf`.

D'Acunto, Francesco, Daniel Hoang, Michael Weber. 2016. The effect of unconventional fiscal policy on consumption expenditure. Tech. rep., National Bureau of Economic Research.

Edwards, Tom. 2018. What's to blame for a fall in passenger numbers? URL `https://www.bbc.com/news/uk-england-london-42622891`.

Esfahani, Peyman Mohajerin, Soroosh Shafieezadeh-Abadeh, Grani A Hanasusanto, Daniel Kuhn.

2018. Data-driven inverse optimization with imperfect information. Mathematical Programming 167(1) 191–234.

Fan, Wei, Randy B Machemehl. 2006. Optimal transit route network design problem with variable transit demand: genetic algorithm approach. Journal of Transportation Engineering 132(1) 40–51.

Farahani, Reza Zanjirani, Elnaz Miandoabchi, Wai Yuen Szeto, Hannaneh Rashidi. 2013. A review of urban transportation network design problems. European Journal of Operational Research 229(2) 281–302.

Federal Highway Administration. 2017. National household travel survey.

Federal Highway Administration (FHA). 2006. Federal Highway Administration university course on bicycle and pedestrian transportation. URL https://www.fhwa.dot.gov/publications/research/safety/pedbike/05085/chapt15.cfm.

Federal Transit Administration. 2017. NTD 2017 Data Tables.

Federal Transit Administration. 2020. 2019 annual database revenue sources. URL https://www.transit.dot.gov/ntd/data-product/2019-annual-database-revenue-sources.

FHA. 2018. HPMS public release. URL https://www.fhwa.dot.gov/policyinformation/hpms/shapefiles.cfm.

Fisk, Caroline. 1980. Some developments in equilibrium traffic assignment. Transportation Research Part B: Methodological 14(3) 243–255.

Fortuny-Amat, José, Bruce McCarl. 1981. A representation and economic interpretation of a two-level programming problem. The Journal of the Operational Research Society 32(9) 783–792.

Fukushima, Masao. 1984. On the dual approach to the traffic assignment problem. Transportation Research Part B: Methodological 18(3) 235–245.

Gagnepain, Philippe, Marc Ivaldi. 2002. Incentive regulatory policies: the case of public transit systems in france. RAND Journal of Economics 605–629.

Gallo, Mariano, Bruno Montella, Luca D'Acierno. 2011. The transit network design problem with elastic demand and internalisation of external costs: An application to rail frequency optimisation. Transportation Research Part C: Emerging Technologies 19(6) 1276–1305.

Goodall, Warwick, Tiffany Dovey, Justine Bornstein, Brett Bonthron. 2017. The rise of mobility as a service. Deloitte Rev 20 112–129.

Google Maps. 2021. The Directions API overview. URL `https://developers.google.com/maps/documentation/directions/overview`.

Goulder, Lawrence H. 1995. Environmental taxation and the double dividend: a reader's guide. International tax and public finance 2(2) 157–183.

Guan, Peiqiu, Jing Zhang, Vineet M Payyappalli, Jun Zhuang. 2018. Modeling and validating public–private partnerships in disaster management. Decision Analysis 15(2) 55–71.

Guan, Peiqiu, Jun Zhuang. 2015. Modeling public–private partnerships in disaster management via centralized and decentralized models. Decision Analysis 12(4) 173–189.

Guo, Pengfei, Robin Lindsey, Zhe George Zhang. 2014. On the downs–thomson paradox in a self-financing two-tier queuing system. Manufacturing & Service Operations Management 16(2) 315–322.

Guo, Pengfei, Christopher S Tang, Yulan Wang, Ming Zhao. 2019. The impact of reimbursement policy on social welfare, revisit rate, and waiting time in a public healthcare system: Fee-for-service versus bundled payment. Manufacturing & Service Operations Management 21(1) 154–170.

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

Gupta, Diwakar, Mili Mehrotra. 2015. Bundled payments for healthcare services: Proposer selection and information sharing. Operations Research 63(4) 772–788.

Gur, Yonatan, Lijian Lu, Gabriel Y Weintraub. 2017. Framework agreements in procurement: An auction model and design recommendations. Manufacturing & Service Operations Management 19(4) 586–603.

Hall, Michael A. 1978. Properties of the equilibrium state in transportation networks. Transportation Science 12(3) 208–216.

Hamilton, Timothy L, Casey J Wichman. 2018. Bicycle infrastructure and traffic congestion:

Evidence from dc's capital bikeshare. Journal of Environmental Economics and Management 87 72–93.

Hansson, Lisa. 2010. Solving procurement problems in public transport: Examining multi-principal roles in relation to effective control mechanisms. Research in Transportation Economics 29(1) 124–132.

He, Pu, Fanyin Zheng, Elena Belavina, Karan Girotra. 2021. Customer preference and station network in the London bike-share system. Management Science 67(3) 1392–1412.

Hearn, Donald W, Motakuri V Ramana. 1998. Solving congestion toll pricing models. Equilibrium and Advanced Transportation Modelling. Springer, 109–124.

Hodges, Tina, et al. 2010. Public transportation's role in responding to climate change. United States. Federal Transit Administration .

Homonoff, Tatiana A. 2018. Can small incentives have large effects? the impact of taxes versus bonuses on disposable bag use. American Economic Journal: Economic Policy 10(4) 177–210.

Hong, Mingyi, Xiangfeng Wang, Meisam Razaviyayn, Zhi-Quan Luo. 2017. Iteration complexity analysis of block coordinate descent methods. Mathematical Programming 163(1-2) 85–114.

Hood, Jeffrey, Elizabeth Sall, Billy Charlton. 2011. A GPS-based bicycle route choice model for San Francisco, California. Transportation Letters 3(1) 63–75.

Illinois Department of Transportation. 2018. 2018 Illinois travel statistics.

Illinois DOT. 2018. URL https://www.idot.illinois.gov/Assets/uploads/files/Doing-Business/Manuals-Guides-&-Handbooks/Highways/Local-Roads-and-Streets/Local%20Roads%20and%20Streets%20Manual.pdf.

Inrix. 2020. Congestion costs each American nearly 100 hours, $1,400 a year. URL https://inrix.com/press-releases/2019-traffic-scorecard-us/.

Jennrich, Robert I. 1969. Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 40(2) 633–643.

Johnson, Gretchen, Aaron Johnson. 2014. Bike lanes don't cause traffic jams if you're smart about where you build them. URL https://fivethirtyeight.com/features/bike-lanes-dont-cause-traffic-jams-if-youre-smart-about-where-you-build-them/.

Jou, Rong-Chang, Yu-Chiun Chiou, Ke-Hong Chen, Hao-I Tan. 2012. Freeway drivers' willingness-to-pay for a distance-based toll rate. Transportation Research Part A: Policy and Practice 46(3) 549–559.

Kabra, Ashish, Elena Belavina, Karan Girotra. 2020. Bike-share systems: Accessibility and availability. Management Science 66(9) 3803–3824.

Kamp, Jon. 2020. Cities Offer Free Buses in Bid to Boost Flagging Ridership. URL https://www.wsj.com/articles/cities-waive-bus-fares-in-bid-to-boost-ridership-11579007895.

Kang, Chao-Chung, Tsun-Siou Lee, Szu-Chi Huang. 2013. Royalty bargaining in public–private partnership projects: insights from a theoretic three-stage game auction model. Transportation research part E: logistics and transportation review 59 1–14.

Khany, Sam. 2022. Personal communication. Senior Transportation Planner, City of Vancouver .

Kirk, Robert S, William J Mallett. 2020. Highway and public transit funding issues. Congressional Research Service. URL https://crsreports.congress.gov/product/pdf/download/IF/IF10495/IF10495.pdf/.

LA Metro. 2016. Metro first/last mile.

LA Times. 2020. Bicycles have enjoyed a boom during the pandemic. Will it last as car traffic resumes? URL https://www.latimes.com/california/story/2020-06-25/bicycle-business-is-exploding-during-covid-19-will-it-last.

Lee, Donald KK, Stefanos A Zenios. 2012. An evidence-based incentive system for medicare's end-stage renal disease program. Management science 58(6) 1092–1105.

Lee, Young-Jae, Vukan R Vuchic. 2005. Transit network design with variable demand. Journal of Transportation Engineering 131(1) 1–10.

Leonard, Christian. 2019. Study: Irvine workers spend nearly $10,000 a year in commute. URL https://www.nbclosangeles.com/news/local/irvine-workers-spend-10000-commuting-565069612.html.

Levi, Retsef, Georgia Perakis, Gonzalo Romero. 2017. On the effectiveness of uniform subsidies in increasing market consumption. Management Science 63(1) 40–57.

Lin, Henry, Tim Roughgarden, Éva Tardos, Asher Walkover. 2011. Stronger bounds on Braess's paradox and the maximum latency of selfish routing. SIAM Journal on Discrete Mathematics 25(4) 1667–1686.

Linton, Joe. 2016. New map shows metro's 20,000 parking spaces, mostly free. URL https://la.streetsblog.org/2016/09/23/new-map-shows-metros-20000-parking-spaces-mostly-free/.

Liu, Haoxiang, WY Szeto, Jiancheng Long. 2019. Bike network design problem with a path-size logit-based equilibrium constraint: Formulation, global optimization, and matheuristic. Transportation research part E: logistics and transportation review 127 284–307.

Liu, Sheng, Zuo-Jun Max Shen, Xiang Ji. 2021. Urban bike lane planning with bike trajectories: Models, algorithms, and a real-world case study. Manufacturing & Service Operations Management .

Lodi, Andrea, Enrico Malaguti, Nicolás E Stier-Moses, Tommaso Bonino. 2016. Design and control of public-service contracts and an application to public transportation systems. Management Science 62(4) 1165–1187.

Los Angeles County Metropolitan Transportation Authority. 2019. La metro launches partnership with via to provide on-demand service to three busy transit stations. URL https://www.metro.net/news/simple_pr/la-metro-launches-partnership-provide-demand-servi/.

Lou, Yin, Chengyang Zhang, Yu Zheng, Xing Xie, Wei Wang, Yan Huang. 2009. Map-matching for low-sampling-rate GPS trajectories. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. GIS '09, Association for Computing Machinery, New York, NY, USA, 352–361. doi:10.1145/1653771.1653820.

Lyft, Inc. 2021. Los angeles area ride costs. URL https://www.lyft.com/pricing/LAX.

Ma, Guangrui, Michael K Lim, Ho-Yin Mak, Zhixi Wan. 2019. Promoting clean technology adoption: To subsidize products or service infrastructure? Service Science 11(2) 75–95.

Magnanti, Thomas L, Richard T Wong. 1984. Network design and transportation planning: Models and algorithms. Transportation Science 18(1) 1–55.

Mallett, William J. 2018. Trends in public transportation ridership: implications for federal policy. Washington, DC: Congressional Research Service .

Marshall, Wesley E, Nicholas N Ferenchak. 2019. Why cities with high bicycling rates are safer for all road users. Journal of Transport & Health 13 100539.

Mauttone, Antonio, Gonzalo Mercadante, María Rabaza, Fernanda Toledo. 2017. Bicycle network design: Model and solution algorithm. Transportation Research Procedia 27 969–976.

McFadden, Daniel, et al. 1973. Conditional logit analysis of qualitative choice behavior .

Miandoabchi, Elnaz, Reza Zanjirani Farahani, Wout Dullaert, Wai Yuen Szeto. 2012. Hybrid evolutionary metaheuristics for concurrent multi-objective design of urban road and public transit networks. Networks and Spatial Economics 12(3) 441–480.

Mills, Kevin. 2021. Analysis: Bipartisan infrastructure bill passes with new opportunities for trails, walking and biking. URL `https://www.railstotrails.org/trailblog/2021/november/06/analysis-bipartisan-infrastructure-bill-passes-with-new-opportunities-for-trails-walking-and-biking/`.

Mitra, Raktim, Raymond A Ziemba, Paul M Hess. 2017. Mode substitution effect of urban cycle tracks: Case study of a downtown street in Toronto, Canada. International Journal of Sustainable Transportation 11(4) 248–256.

Monsere, Christopher, Jennifer Dill, Nathan McNeil, Kelly J Clifton, Nick Foster, Tara Goddard, Mathew Berkow, Joe Gilpin, Kim Voros, Drusilla van Hengel, et al. 2014. Lessons from the green lanes: Evaluating protected bike lanes in the US .

Murray, Brian, Nicholas Rivers. 2015. British columbia's revenue-neutral carbon tax: A review of the latest "grand experiment" in environmental policy. Energy Policy 86 674–683.

National Household Travel Survey. 2017. URL `https://nhts.ornl.gov/`.

Nguyen, Sang. 1974. An algorithm for the traffic assignment problem. Transportation Science 8(3) 203–216.

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

OSM. 2021. Download OpenStreetMap data for this region: Illinois. URL https://download.geofabrik.de/north-america/us/illinois.html.

Parks, Jamie, Paul Ryus, Alison Tanaka, Chris Monsere, Nathan McNeil, Jennifer Dill, William Schultheiss. 2012. District Department of Transportation bicycle facility evaluation. URL https://ddot.dc.gov/sites/default/files/dc/sites/ddot/publication/attachments/ddot_bike_evaluation_summary_final_report_part1_0.pdf.

Patriksson, Michael. 2015. The traffic assignment problem: models and methods. Courier Dover Publications.

Rastogi, Rajat, KV Krishna Rao. 2003. Travel characteristics of commuters accessing transit: Case study. Journal of Transportation Engineering 129(6) 684–694.

Roughgarden, Tim, Éva Tardos. 2002. How bad is selfish routing? Journal of the ACM (JACM) 49(2) 236–259.

Saban, Daniela, Gabriel Y Weintraub. 2019. Procurement mechanisms for assortments of differentiated products. Working paper, Stanford University, California .

Schwieterman, Joseph, Mallory Livingston. 2018. Partners in transit: A review of partnerships between transportation network companies and public agencies in the united states .

Scott, Darren M, Wei Lu, Matthew J Brown. 2021. Route choice of bike share users: Leveraging GPS data to derive choice sets. Journal of Transport Geography 90 102903.

Singhvi, Divya, Somya Singhvi, Peter I Frazier, Shane G Henderson, Eoin O'Mahony, David B Shmoys, Dawn B Woodard. 2015. Predicting bike usage for new york city's bike sharing system. Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.

So, Kut C, Christopher S Tang. 2000. Modeling the impact of an outcome-oriented reimbursement policy on clinic, patients, and pharmaceutical firms. Management Science 46(7) 875–892.

Strauss, Jillian, Luis F Miranda-Moreno. 2013. Spatial modeling of bicycle activity at signalized intersections. Journal of Transport and Land Use 6(2) 47–58.

Swamy, Chaitanya. 2012. The effectiveness of Stackelberg strategies and tolls for network congestion games. ACM Transactions on Algorithms (TALG) 8(4) 1–19.

Szeto, Wai Yuen, Xiaoqing Jaber, Margaret O'Mahony. 2010. Time-dependent discrete network

design frameworks considering land use. Computer-Aided Civil and Infrastructure Engineering 25(6) 411–426.

Taylor, Terry A, Wenqiang Xiao. 2014. Subsidizing the distribution channel: Donor funding to improve the availability of malaria drugs. Management Science 60(10) 2461–2477.

Thai, Jérôme, Rim Hariss, Alexandre Bayen. 2015. A multi-convex approach to latency inference and control in traffic equilibria from sparse data. 2015 American Control Conference (ACC). IEEE, 689–695.

The Economist. 2019. Public transport is in decline in many wealthy cities. URL https://www.economist.com/international/2018/06/21/public-transport-is-in-decline-in-many-wealthy-cities.

The White House. 2021. Fact sheet: The bipartisan infrastructure deal. URL https://www.whitehouse.gov/briefing-room/statements-releases/2021/11/06/fact-sheet-the-bipartisan-infrastructure-deal/.

Thorsten, Koska, Frederic Rudolph. 2016. The role of walking and cycling in reducing congestion. URL https://epub.wupperinst.org/frontdoor/deliver/index/docId/6597/file/6597_Reducing_Congestion.pdf.

Tomer, Adie, Elizabeth Kneebone, Robert Puentes, Alan Berube. 2011. Missed opportunity: Transit and jobs in metropolitan america. URL https://www.brookings.edu/wp-content/uploads/2016/06/0512_jobs_transit.pdf.

Transport for London. 2019. Congestion charge (official). URL https://tfl.gov.uk/modes/driving/congestion-charge.

Uber. 2020. Make your train with uber transit. URL https://www.uber.com/newsroom/make-my-train/.

UN. 2015. Make cities and human settlements inclusive, safe, resilient and sustainable. URL https://sdgs.un.org/goals/goal11.

UN. 2019. 2018 revision of world urbanization prospects. URL https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf.

United States Bureau of Public Roads. 1964. Traffic assignment manual for application with a

large, high speed computer, vol. 37. US Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division.

US Census Bureau. 2018a. Means of transportation to work by selected characteristics. American Community Survey .

US Census Bureau. 2020. American community survey 5-year data (2009-2019). URL https://www.census.gov/data/developers/data-sets/acs-5year.html.

US Census Bureau, Center for Economic Studies. 2018b. Us census bureau center for economic studies publications and reports page. URL https://lehd.ces.census.gov/data/.

US DOT. 2009. Assessing the full costs of congestion on surface transportation systems and reducing them through pricing. URL https://www.transportation.gov/office-policy/transportation-policy/assessing-full-costs-congestion-surface-transportation-systems.

US DOT. 2017. Traffic congestion and reliability: Linking solutions to problems. URL https://ops.fhwa.dot.gov/congestion_report_04/chapter4.htm.

Van Ommeren, Jos, Gerard J Van den Berg, Cees Gorter. 2000. Estimating the marginal willingness to pay for commuting. Journal of regional science 40(3) 541–563.

Wang, Hua, Xiaoning Zhang. 2016. Joint implementation of tradable credit and road pricing in public-private partnership networks considering mixed equilibrium behaviors. Transportation Research Part E: Logistics and Transportation Review 94 158–170.

Wardrop, John Glen. 1952. Road paper. some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers 1(3) 345.

Wei, Keji, Vikrant Vaze, Alexandre Jacquillat. 2021. Transit planning optimization under ride-hailing competition and traffic congestion. Transportation Science .

Weimer, David L, Aidan R Vining. 2017. Policy analysis: Concepts and practice. Routledge.

Wu, Chien-Fu. 1981. Asymptotic theory of nonlinear least squares estimation. The Annals of Statistics 9(3) 501–513.

Xiao, Feng, HM Zhang. 2014. Pareto-improving and self-sustainable pricing for the morning commute with nonidentical commuters. Transportation Science 48(2) 159–169.

Xiao, Ping, Ruli Xiao, Yitian Liang, Xinlei Chen, Wei Lu. 2019. The effects of a government's subsidy program: Accessibility beyond affordability. Management Science .

Xu, Shaoji, Jianzhong Zhang. 1995. An inverse problem of the weighted shortest path problem. Japan Journal of Industrial and Applied Mathematics 12(1) 47–59.

Xu, Yangyang, Wotao Yin. 2013. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM Journal on Imaging Sciences 6(3) 1758–1789.

Yang, Nan, Yong Long Lim. 2018. Temporary incentives change daily routines: Evidence from a field experiment on singapore's subways. Management Science 64(7) 3365–3379.

Yu, Jiayi Joey, Christopher S Tang, Zuo-Jun Max Shen. 2018. Improving consumer welfare and manufacturer profit via government subsidy programs: Subsidizing consumers or manufacturers? Manufacturing & Service Operations Management 20(4) 752–766.

Zhan, Xianyuan, Samiul Hasan, Satish V Ukkusuri, Camille Kamga. 2013. Urban link travel time estimation using large-scale taxi data with partial information. Transportation Research Part C: Emerging Technologies 33 37–49.

Zhang, Jianzhong, Mao-Cheng Cai. 1998. Inverse problem of minimum cuts. Mathematical Methods of Operations Research 47(1) 51–58.

Zhang, Jianzhong, Zhongfan Ma. 1996. A network flow method for solving some inverse combinatorial optimization problems. Optimization 37(1) 59–72.

Zhang, Jianzhong, Zhongfan Ma, Chao Yang. 1995. A column generation method for inverse shortest path problems. Zeitschrift für Operations Research 41(3) 347–358.

Zhang, Jing, Ioannis Ch Paschalidis. 2017. Data-driven estimation of travel latency cost functions via inverse optimization in multi-class transportation networks. 2017 IEEE 56th Annual Conference on Decision and Control (CDC). IEEE, 6295–6300.

Zhang, Jing, Sepideh Pourazarm, Christos G Cassandras, Ioannis Ch Paschalidis. 2018. The price of anarchy in transportation networks: Data-driven evaluation and reduction strategies. Proceedings of the IEEE 106(4) 538–553.

Zhao, Qi, Arion Stettner, Ed Reznik, Daniel Segrè, Ioannis Ch Paschalidis. 2015. Learning cellular

objectives from fluxes by inverse optimization. 2015 54th IEEE Conference on Decision and Control (CDC). IEEE, 1271–1276.

Zhu, Jing, Yingling Fan. 2018. Daily travel behavior and emotional well-being: Effects of trip mode, duration, purpose, and companionship. Transportation Research Part A: Policy and Practice 118 360–373.