Planning Bike Lanes with Data: Ridership, Congestion, and Path Selection

Sheng Liu
Rotman School of Management, University of Toronto, sheng.liu@rotman.utoronto.ca.

Auyon Siddiq
Anderson School of Management, University of California, Los Angeles, auyon.siddiq@anderson.ucla.edu.

Jingwei Zhang
Anderson School of Management, University of California, Los Angeles, jingwei.zhang.phd@anderson.ucla.edu.

Urban infrastructure is vital for sustainable cities. In recent years, municipal governments have invested heavily in the expansion of bike lane networks to meet growing demand, promote ridership, and reduce emissions. However, re-allocating road capacity to cycling is often contentious due to the risk of amplifying traffic congestion. In this paper, we develop a method for planning bike lanes that accounts for ridership and congestion effects. We first present a procedure for estimating parameters of a traffic equilibrium model, which combines an inverse optimization method for predicting driving times with an instrumental variables method for estimating a commuter mode choice model. We then formulate a prescriptive model that selects paths in a road network for bike lane installation while endogenizing cycling demand and driving travel times. We conduct an empirical study on the City of Chicago that brings together several datasets that describe the urban environment – including the road and bike lane networks, vehicle flows, commuter mode choices, bike share trips, driving and cycling routes, demographic features, and points of interest – with the goal of estimating 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 cycling ridership from 3.6% to 6.1%, with at most an 9.4% increase in driving times. We also find that three intuitive heuristics for bike lane planning can lead to lower ridership and worse congestion outcomes, highlighting the value of a holistic and data-driven approach to urban infrastructure planning.

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 invested heavily into the installation of new bike lanes.[1] For example, between 2014 to 2019, New York City increased its 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.[2] 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 so is non-trivial due to the opposing forces alluded to above: While adding bike lanes might increase the attractiveness of cycling and take

[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).

[2] In special cases, road capacity for automotive vehicles can be preserved by instead removing on-street parking, albeit with negative impacts on businesses and residences (Federal Highway Administration 2006).

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 these competing 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 the impacts on ridership (Khany 2022).

1.1. Contributions

Where should bike lanes be built to promote cycling ridership while mitigating congestion effects? We address this question by developing a method 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,[3] and is illustrated through an extensive empirical study using data from the City of Chicago.

Our method consists of three main steps. In the first step, we estimate a congestion (i.e., driving travel time) function from features of the road network and observed vehicle flows. While there is a rich literature on traffic equilibrium models (following the celebrated 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 such models. Our procedure is based on the paradigm of inverse optimization, which refers to estimation of an optimization model's parameters from (potentially noisy) observations of its solutions (Chan et al. 2021); accordingly, we assume traffic data are observations of Wardrop equilibrium network flows.[4] The estimation procedure 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 estimation problem a challenging non-convex optimization problem. To obtain solutions, we consider an approximation based on Lagrangian relaxation that admits a multi-convex structure, which allows us to generate reliable estimates through an iterative solution procedure.

In the second step, taking the congestion parameters as input, we estimate a multinomial logit mode choice model in which driving and cycling utilities depend on the locations of bike lanes and predicted equilibrium travel times, among other covariates. In particular, we use an instrumental

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

[4] In particular, our estimator takes the form of a mathematical program with equilibrium constraints (MPEC), and thus falls into the optimization-based framework for structural estimation proposed by Su and Judd (2012).

variables approach to address endogeneity in the locations of existing bike lanes and driving travel times, closely following established methods from the literature (Berry et al. 1995, Davis 2006, Kabra et al. 2020, He et al. 2021).

In the third step, we embed the estimated traffic congestion and mode choice models into a prescriptive "path selection" model that identifies paths in the road network for bike lane installation. The model endogenizes the mode choice of all commuters 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 bound the suboptimality due to the linearization as a function of network and demand parameters.

We use our method to conduct an empirical study using data from the City of Chicago. Our study combines multiple datasets, including Chicago's road and bike lane network, observed traffic flows, commuter mode choices, bike share trips, taxi trips, driving and cycling routes, demographics, and points of interest in the city. We estimate that adding 25 miles of bike lanes as prescribed by our model can increase cycling ridership from 3.6% to 6.1% in downtown Chicago, without increasing the travel time of any driving route by more than 9.4%.

To evaluate our modeling approach, we benchmark 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.

1.2. Related Literature

Our paper 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 estimation problem takes the form of an inverse optimization model 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 (Dafermos and Sparrow 1969). 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. (2021) 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, our paper 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 estimation problem is non-convex. However, we show in this paper that our estimates are 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 estimation problem, 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.

Our paper 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. 2022). 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 area 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.

1.3. Outline

We provide here a brief overview of the paper. §2 outlines model fundamentals, including our approach for modeling the traffic equilibrium, congestion (i.e., driving times), and commuters' mode choices. §3 describes the data environment and presents an estimation method for recovering parameters of the traffic model from features of the road network, observed vehicle flows, and origin-destination demands. Notably, our focus is on a setting where equilibrium travel times are unobserved; this requires endogenizing travel time within the estimator, resulting in a mathematical program with equilibrium constraints (MPEC). §4 presents a prescriptive model for bike lane expansion, an approximation scheme to improve computational tractability, and a corresponding upper bound on the approximation error. §5 presents an empirical study for the City of Chicago. §6 discusses policy implications and concludes.

2. Model: Traffic Congestion and Commuters' Mode Choices

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, §2.1 formalizes the road network, §2.2 presents a model of traffic congestion, and §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.

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 $i,j \in \mathcal{I}, i \neq j$ 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 $p \in \mathcal{P}_w^D$, $w \in \mathcal{W}$ 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.

2.2. Wardrop Equilibrium 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 the mass of drivers 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 $\mathbf{v} \in \mathbb{R}^{|\mathcal{S}^D|}$ be the vectors of ϕ_p and v_s, respectively. Then for fixed $\mathbf{d}^D$, $(\boldsymbol{\phi}, \mathbf{v})$ are *feasible* flows if they satisfy

$$\sum_{p \in \mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{1a}$$

$$v_s = \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p, \quad s \in \mathcal{S}^D, \tag{1b}$$

$$(\boldsymbol{\phi}, \mathbf{v}) \geq \mathbf{0}. \tag{1c}$$

For convenience, we define

$$\Omega\left(\mathbf{d}^D\right) = \{(\boldsymbol{\phi}, \mathbf{v}) \,|\, (\boldsymbol{\phi}, \mathbf{v}) \text{ satisfies (1)}\} \tag{2}$$

to be the set of all feasible flows given driving demand $\mathbf{d}^D$. The equations in (1) assumes 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)).[5]

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 be altered by the presence of bike lanes. We define for each segment a congestion function α_s, which depends linearly on segment features:

$$\alpha_s\left(\mathbf{x}, \boldsymbol{\theta}\right) = \boldsymbol{\theta}^\top \mathbf{q}_s\left(\mathbf{x}\right). \tag{3}$$

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

$$t_s^D(\mathbf{x}, \boldsymbol{\theta}, v_s) = \alpha_s\left(\mathbf{x}, \boldsymbol{\theta}\right) \cdot v_s + T_s, \tag{4}$$

[5] An alternative approach is the *link-flow* formulation, which requires variables for each segment and OD pair in the network (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.

where T_s is the free-flow travel time. We impose the mild assumption that $\alpha_s(\mathbf{x}, \boldsymbol{\theta}) > 0$, which implies driving times strictly increase in traffic flow v_s.[6] The total travel time along each path $p \in \mathcal{P}^D$ is then given by

$$t_p^D(\mathbf{x}, \boldsymbol{\theta}, \mathbf{v}) = \sum_{s \in \mathcal{S}_p^D} t_s^D(\mathbf{x}, \boldsymbol{\theta}, v_s). \tag{5}$$

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 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 $t_w^D(\mathbf{x}, \boldsymbol{\theta})$ denote the equilibrium travel time on OD pair w. Then Wardrop's first principle is equivalent to the following set of conditions:

$$t_w^D(\mathbf{x}, \boldsymbol{\theta}) - t_p^D(\mathbf{x}, \boldsymbol{\theta}, \mathbf{v}^*(\boldsymbol{\theta})) \left\{ \begin{array}{l} = 0, \text{ if } \phi_p^*(\mathbf{x}, \boldsymbol{\theta}) > 0 \\ \leq 0, \text{ if } \phi_p^*(\mathbf{x}, \boldsymbol{\theta}) = 0 \end{array} \right\}, \quad p \in \mathcal{P}_w^D,\ w \in \mathcal{W}. \tag{6}$$

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

$$\Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta}) = \{(\boldsymbol{\phi}, \mathbf{v}) | (\boldsymbol{\phi}, \mathbf{v}) \in \Omega\left(\mathbf{d}^D\right) \text{ and } (\boldsymbol{\phi}, \mathbf{v}) \text{ satisfies (6)}\} \tag{7}$$

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 $\alpha_s(\mathbf{x}, \boldsymbol{\theta})$.

The traffic model outlined above rests on the strong assumption that every driver chooses the fastest route in equilibrium and ignores all other route characteristics, which may not hold in practice. Nonetheless, our empirical results suggest that Wardrop equilibrium provides a reasonable approximation of *network-level* traffic dynamics (see Figure 7 in Appendix A), even if it does not accurately reflect individual drivers' route choices, which aligns with prior empirical work (Yildirimoglu and Kahraman 2018). In Appendix G, we discuss a variation of our model based on the *stochastic user equilibrium (SUE)* problem (Prashker and Bekhor 2004), in which drivers account for route characteristics other than travel time and have idiosyncratic preferences over routes.

[6] The most well-known congestion function takes the form of $t_s^D(\mathbf{x}, v_s) = T_s \cdot (1 + \alpha(v_s/m_s)^4)$, 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 estimation problem. Linear congestion functions have appeared elsewhere in the literature (Roughgarden and Tardos 2002, Lin et al. 2011, Swamy 2012).

2.3. Commuter Mode Choice

We now describe how commuters select among alternate travel modes. Following the literature on mode choice (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 utility of a commuter who wishes to travel by mode m on OD pair w. Then the utility of driving, cycling and outside options on OD pair w is

$$u_w^D(\mathbf{x}, \boldsymbol{\theta}) = \beta_0^D + \beta_1^D \cdot t_w^D(\mathbf{x}, \boldsymbol{\theta}) + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D + \xi_w^D + \epsilon_w^D, \tag{8a}$$

$$u_w^C(\mathbf{x}) = \beta_0^C + \beta_1^C \cdot t_w^C + \beta_2^C \cdot \rho_w(\mathbf{x}) + (\tilde{\boldsymbol{\beta}}^C)^\top \mathbf{X}_w^C + \xi_w^C + \epsilon_w^D, \tag{8b}$$

$$u_w^O(\mathbf{x}) = \beta_1^O \cdot t_w^O + (\tilde{\boldsymbol{\beta}}^O)^\top \mathbf{X}_w^O + \epsilon_w^O. \tag{8c}$$

The driving utility $u_w^D(\mathbf{x}, \boldsymbol{\theta})$ depends on the driving time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$ and a vector of additional covariates $\mathbf{X}_w^D$.[7] Note that the equilibrium conditions (6) imply for each $w \in \mathcal{W}$, the driving time is the same for all $p \in \mathcal{P}_w^D$, and (8b) implies that all drivers on OD pair w share the same utility regardless of their selected path, which depends primarily on the travel time.

The cycling utility $u_w^C(\mathbf{x})$ depends on the cycling time t_w^C, bike lane *coverage* $\rho_w(\mathbf{x})$ (i.e., the proportion of the path that is covered by a bike lane), and additional covariates $\mathbf{X}_w^C$. Cycling time and bike lane coverage have both been observed to be significant factors in cycling adoption (Hood et al. 2011, Broach et al. 2012).[8] For tractability, we assume that for each OD pair $w \in \mathcal{W}$ all cyclists travel along the same path p_w^C.

To define bike lane coverage, let ℓ_s be the length of segment s. Then the coverage for OD pair w is given by

$$\rho_w(\mathbf{x}) = \frac{1}{L_w} \sum_{s \in \mathcal{S}_w^C} \ell_s \cdot x_s, \tag{9}$$

where $\mathcal{S}_w^C$ is the set of segments that lie on cycling path p_w^C. Note that cycling utility does not depend on the congestion parameter $\boldsymbol{\theta}$, which influences driving times only. Lastly, we assume the utility of the outside option depends on the travel time by public transit t_w^O and additional covariates $\mathbf{X}_w^O$.

In the utility functions above, ξ_w^D and ξ_w^C are unobservable components of utility for driving and cycling that are common to all users for OD pair w. For instance, ξ_w^D may include factors that influence the inconvenience of driving for the OD pair w, such as the presence of traffic lights along

[7] The covariates $\mathbf{X}_w^D$, $\mathbf{X}_w^C$, and $\mathbf{X}_w^O$ may include attributes of each mode, socio-economic characteristics of commuters, or other contextual information (Hensher and Greene 2003).

[8] 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.

the driving routes. Similarly, ξ_w^C may represent unobservables associated with the degree of bike-friendliness, such as pavement smoothness along the cycling route. The ϵ_w^m's are idiosyncratic error terms that are independent and identically distributed and follow an extreme value distribution.

Let $\tilde{d}_w$ be the total demand across all travel modes for OD pair w. Based on the utilities defined in (8), the demand for mode m on OD pair w is then given by

$$d_w^m(\mathbf{x}, \boldsymbol{\theta}) = \tilde{d}_w \cdot P_w^m = \tilde{d}_w \cdot \frac{e^{u_w^m(\mathbf{x},\boldsymbol{\theta})}}{e^{u_w^C(\mathbf{x})} + e^{u_w^D(\mathbf{x},\boldsymbol{\theta})} + e^{u_w^O}}, \tag{10}$$

where P_w^m is the share of consumers choosing mode m between w. We assume $\tilde{d}_w$ is exogenous and does not depend on $\mathbf{x}$ or $\boldsymbol{\theta}$. Note that the MNL model (10) implies $\tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m(\mathbf{x}, \boldsymbol{\theta})$. Having defined how the equilibrium traffic assignment depends on $\mathbf{x}$, $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, in §3 we address how to estimate the parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ from data.

The specification of the commuter choice model above captures the key trade-offs in our study, namely, the impact of congestion and bike lane coverage on the cycling and driving utilities. Notably, we do not include vehicle flows in the cycling utility function, which may be a significant variable (e.g., if traffic alters cyclists' perceptions of safety). In particular, because the vehicle flows $\mathbf{v}^*$ are endogenous, including them in the cycling utility would introduce a large number of non-convex, bi-linear terms into the §4 optimization model (as a result of $\mathbf{v}^*$ being multiplied with other decision variables in the formulation), which would greatly undermine the computational tractability of the optimization step. As a consequence, we assume throughout that cyclists' utilities are independent of vehicle flows.

3. Estimation

Our estimation procedure consists of two steps. In the first step, we estimate the congestion parameters $\boldsymbol{\theta}$ that define the Wardrop equilibrium 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 $\boldsymbol{\beta}$.

This section is organized as follows. In §3.1, we define an estimation problem for recovering the congestion parameters $\boldsymbol{\theta}$ from appropriate road network and vehicle flow data. In §3.2, we show that the estimation problem for $\boldsymbol{\theta}$ can be posed as an inverse optimization model and present an accompanying solution procedure. In §3.3, we address estimating the mode choice parameters $\boldsymbol{\beta}$, taking the estimated values of $\boldsymbol{\theta}$ as input; in particular, we note endogeneity challenges in estimating $\boldsymbol{\beta}$, which we address through the use of instrumental variables in our empirical study.

3.1. Estimation of Wardrop Congestion Parameters

The estimation procedure for $\boldsymbol{\theta}$ takes road segment-level data of the form

$$(\mathbf{q}_s, z_s), \quad s = 1, \ldots, n, \tag{11}$$

where $\mathbf{q}_s$ are the features of segment s, z_s is a noisy observation of the equilibrium vehicle flow on segment s, and n is the number of segments in the network (equivalently, $|\mathcal{S}|$). Additionally, for each OD pair $w \in \mathcal{W}$, the procedure takes as input the observed driving demand d_w^D, the driving path sets $\mathcal{P}_w^D$, and the free-flow travel time for each path $p \in \mathcal{P}_w^D$ (i.e., $\sum_{s \in \mathcal{S}_p^D} T_s$).

Notably, our focus is on a setting where the path-level equilibrium driving times (t_p^D in (5)) are *not* observed; instead, these variables are endogenized within the estimation procedure. If the path-level driving times t_p^D are observed, then $\boldsymbol{\theta}$ can be estimated more directly by linear regression or a generalized method of moments (GMM) estimator. We briefly discuss the relative merits of our approach compared to more standard methods at the end of §3.2.

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:

$$L_n(\boldsymbol{\theta}) = \underset{\boldsymbol{\phi}, \mathbf{v}}{\text{minimize}} \ \|\mathbf{v} - \mathbf{z}\|_2^2 \tag{12a}$$

$$\text{subject to } (\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\boldsymbol{\theta}), \tag{12b}$$

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 the loss function over a parameter set Θ:

$$\hat{\boldsymbol{\theta}}_n \in \underset{\boldsymbol{\theta} \in \Theta}{\text{argmin}} \ L_n(\boldsymbol{\theta}). \tag{13}$$

The estimation problem (13) can be viewed as an analog to non-linear least squares regression (Jennrich (1969), Wu (1981)). The main distinction between (13) and traditional non-linear least squares regression is that the response variables $\mathbf{z}$ 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. As a result, enforcing the equilibrium conditions (12b) makes solving (13) computationally non-trivial.

To see the intuition behind how $\boldsymbol{\theta}$ is identified, note that variation throughout the network in the OD driving demands d_w^D and road segment features $\mathbf{q}_s$ induces variation over the road segments in

the observed vehicle flows z_s and travel times t_s^D. Because all travel times are determined by the common congestion function $\alpha_s(\boldsymbol{\theta})$ given in (3), this variation in OD driving demands and segment features combined with the structure of the road network helps pin down $\boldsymbol{\theta}$.

Under appropriate conditions, it can be shown that the sequence of estimates $\hat{\boldsymbol{\theta}}_n$ from solving (13) are statistically consistent, in that they converge to ground-truth, data-generating parameters as the size of the network tends to infinity. In Appendix A.1, we present sufficient conditions for the consistency of the estimates $\hat{\boldsymbol{\theta}}_n$, including an identifiability condition related to the structure of the road network.

3.2. Inverse Optimization Formulation and Solution Method

Next, we present an exact reformulation of the estimation problem (13) and describe a corresponding solution method. Our reformulation relies on expressing (13) as an inverse optimization problem (Chan et al. 2021). In particular, we leverage a classical result that shows the vehicle flows in Wardrop equilibrium can be attained at the solution to a convex optimization problem (Dafermos and Sparrow 1969). As a consequence, the problem of estimating $\boldsymbol{\theta}$ from data of the form (11) amounts to solving an inverse optimization problem. In general, inverse optimization problems are themselves mathematical programs, and are often non-convex as a consequence of enforcing optimality or equilibrium conditions, which is the case in our setting as well.

To express (13) as an inverse optimization problem, we develop an explicit representation of the equilibrium conditions $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\boldsymbol{\theta})$. To that end, consider the following "Wardrop least-squares" formulation:

$$
\begin{aligned}
\underset{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b}}{\text{minimize}} \quad & \|\mathbf{v}-\mathbf{z}\|_2^2 && \text{(14a)}\\
\text{subject to} \quad & \nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} = 0, && \text{(14b)}\\
& \sum_{\{w\in\mathcal{W}\mid p\in\mathcal{P}_w^D\}} b_w \le \nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right), \quad p\in\mathcal{P}^D, && \text{(14c)}\\
\text{(WLS)} \qquad & \sum_{p\in\mathcal{P}_w^D} \phi_p = d_w^D, \quad w\in\mathcal{W}, && \text{(14d)}\\
& v_s = \sum_{\{p\in\mathcal{P}^D\mid s\in\mathcal{S}_p^D\}} \phi_p, \quad s\in\mathcal{S}^D, && \text{(14e)}\\
& \boldsymbol{\theta}\in\Theta, && \text{(14f)}
\end{aligned}
$$

where

$$
g(\boldsymbol{\phi},\boldsymbol{\theta}) = \sum_{s\in\mathcal{S}} \left(\frac{1}{2}\cdot\alpha_s(\mathbf{x},\boldsymbol{\theta})\cdot\left[\sum_{\{p\in\mathcal{P}^D\mid s\in\mathcal{S}_p^D\}}\phi_p\right]^2 + T_s\cdot\left[\sum_{\{p\in\mathcal{P}^D\mid s\in\mathcal{S}_p^D\}}\phi_p\right]\right). \tag{15}
$$

We now present a result that establishes an equivalence between the estimation problem given in (13) and the optimization problem WLS:

Proposition 1. *Let $\boldsymbol{\theta}^*$ be attained at an optimal solution to WLS. Then $\boldsymbol{\theta}^* \in \operatorname{argmin}_{\theta \in \Theta} L_n(\boldsymbol{\theta})$.*

All proofs are contained in Appendix H. At a high level, Proposition 1 relies on a well-known result which shows that equilibrium network flows $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\boldsymbol{\theta})$ correspond to the solution to a convex optimization problem (Dafermos and Sparrow 1969). Expressing the optimality conditions of that convex program as a variational inequality and invoking linear programming duality[9] then yields the constraints (14b) and (14c), which are together equivalent to the Wardrop equilibrium conditions (6). This implies that a solution $(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{v}, \mathbf{b})$ that satisfies (14b)-(14e) must also satisfy $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\boldsymbol{\theta})$, which yields the equivalence between (13) and (14). We note here that formulation (14) is also a mathematical program with equilibrium constraints (MPEC), and as a consequence falls into the general MPEC-based framework for structural estimation proposed by Su and Judd (2012).

The key challenge in obtaining solutions to (14) is that constraint (14b) 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 (14b) by introducing a penalty parameter $\lambda > 0$ and the auxiliary variable ϵ, which yields the following approximation of WLS:

$$\underset{\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{v}, \mathbf{b}, \epsilon}{\text{minimize}} \ \|\mathbf{v} - \mathbf{z}\|_2^2 + \lambda \cdot \epsilon \tag{16a}$$

(WLS-A) $$\text{subject to } \nabla_{\boldsymbol{\phi}} g(\boldsymbol{\phi}, \boldsymbol{\theta})^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} \leq \epsilon, \tag{16b}$$

$$(14c) - (14f).$$

In the approximation WLS-A, the loss (16a) can be interpreted as a linear combination of the "flow error" $\|\mathbf{v} - \mathbf{z}\|_2^2$ and the "Wardrop error" ϵ.[10]

In addition to circumventing the numerical issues of WLS, the approximation 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 (15), 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 inverse optimization problem WLS-A are provided in Appendix A.3. In

[9] See Aghassi et al. (2006) for details on the on the reformulation of variational inequalities using linear programming duality.

[10] Note that multiplying both sides of (14c) by $\boldsymbol{\phi}$ yields $\nabla_{\boldsymbol{\phi}} g(\boldsymbol{\phi}, \boldsymbol{\theta})^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} \geq 0$, which implies the Wardrop error ϵ is always non-negative.

Appendix F, we numerically evaluate the validity of the approximation using synthetic data that mimic the empirical data distribution from our case study in §5.

As noted earlier, our method is tailored for a setting where the equilibrium path-level driving times t_p^D are unobserved, which requires endogenizing them within the estimation procedure. If suitable data for t_p^D is available, then $\boldsymbol{\theta}$ can be directly estimated from (4) and (5) through linear regression or a generalized method of moments (GMM) estimator, without the need to enforce Wardrop equilibrium conditions during estimation. Compared to these standard econometric techniques, our method has milder data requirements, but comes at the cost of a more complex estimation procedure and a higher computational burden. Our method also naturally handles missing observations for the segment-level vehicle flows z_s, which would need to be imputed if applying regression or GMM, even if the path times t_p^D are observed. Lastly, our MPEC-based method does not easily allow for statistical inference (e.g., confidence intervals for $\boldsymbol{\theta}$),[11] which is otherwise straightforward for linear regression and GMM under appropriate conditions.

3.3. Estimation of Commuter Mode Choice Model

Given the congestion parameters $\boldsymbol{\theta}$, estimating the mode choice model parameters $\boldsymbol{\beta}$ outlined in §2.3 follows from standard methods. Specifically, let $\tilde{\boldsymbol{\theta}}$ be the estimate from solving WLS-A. Then the corresponding equilibrium segment flows $\mathbf{v}^*(\tilde{\boldsymbol{\theta}})$ are given by solving a convex optimization problem (see the proof of Proposition 1 in Appendix H). 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 (6) that the equilibrium travel time for each OD pair $w \in \mathcal{W}$ is given by

$$t_w^D(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}) = \min_{p \in \mathcal{P}_w^D} \left\{ t_p^D \left(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}}, \mathbf{v}^*(\tilde{\boldsymbol{\theta}}) \right) \right\}. \tag{17}$$

Therefore, the status quo bike lane network $\tilde{\mathbf{x}}$ and the estimate $\tilde{\boldsymbol{\theta}}$ together pin down the travel times $t_w^D(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$, and thus the driving utilities $u_w^D(\tilde{\mathbf{x}}, \tilde{\boldsymbol{\theta}})$. Similarly, the cycling utility under the existing bike lane network is given by $u_w^C(\tilde{\mathbf{x}})$, where the coverage $\rho_w(\tilde{\mathbf{x}})$ is determined by (9).[12]

Because we observe commuters' mode choices aggregated as OD-level mode shares (i.e., P_w^m), we estimate $\boldsymbol{\beta}$ following the literature of linear logit models (Theil 1969, McFadden et al. 1973, Malhotra 1984, Cooper 1993). Specifically, by taking the logarithm of both sides of (10), we obtain

$$\log\left(P_w^D\right) - \log\left(P_w^O\right) = \left(u_w^D(\mathbf{x}, \boldsymbol{\theta}) - u_w^O(\mathbf{x}, \boldsymbol{\theta})\right)$$

[11] In general, obtaining confidence intervals from MPEC-based estimators is both analytically and computationally challenging, and is a relatively open question (see, e.g., Reich and Judd (2020)).

[12] 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 §5.

$$= \beta_0^D + \beta_1^D \cdot t_w^D(\mathbf{x}, \boldsymbol{\theta}) + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D - (\beta_1^O \cdot t_w^O + (\tilde{\boldsymbol{\beta}}^O)^\top \mathbf{X}_w^O) + \xi_w^D, \tag{18a}$$

$$\log\left(P_w^C\right) - \log\left(P_w^O\right) = \left(u_w^C(\mathbf{x}, \boldsymbol{\theta}) - u_w^O(\mathbf{x}, \boldsymbol{\theta})\right)$$
$$= \beta_0^C + \beta_1^C \cdot t_w^C(\mathbf{x}, \boldsymbol{\theta}) + \beta_2^C \cdot \rho_w(\mathbf{x}) + (\tilde{\boldsymbol{\beta}}^C)^\top \mathbf{X}_w^C - (\beta_1^O \cdot t_w^O + (\tilde{\boldsymbol{\beta}}^O)^\top \mathbf{X}_w^O) + \xi_w^C, \tag{18b}$$

which makes the estimation of $\boldsymbol{\beta}$ amenable to linear regression. However, the use of ordinary least squares (OLS) regression may lead to biased estimates of β_1^D and β_2^C due to potential endogeneity of the driving travel time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$ and bike lane coverage $\rho_w(\mathbf{x})$, respectively. In particular, the demand for driving influences the driving time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$, leading to correlation between $t_w^D(\mathbf{x}, \boldsymbol{\theta})$ and the error term ξ_w^D. Similarly, there may be correlation between the error term ξ_w^C and bike lane coverage $\rho_w(\mathbf{x})$, due to, for example, city planners selecting the locations of bike lanes based on existing cycling demand. In our empirical study, we address these sources of endogeneity using instrumental variables, which follows prior work on cycling infrastructure planning (Kabra et al. 2020, He et al. 2021) and is discussed further in §5.3.

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 $\mathbf{x}$ from the perspective of a city planner. §4.1 formulates the main optimization model for selecting bike lanes, §4.2 develops an approximation scheme, and §4.3 discusses additional constraints that can be included to limit the impact on traffic congestion.

4.1. Optimization Model for Bike Lane Path Selection

Our prescriptive model seeks to maximize cycling ridership while satisfying constraints on bike lane continuity, budget, and the traffic equilibrium. Next, we describe each component of the prescriptive model, starting with the constraints.

Continuity. Continuity of bike lanes is often a priority of city planners (Federal Highway Administration 2006, Bao et al. 2017). To that end, we consider bike lane design 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 installed along the entirety of 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 and only if a bike lane is located on segment s, we enforce coherence between $\mathbf{x}$ and $\mathbf{y}$ using the constraints

$$x_s \geq y_w, \quad s \in S_w^C,\, w \in \mathcal{W}, \tag{19a}$$

$$x_s \leq \sum_{\{w \in \mathcal{W} | s \in \mathcal{S}_w^C\}} y_w, \quad s \in \mathcal{S}. \tag{19b}$$

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

$$\sum_{s \in \mathcal{S}} \ell_s \cdot x_s \leq B. \tag{20}$$

Note that the model can accommodate pre-existing bike lanes by adding the constraints $x_s = 1$ for all road segments where a bike lane already exists, and dropping those segments from the summation in (20) so they do not consume the bike lane budget B.

Traffic equilibrium. For ease of exposition, let $\mathbf{d} = (\mathbf{d}^C, \mathbf{d}^D, \mathbf{d}^O)$ represent all mode-specific demands. To endogenize traffic congestion and cycling ridership within the optimization model, we include constraints that force the network flows $(\boldsymbol{\phi}, \mathbf{v})$ to abide by Wardrop equilibrium and the demands $\mathbf{d}$ to follow the MNL mode choice model. To that end, we define

$$\Gamma(\mathbf{x}) = \{(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \mid \mathbf{d} \text{ satisfies (10) and } (\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{x}, \mathbf{d}^D, \boldsymbol{\theta})\}. \tag{21}$$

The set $\Gamma(\mathbf{x})$ represents all equilibrium traffic assignments under the bike lane network $\mathbf{x}$, in addition to the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$. 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

$$(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x}). \tag{22}$$

As discussed in §4.2 below, enforcing this equilibrium condition will be the main computational challenge in solving the optimization model.

Cycling ridership. Lastly, we assume the city planner's goal is to maximize total cycling ridership, which 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 the *bike lane path selection* problem (BL) can be written as

$$\text{(BL)} \qquad \begin{aligned}
&\underset{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v}}{\text{maximize}} && \sum_{w \in \mathcal{W}} d_w^C && \text{(23a)}\\
&\text{subject to} && x_s \geq y_w, \quad s \in S_w^C,\, w \in \mathcal{W}, && \text{(23b)}\\
& && x_s \leq \sum_{\{w \in \mathcal{W} \mid s \in \mathcal{S}_w^C\}} y_w, \quad s \in \mathcal{S}, && \text{(23c)}\\
& && \sum_{s \in \mathcal{S}} \ell_s \cdot x_s \leq B, && \text{(23d)}\\
& && (\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x}), && \text{(23e)}\\
& && \mathbf{x} \in \{0,1\}^{|\mathcal{S}|}, && \text{(23f)}\\
& && \mathbf{y} \in \{0,1\}^{|\mathcal{W}|}. && \text{(23g)}
\end{aligned}$$

The formulation above is computationally challenging because the decision variables $\mathbf{x}, \mathbf{y}$ are binary and the traffic equilibrium condition (23e) is non-convex and non-linear as a result of depending on multinomial logit choice probabilities (see (10)). In the next section, we show that BL 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.

4.2. Linear Approximation and Suboptimality Bound

The approximation of BL consists of two steps: First, we reformulate (23) exactly by leveraging an equivalence between the equilibrium condition (23e) 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. To reformulate (23) exactly, we first present a result that shows the equilibrium traffic assignments $\Gamma(\mathbf{x})$ is equivalent to the set of optimal solutions to a convex optimization problem.

Proposition 2 (User equilibrium). *For any bike lane network* $\mathbf{x}$, $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x})$ *if and only if* $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v})$ *solves the convex optimization problem*

$$\underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v}}{\text{minimize}}\ S(\mathbf{x}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{v}) \tag{24a}$$

$$\text{subject to } \tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m,\ w \in \mathcal{W}, \tag{24b}$$

$$(\boldsymbol{\phi}, \mathbf{v}) \in \Omega\left(\mathbf{d}^D\right), \tag{24c}$$

where

$$S(\mathbf{x}, \boldsymbol{\theta}, \mathbf{d}, \mathbf{v}) = \sum_{w \in \mathcal{W}} \left(\sum_{m \in \{O,C\}} \int_0^{d_w^m} (-u_w^m(\mathbf{x})) \cdot dd + \int_0^{d_w^D} -\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D\right) \cdot dd \right) - \beta_1^D \sum_{s \in \mathcal{S}} \left(\int_0^{v_s} t_s^D(\mathbf{x}, \boldsymbol{\theta}, v) \cdot dv \right) \tag{25a}$$

$$+ \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} d_w^m \cdot \log(d_w^m). \tag{25b}$$

Further, the equilibrium mode demands $\mathbf{d}$ *and traffic flows* $\mathbf{v}$ *attained at a solution to* (24) *are unique.*

For conciseness, we suppress the dependence of the function $S(\cdot)$ on $\boldsymbol{\theta}$ in the remainder of this section. Proposition 2 follows closely from Abdulaal and LeBlanc (1979), Dafermos (1982) and Fisk (1980).[13] Following the literature, we refer to formulation (24) as the *user equilibrium* problem because it specifies equilibrium mode choices and driving routes. The objective $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is composed of two parts: a system-wide generalized cost term (25a), which is the sum of "disutilites" (i.e., negative of utilities) of all commuters, and an entropy term (25b). Note that $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is

[13] 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.

related to the observable road segment features $\mathbf{q}(\mathbf{x})$ and traffic flow $\mathbf{v}$ via the driving time function $t_s^D(\mathbf{x}, \boldsymbol{\theta}, v_s)$. Further, because $t_s^D(\mathbf{x}, \boldsymbol{\theta}, v_s)$ strictly increases in v_s, $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is strictly convex in $(\mathbf{d}, \mathbf{v})$.

Proposition 2 allows us to replace the equilibrium condition (23e) in the bike lane path selection model BL with the equivalent condition

$$(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \underset{\mathbf{d}, (\boldsymbol{\phi}, \mathbf{v}) \in \Omega(\mathbf{d}^D)}{\operatorname{argmin}} S(\mathbf{x}, \mathbf{d}, \mathbf{v}), \tag{26}$$

This substitution transforms problem (BL) 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 (10) in the path selection model and instead work with the convex problem (26), which is more tractable with respect to linearization. Further, as discussed below, the linearization of (26) yields a bi-level program with a linear lower level problem, which can be reformulated exactly as a MILP (Fortuny-Amat and McCarl 1981).

The uniqueness result in Proposition 2 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 BL as a MILP. Note that the upper-level objective function and constraints (23a)-(23d), (23f), and (23g) are all linear, so we focus on approximating (23e), which by Proposition 2 is equivalent to the user equilibrium problem (24). First, note that by evaluating each term in $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$ defined in (25), it can be re-written as

$$\begin{aligned} S(\mathbf{x}, \mathbf{d}, \mathbf{v}) = \sum_{w \in \mathcal{W}} & -\left(u_w^O d_w^O + u_w^C d_w^C + \left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}^D\right) d_w^D\right) \\ & - \beta_1^D \sum_{s \in \mathcal{S}} (T_s \cdot v_s) - \beta_1^D \sum_{s \in \mathcal{S}} \underbrace{\frac{1}{2} \alpha_s(\mathbf{x}) \cdot v_s^2}_{\xi_s(v_s)} + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \underbrace{d_w^m \log(d_w^m)}_{\psi(d_w^m)}, \end{aligned} \tag{27}$$

where, for fixed $\mathbf{x}$, the non-linearity is due to the final two terms, labeled $\xi_s(v_s)$ and $\psi(d_w^m)$. Each of these terms is convex in $\mathbf{v}$ and $\mathbf{d}$, respectively, which allows us to approximate them using well-known techniques from linear programming (see, e.g., §1.3 of Bertsimas and Tsitsiklis (1997)).

To linearize $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$, we approximate $\xi_s(v_s)$ with a piecewise linear function, where

$$\xi_s^r(v_s) = \alpha_s(\mathbf{x}) \cdot v_s^r \cdot v_s - \frac{1}{2} \cdot \alpha_s(\mathbf{x}) \cdot (v_s^r)^2 \tag{28}$$

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 $\xi_s^r(v_s)$ is also linear in $\mathbf{x}$, which is important for approximating BL as a MILP. Similarly, we approximate $\psi(d_w^m)$ with a piecewise linear function, where

$$\psi_w^r(d_w^m) = [1+\log(d_w^r)]\cdot d_w^m - d_w^r \tag{29}$$

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 (28) and (29), the linear approximation of the user equilibrium problem (24) can be written as

$$\underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}}{\text{minimize}}\ S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) \tag{30a}$$

$$\text{subject to } \tilde{d}_w = \sum_{m\in\mathcal{M}} d_w^m, \quad w\in\mathcal{W}, \tag{30b}$$

$$(\boldsymbol{\phi},\mathbf{v})\in\Omega(\mathbf{d}^D), \tag{30c}$$

$$\xi_s \geq \xi_s^r(\mathbf{x},v_s), \quad r\in\mathcal{R}_\xi, s\in\mathcal{S}, \tag{30d}$$

$$\psi_w^m \geq \psi_w^r(d_w^m), \quad r\in\mathcal{R}_\psi, w\in\mathcal{W}, m\in\mathcal{M}, \tag{30e}$$

where

$$\begin{aligned} S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) = \sum_{w\in\mathcal{W}} -\left(u_w^O d_w^O + u_w^C d_w^C + \left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}^D\right) d_w^D\right) \\ - \beta_1^D \sum_{s\in\mathcal{S}} T_s\cdot v_s - \beta_1^D \sum_{s\in\mathcal{S}} \xi_s + \sum_{w\in\mathcal{W}}\sum_{m\in\mathcal{M}} \psi_w^m. \end{aligned} \tag{31}$$

The linear approximation of BL is then given by

$$\underset{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}}{\text{maximize}} \sum_{w\in\mathcal{W}} d_w^C \tag{32a}$$

(BL-A) $$\text{subject to (23b)} - \text{(23d)}, \tag{32b}$$

$$(\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) = \underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}}{\text{argmin}}\ S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}), \tag{32c}$$

$$\text{subject to (30b)} - \text{(30e)}.$$

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

Theorem 1 (**Suboptimality bound**). *Let $\mathbf{x}^*$ and $\bar{\mathbf{x}}$ be optimal solutions to the exact bike lane planning formulation BL and the approximation BL-A, respectively. Then*

$$C(\mathbf{x}^*) - C(\bar{\mathbf{x}}) \leq 2\cdot\mu_1\cdot\sqrt{|\mathcal{W}|}\left(\mu_2\cdot\frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi|-1} + \mu_3\cdot\frac{\sqrt{3\cdot|\mathcal{W}|}}{|\mathcal{R}_\psi|-1}\right), \tag{33}$$

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

From Theorem 1, 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 1 can be made small using a modest number of linear segments. For example, in our empirical study in §5, we use $|\mathcal{R}_\psi| = |\mathcal{R}_\xi| = 15$, which yields an optimality gap in the cycling share of 3.2% (see Appendix D.1 for details).

Finally, to solve BL-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 D.

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 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 to BL-A the constraint

$$\sum_{w\in\mathcal{W}} d_w^D \cdot t_w^D(\mathbf{x},\mathbf{v}) \le (1+\tau)\sum_{w\in\mathcal{W}} \tilde{d}_w^D \cdot t_w^D(\tilde{\mathbf{x}},\tilde{\mathbf{v}}), \tag{34}$$

where $\tilde{d}_w^D$, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{v}}$ are the status quo 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:

$$t_p^D(\mathbf{x},\mathbf{v}) \le (1+\tau)\cdot t_p^D(\tilde{\mathbf{x}},\tilde{\mathbf{v}}), \quad p\in\mathcal{P}^D. \tag{35}$$

The choice of congestion control depends on the priorities of city planners. In particular, constraint (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, we adopt this latter set of constraints, which allows for more granular control of congestion impacts. Note that Theorem 1 applies to the formulation (32), and may no longer hold if either of the constraints (34) or (35) are added.

5. Empirical Study: Expanding Chicago's Cycling Infrastructure

We choose 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 has experienced 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 lanes are typically planned based on current ridership and population density, with little consideration of 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 this 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. More broadly, our study demonstrates how data-driven methods might support city planners in making decisions regarding urban infrastructure.

The remainder of this section is organized as follows. §5.1 provides an overview of the data used in the study and describes how the data enter into our modeling framework. §5.2 specifies the traffic congestion model and presents the estimates of the congestion parameters $\boldsymbol{\theta}$, which are validated through an out-of-sample travel time prediction task. §5.3 specifies the mode choice model, describes its estimation using instrumental variables, and presents estimates of the mode choice parameters $\boldsymbol{\beta}$. §5.4 presents the results from the bike lane optimization model, which are estimates of the change in cycling ridership and traffic congestion resulting from our model's prescriptions. Lastly, §5.5 validates our method through a comparison with three intuitive benchmark methods for bike lane planning.

5.1. Data Description and Sources

Our study focuses on a contiguous region of downtown Chicago, depicted as the shaded area in Figure 1. We select this region due to the availability of detailed bike share trip 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 1), so to maximize the data available to the estimation procedure. Next, we provide an overview of the data used in our study and their sources.

1. Road network topology and features. To construct the road network, we first retrieve the locations of 801 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.

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). These characteristics are used to construct the feature vector $\mathbf{q}_s$ for each

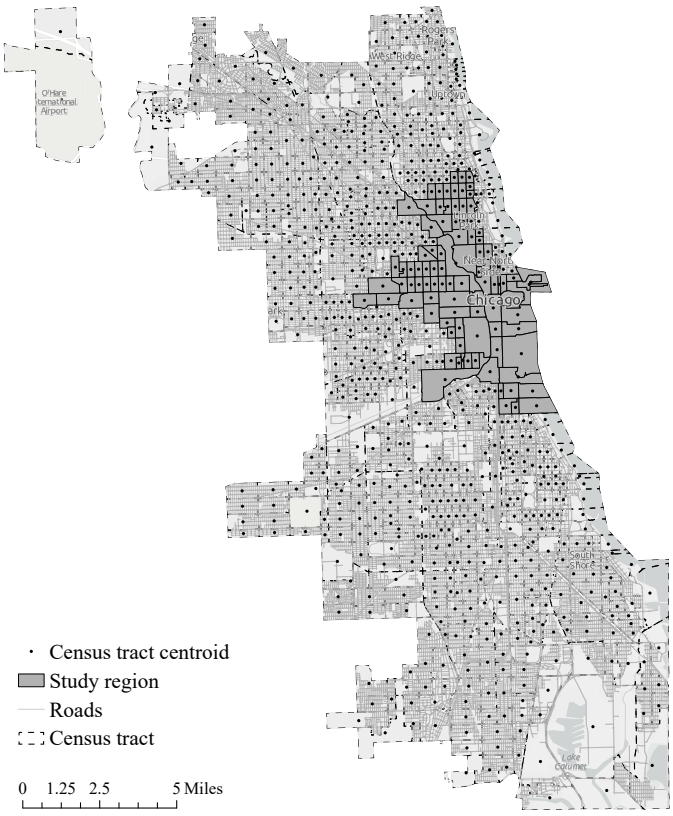


Figure 1 Study region: City of Chicago.

road segment, which is an input to the estimation procedure described in §3.1. Segment length l_s and number of lanes n_s were extracted from OpenStreetMap. The lengths of the road segments are primarily determined by the city's well-defined block pattern (City of Chicago Planning and Policy Division 2003). Lane widths w_s 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 throughout the network from the Highway Performance Monitoring System (HPMS) (FHA 2018), which correspond to the observed flows $\mathbf{z}$ in the estimation procedure described in §3.1. 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. We examine morning rush hours

because they represent a sizeable share of all traffic (24% of all vehicle flows, Illinois Department of Transportation (2018)), and commuters are the main beneficiaries of improvements to cycling infrastructure (Chicago DOT 2020). We match the road segments in the HPMS dataset onto the OpenStreetMap 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 (out of 31,815).

We note here that the vehicle flow data are annual averages for each hour of the day, and that we do not observe day-to-day variations in traffic. While daily vehicle flows would provide useful variation, the use of annual average flows is well-suited for our purposes because of our focus on urban infrastructure planning, which are long-term and usually permanent decisions. Further, we focus on traffic congestion during morning rush hours, which is primarily commuter traffic and thus relatively stable.

4. Travel mode choices. To construct the mode-specific demands d_w^m for each OD pair $w \in \mathcal{W}$, we first obtain commuter flow data from the Origin-Destination Employment Statistics dataset maintained by the Longitudinal Employer-Household Dynamics (LEHD) (US Census Bureau 2018). The dataset specifies total commuting flows between pairs of census tracts, which correspond to the total commute demand $\tilde{d}_w$ in the mode choice model. We only focus on OD pairs with a strictly positive number of commuters and where the origin and destination census tracts are distinct.[14] This resulted in a total of 146,847 OD pairs in our study region with strictly positive demand (out of a possible $801 \times 800 = 640{,}800$ OD pairs, where 801 is the number of census tracts), denoted by $\mathcal{W}$.

We retrieve census tract-level mode shares from the 2014 - 2019 American Community Survey (ACS) (US Census Bureau 2020), which specifies mode shares of driving, cycling, and all other options for commuters residing within each census tract. Multiplying the mode shares with the total commuter volume provides an estimate of the total demand for each mode of transport *originating* at each census tract. Because the census data only defines mode demands at the origin of each OD pair, we determine OD demands for driving (d_w^D) by imputing destinations jointly while estimating $\boldsymbol{\theta}$. Appendix F presents experimental results using synthetic data that show this procedure generates accurate estimates of $\boldsymbol{\theta}$.

5. Bike share trips. We computed the cycling demand d_w^C for each OD pair w using cycling trip data made available by Chicago's bike share system Divvy (Divvy 2021). The Divvy trip data contains information on individual bike rentals, including timestamps indicating the beginning and

[14] Within-census tract trips account for approximately 7% of all commuter trips. These trips cannot be accounted for in our model because they correspond to a distance of zero. With respect to cycling specifically, approximately 6.2% (47,919 out of 791,034) Divvy trips are made within census tract, with the median cycling time being 4.5 minutes.

end of each rental, and the addresses of the bike share stations where each rental was taken from and returned to. We selected all trips taken in 2018 during the morning rush hour period (6am to 10am), resulting in a total of 791,034 bike share trips. We geocoded the addresses of all 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. Then for each origin census tract, we multiplied the destination trip-shares obtained from Divvy data with the total cycling demand originating at each census tract (obtained from the 2014 - 2019 ACS commute data described above) to determine the OD cycling demand d_w^C. A visualization of all Divvy stations in our study region is presented in Figure in Appendix B.[15]

6. Demographics and points of interest. Our mode choice model includes several covariates corresponding to demographics and characteristics of the urban environment. In particular, for each census tract $i \in \mathcal{I}$, we retrieved the median household income (hi_i), the vehicle ownership rate (i.e., fraction of households with at least one vehicle) (vo_i) and population density (pd_i), all obtained from the 2014 - 2019 ACS data. We also queried the Google Places API to identify "points of interest" in each census tract, including parks, restaurants, shopping malls, supermarkets, bakeries, bars, universities, hospitals, and libraries. The presence of each of these points of interest at the origin and destination of OD pair w are included as covariates in the mode choice model using a dummy variable, and are collectively represented by the binary vector $\mathbf{pi}_w$. Further, we also obtain the number of public transportation stations (subways, buses, trains, and light rails) in each census tract by querying the Google Places API. We include the total count of transit station in the origin census tract of each OD pair w, denoted by ts_w, as a covariate in the outside option utility (see §5.3).

7. Driving routes, cycling routes, and public transit time. 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 Directions API (Google Maps 2021). Specifically, we constructed $\mathcal{P}_w^D$ by obtaining the top three 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.[16] In addition to providing routes, the Google Directions 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^D, $s \in \mathcal{S}$, along with the public transit time t_w^O.

[15] Our computation of d_w^C makes the assumption that for each origin census tract, the distribution over destination census tracts is the same for Divvy and non-Divvy cyclists; violations of this assumption would introduce errors into our results. However, this is less likely to be a strong assumption in our setting given that our study region contains a high concentration of Divvy stations, as shown in Figure 8 in Appendix B.

[16] Our assumption that all commuter demand begins and ends at the centroid of a census tract is an approximation of true commute patterns. For context, the average radius of a census tract in the full area of Figure 1 is 500 meters, and the mean and median straight-line distances between OD pairs with positive demand are 9,062 and 7,868 meters, respectively.

Similarly, for each OD pair w, we determined the cycling path p_w^C and the corresponding cycling time t_w^C by retrieving the topmost recommended cycling path from the Google Directions API. We then constructed $\mathcal{S}_w^C$ (the set of road segments that constitute the cycling path p_w^C) by matching the recommended cycling route with the road network. This matching process also allowed us to compute the bike lane coverage $\rho_w(\tilde{\mathbf{x}})$ for each path p_w^C.

To capture the effect of elevation changes on cycling utility, we also determined the *average elevation change* (aec_w) along each cycling path p_w^C, $w \in \mathcal{W}$, which is a measure of the total vertical distance, either uphill or downhill, that must be traversed by cyclists on the cycling path. For each path w, we computed elevation changes by dividing each cycling path into 99 equally-distant segments and querying the altitude at each end point from the Google Elevation API, which generated the elevation vector $[\text{el}_{w,1}, \cdots, \text{el}_{w,100}]$. The average elevation change for path w is then given by $\text{aec}_w = \frac{\sum_{i=1}^{99} |\text{el}_{w,i+1} - \text{el}_{w,i}|}{99}$, given in units of meters.

8. Taxi trips. We validate our estimates of the congestion parameter $\boldsymbol{\theta}$ using taxi trip data from Chicago. These records report the origin and destination census tracts, and duration of all taxi trips made within city boundaries (Chicago Data Portal 2021). For consistency with the traffic flow data described above, we selected all taxi trips taken during morning rush hours (6am to 10am) on weekdays in 2018 between our set of OD pairs $\mathcal{W}$, resulting in 1,408,833 trips between 3,698 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.

5.2. Estimation and Validation of Congestion Parameter $\boldsymbol{\theta}$

We now describe the estimation of the congestion parameters $\boldsymbol{\theta}$, beginning by specifying the congestion functions $\alpha_s(\mathbf{x}, \boldsymbol{\theta})$. First, for each road segment $s \in \mathcal{S}$, we construct the feature vector

$$\mathbf{q}_s(x_s) = \begin{cases} \left(1, \ell_s, n_s \cdot w_s, \frac{\ell_s}{n_s \cdot w_s}\right), & \text{if } x_s = 0, \\ \left(1, \ell_s, n_s \cdot w_s - \Delta w, \frac{\ell_s}{n_s \cdot w_s - \Delta w}\right), & \text{if } x_s = 1, \end{cases} \tag{36}$$

where ℓ_s is the segment's length, n_s is the number of lanes, and w_s is the width of each lane. 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

$$\alpha_s(x_s, \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{q}_s(x_s) = \theta_0 + \theta_1 \cdot \ell_s + \theta_2 (n_s \cdot w_s - \Delta w \cdot x_s) + \theta_3 \left(\frac{\ell_s}{n_s \cdot w_s - \Delta w \cdot x_s}\right). \tag{37}$$

Following the model outlined in §2.2, the segment travel time is then

$$t_s^D(x_s, \boldsymbol{\theta}, v_s) = \alpha_s(x_s, \boldsymbol{\theta}) \cdot v_s + T_s, \tag{38}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)$ is the parameter vector to be estimated.[17] As noted in §5.1, only the driving demand at the origin of each OD pair is available from the American Community Survey data. We therefore slightly modify WLS-A by jointly imputing the driving demand d_w^D with the congestion parameter $\boldsymbol{\theta}$ (see Appendix A for details). In Appendix F, 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 ($|\mathcal{S}^{\text{obs}}| = 12,447$), 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 estimation problem 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 time of taxi trips in Chicago (see §5.1).[18] In particular, we matched each of the 1,408,833 taxi trips that began between 6am and 10am in 2018 to an OD pair based on the trip's start and end locations. We then selected all OD pairs for which there was at least one trip and obtained 3,698 OD pairs. We measure predictive performance by comparing the model's predicted travel time with the average taxi trip duration for each of these OD pairs.

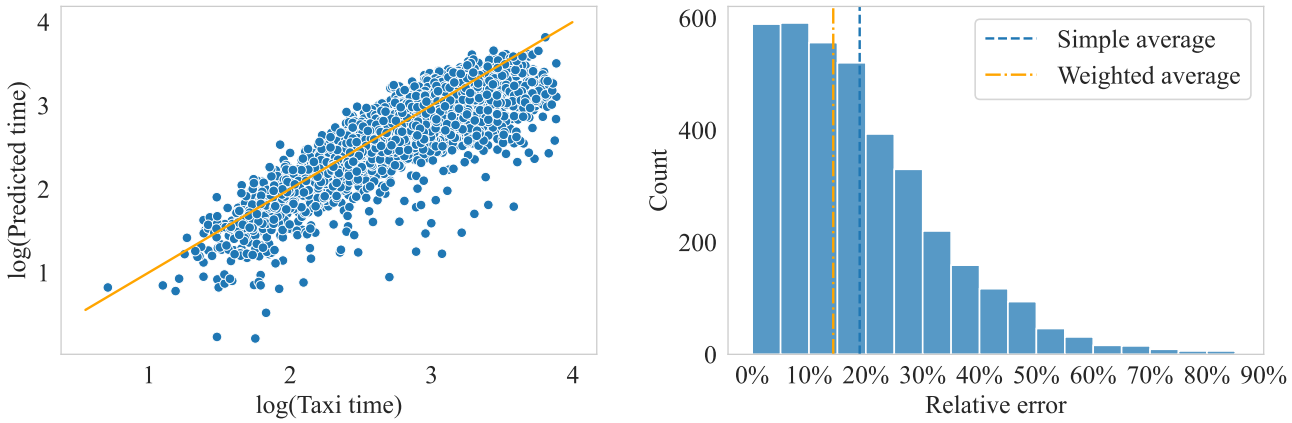
Among the values of λ tested, we found $\lambda = 2500$ to yield the minimal prediction error in taxi travel times, with an average absolute error of 1.34 minutes and an average relative error weighted by number of taxi trips of 14% (see Appendix A.4 for additional details on the selection of λ). Figure 2 depicts predictive performance on the taxi trip data. In Figure 2(a), each dot is an OD pair, and the correlation is 2 is 0.85, suggesting a sensible model fit.[19] Figure 2(b) shows the distribution of prediction errors, with and without weighting each OD pair by the number of taxi trips for that pair.

Table 1 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. An important caveat to the estimates in Table 1 is that there may exist unobserved variables (e.g., speed limits, road conditions) that

[17] Although our specification for the congestion function $\alpha_s(x_s, \boldsymbol{\theta})$ in (37) depends on physical features of the road network that are likely to influence driving times (i.e., length and width), our approach emphasizes predictive performance and should not be interpreted as a structural model.

[18] 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. For this reason, we use the taxi trips as our out-of-sample validation dataset.

[19] Our method'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 suggests traffic congestion has a substantial impact on travel times.



(a) Predicted vs. actual taxi trip times.

(b) Distribution of prediction errors.

Figure 2 Driving time prediction errors on out-of-sample taxi trip data.

influence travel time and are also correlated with road features. The predictions in our counterfactual study should be interpreted in light of this potential bias.

Table 1 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

5.3. Estimation of Mode Choice Parameters $\boldsymbol{\beta}$

Next, we estimate the mode choice parameters $\boldsymbol{\beta}$ using the estimated congestion parameter $\boldsymbol{\theta}$ as input. In estimating $\boldsymbol{\beta}$, we restrict attention to the sub-network for which Divvy trip data is available (shaded region of Figure 1), which corresponds to 5,272 OD pairs, 6,101 road segments, and 14,632 driving paths. We use the following specification for the utility of driving, cycling and outside options on OD pair w:

$$u_w^D(\mathbf{x}, \boldsymbol{\theta}) = \beta_0^D + \beta_1^D \cdot t_w^D(\mathbf{x}, \boldsymbol{\theta}) + \beta_2^D \cdot \mathrm{vo}_w + \beta_3^D \cdot \mathrm{hi}_w + \boldsymbol{\beta}_4^D \cdot \mathbf{pi}_w + \xi_w^D + \epsilon_w^D \tag{39a}$$

$$u_w^C(\mathbf{x}) = \beta_0^C + \beta_1^C \cdot t_w^C + \beta_2^C \cdot \rho_w(\mathbf{x}) + \beta_3^C \cdot \mathrm{aec}_w + \boldsymbol{\beta}_4^C \cdot \mathbf{pi}_w + \xi_w^C + \epsilon_w^D \tag{39b}$$

$$u_w^O = \beta_1^O \cdot t_w^O + \beta_2^O \cdot \mathrm{ts}_w + \epsilon_w^O \tag{39c}$$

The driving utility $u_w^D(\mathbf{x}, \boldsymbol{\theta})$ depends on the driving time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$, vehicle ownership rate at the origin vo_w, the average median household income over the origin and destination census tracts hi_w, and a dummy variable vector of the presence of various points of interest at the origin and destination $\mathbf{pi}_w$, following Kabra et al. (2020). The cycling utility $u_w^C(\mathbf{x})$ depends on the cycling

time t_w^C, bike lane coverage $\rho_w(\mathbf{x})$, the average elevation change aec_w along the cycling route p_w^C, and the dummy vector for points of interest $\mathbf{pi}_w$. In addition, we assume the outside option utility depends on the travel time by public transit t_w^O and the number of public transit stations (subways, buses, trains and light rail) in the origin census tract, denoted by ts_w. Table 2 presents the summary statistics of the data used in the mode choice model.

Table 2 Summary of data for estimating mode choice parameters β.

		Units	Mean	SD	Min.	25%	Median	75%	Max.
Cycling time	t_w^C	minutes	21.37	10.69	0.98	13.31	20.87	28.50	59.13
Bike lane coverage	ρ_w	%	50.82	25.97	0.00	31.55	54.05	71.93	100.0
Average elevation change	aec_w	meters	0.49	0.06	0.01	0.45	0.49	0.53	0.81
Predicted driving times	t_w^D	minutes	14.14	4.62	3.64	10.86	14.15	17.30	30.78
Vehicle ownership rate	vo_w	%	71.08	13.55	25.99	62.10	73.71	81.66	92.22
Public transit travel time	t_w^O	minutes	34.57	16.50	0.00	22.48	33.97	44.62	103.65
No. of transit stations	ts_w	1	17	22	0	3	6	30	127
Median household income	hi_w	10^4 dollars	9.81	2.56	1.82	8.07	9.98	11.64	17.22

5.3.1. Endogeneity and instruments. As discussed in §2.3, because we observe mode shares for each OD pair w, the parameters in (39a)–(39c) can be estimated using linear regression. However, naively applying ordinary least squares (OLS) estimation may lead to biased estimates due to two potential sources of endogeneity, namely, in the bike lane coverage $\rho_w(\mathbf{x})$ and the driving travel time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$. Before discussing the endogeneity issues and our solution via instrumental variables, we first present the OLS estimates in Table 3.[20] As the table shows, the bike lane coverage estimate β_2^C is statistically insignificant and is not informative for capturing bike lanes' value to cyclists. Below, we briefly describe this endogeneity problem and our use of instrumental variables to address it.

Table 3 Estimates (and standard errors) of mode choice parameters using OLS.

Cycling				Driving			
intercept β_0^C	cycling time β_1^C	bike lane coverage β_2^C	average elevation change β_3^C	intercept β_0^D	driving time β_1^D	vehicle ownership β_2^D	household income β_3^D
2.1559	−0.2081***	−0.1425	−2.0991***	4.3919***	−0.2452***	0.4860	0.0265
(1.71)	(0.00)	(0.15)	(0.66)	(1.69)	(0.01)	(0.31)	(0.02)

Outside option					
transit time β_1^O	transit stations β_2^O	Other control variables	Adjusted R^2	Number of observations	
−0.0231***	0.0052***				
(0.00)	(0.00)	Table 12	0.57	10,544	

[20] Note that the number of observations in the regression model is twice of the number of OD pairs, since there are 5,272 observations for both cycling and driving.

Although bike lane coverage may influence cycling demand on each OD pair, it is plausible that city planners select locations for new bike lanes to serve unmet demand (i.e., along paths where cycling ridership is already high but which lack bike lanes), which would lead to correlation between the error term ξ_w^C and bike lane coverage $\rho_w(\mathbf{x})$. Ignoring this potential reverse causality in the location of existing bike lanes would lead to an overestimation of β_1^C, and thus inflated counterfactual predictions of cycling ridership following the addition of new bike lanes. Conversely, it is also plausible that planners choose locations with low ridership in order to promote cycling adoption, which would lead to underestimation of β_1^C. Similarly, with respect to driving travel times, there may exist unobserved factors that lead to low (high) demand for driving on a particular OD pair, resulting in a lower (higher) driving travel time and leading to correlation between the error terms ξ_w^D and driving the travel time $t_w^D(\mathbf{x}, \boldsymbol{\theta})$.

To address the aforementioned endogeneity biases, we employ BLP-like instrumental variables (Berry et al. 1995) that are tailored for analyzing spatially differentiated products (e.g., Davis (2006), Kabra et al. (2020), He et al. (2021)). In particular, our instruments are inspired by Kabra et al. (2020), who use exogenous characteristics of nearby bike-share stations (e.g., points of interest) as instruments for estimating the effect of bike-share availability on usage at a focal bike-share station.

Accordingly, for each OD pair w, we construct instruments using the exogenous characteristics of "nearby" OD pairs. Specifically, for each w, let $\mathcal{Z}_w{}^{(a,b)}$ to be the set of all other OD pairs whose origin is between distance a and b from the origin of w and whose destination is within distance between a and b of the destination of w. Note that each OD pair w is a pair of nodes (i.e., census tracts) (i, j) on the network, and let $\text{dis}(i, j)$ be the Euclidean distance between nodes i and j. Formally, $\mathcal{Z}_w^{(a,b)}$ is then defined as

$$\mathcal{Z}_{w=(i,j)}^{(a,b)} = \{w' = (i', j') \in \mathcal{W} | \text{dis}\,(i', i) \in [a, b]\,, \text{dis}\,(j', j) \in [a, b]\}\,. \tag{40}$$

The instruments for an OD pair w are then given by

$$Z_w^{(a,b)} = \begin{bmatrix} \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \text{pd}_{i'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \text{pd}_{j'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \mathbf{npi}_{i'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \mathbf{npi}_{j'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} (\mathbf{1}^T \cdot \mathbf{npi}_{i'}) \times \text{pd}_{i'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \left(\mathbf{1}^T \cdot \mathbf{npi}_{j'}\right) \times \text{pd}_{j'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \text{hi}_{i'} \\ \frac{1}{\left|\mathcal{Z}_{w=\{i,j\}}^{(a,b)}\right|} \sum_{w'=\{i',j'\} \in \mathcal{Z}_{w=\{i,j\}}^{(a,b)}} \text{hi}_{j'} \end{bmatrix}, \tag{41}$$

where pd_i, $\mathbf{npi}_i$, and hi_i denote the population density, a vector that tracks the number of each type of point of interest, and median household income in census tract i, respectively. Note $(\mathbf{1}^T \cdot \mathbf{npi}_i) \times \text{pd}_i$ is the interaction between the total number of points of interest of all categories and population density.

To be valid, the instruments should be correlated with the endogenous variables, which in our case are bike lane coverage and driving travel time (i.e., satisfy the relevance condition), and should be uncorrelated with the unobserved components ξ_w^m in the utility of the corresponding mode (i.e., satisfy the exclusion criterion). The validity of our proposed instruments follow similarly from the arguments in Berry et al. (1995), Davis (2006), Kabra et al. (2020), and He et al. (2021), which we concisely summarize here for our context. In particular, relevance follows because demographic characteristics and features of the urban environment are likely to be spatially correlated, making it plausible for them to also be correlated with bike lane coverage and driving travel times in nearby OD pairs. Satisfaction of the exclusion restriction follows because the instruments are relatively static characteristics of the study region (e.g., points of interest). As a result, for a focal OD pair, these static characteristics of nearby OD pairs are unlikely to vary in response to unexplained realizations of cycling or driving demand in the focal OD pair. Similarly, these characteristics of nearby OD pairs are also unlikely to directly influence demand for cycling or driving in the focal OD pair, because commuters in the focal OD pair are presumably unexposed to them during their commute.

Similar to He et al. (2021), as a robustness check we consider multiple distance intervals $[a, b]$ when constructing the instrumental set of OD pairs $\mathcal{Z}_w^{(a,b)}$. In particular, we vary $a \in \{500, 1000, 1500\}$ meters and $b \in \{7000, 8000\}$ meters to obtain six different distance intervals $[a, b]$. The distance 500, 1000, 1500, 7000, and 8000 correspond to the 2^{nd}, 6^{th}, 11^{th}, 84^{th}, 90^{th} percentiles of the between-centroid distances of the corresponding census tracts, respectively. We estimate our model using the standard approach of generalized method of moments (GMM), and find the results to be robust to the choice of a and b (see Appendix C). We verify relevance of all instruments by performing the weak-instrument test proposed by Stock and Yogo (2002).

As an additional robustness check, we estimate our model with two alternative instruments. First, we instrument for bike lane coverage and driving travel time using the average value of these same endogenous variables in nearby OD pairs, in the spirit of Hausman-type instruments (Hausman 1996, Nevo 2001). Second, we follow He et al. (2021) and Arora et al. (2022), who construct instruments that are similar to the BLP-like instrument described above, but explicitly utilize the network structure. We find that our estimates are stable across these alternative instruments as well (see Appendix C).

5.3.2. Results. Table 4 reports the estimation results corresponding to $[a, b] = [500, 8000]$, which produces the most conservative (i.e., smallest) estimate of the bike lane coverage coefficient β_2^C among the distance intervals $[a, b]$ we implemented (see Table 10). We focus on the most conservative estimate because it is the value used in the bike lane planning problem described in §5.4 below; Appendix C contains the estimates for the remaining values of a and b, as well as the alternative instruments. Applying the weak instrument test by Stock and Yogo (2002) using $[a, b] = [500, 8000]$ yields a Cragg–Donald statistic of 21.6, which exceeds the critical value of 20.7.[21]

The signs of the estimates shown in Table 4 agree with intuition: The negative coefficients for the travel time by cycling (β_1^C), driving (β_1^D) and public transit (β_1^O) indicate shorter travel times increase the attractiveness of each transportation mode. For cycling specifically, the the results suggest the attractiveness of cycling increases in bike lane coverage ($\beta_2^C > 0$), and that cyclists prefer flatter paths with fewer elevation changes ($\beta_3^C < 0$). Similarly, driving demand is positively correlated with vehicle ownership rate ($\beta_2^D > 0$). Lastly, improved accessibility to transit stations is associated with higher demand for the outside option ($\beta_2^O > 0$). Lastly, the estimated mode choice model yields $R^2 = 0.57$, which suggests a reasonable fit to the data.

Table 4 Estimates (and standard errors) of mode choice parameters from IV estimation.

Cycling				Driving			
intercept β_0^C	cycling time β_1^C	bike lane coverage β_2^C	average elevation change β_3^C	intercept β_0^D	driving time β_1^D	vehicle ownership β_2^D	household income β_3^D
2.3614	−0.2189***	1.8817***	−2.6596***	3.9695**	−0.2088***	0.5964***	0.0291**
(2.43)	(0.01)	(0.79)	(0.64)	(1.55)	(0.04)	(0.22)	(0.01)

Outside option				
transit time β_1^O	transit stations β_2^O	Other control variables	Adjusted R^2	Number of observations
−0.0304***	0.0040***			
(0.01)	(0.00)	Table 12	0.57	10,544

5.4. Bike Lane Expansion: Impact on Cycling Ridership and Traffic Congestion

Next, we use the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ to develop bike lane plans for downtown Chicago and estimate the associated impacts on 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 bike lane optimization model over the largest OD pairs that account for 80% of commute demand in the study region (shaded area of

[21] The critical value of 20.7 corresponds to 24 instruments, two endogenous regressors, and a relative bias level of 0.05 – see Stock and Yogo (2002) for details.

Figure 1). This resulted in 887 candidate paths for bike lane installation, which together comprise 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 (35) to the path selection model BL-A given in (32), which is solved to obtain a recommended network $\mathbf{x}^*$. We then evaluate the model's recommendation by solving the (exact) user equilibrium problem (24) under $\mathbf{x}^*$ and computing the relative increase in cycling ridership over the status quo network $\tilde{\mathbf{x}}$:

$$\frac{\sum_{w \in \mathcal{W}} d_w^C(\mathbf{x}^*) - d_w^C(\tilde{\mathbf{x}})}{\sum_{w \in \mathcal{W}} d_w^C(\tilde{\mathbf{x}})}. \tag{42}$$

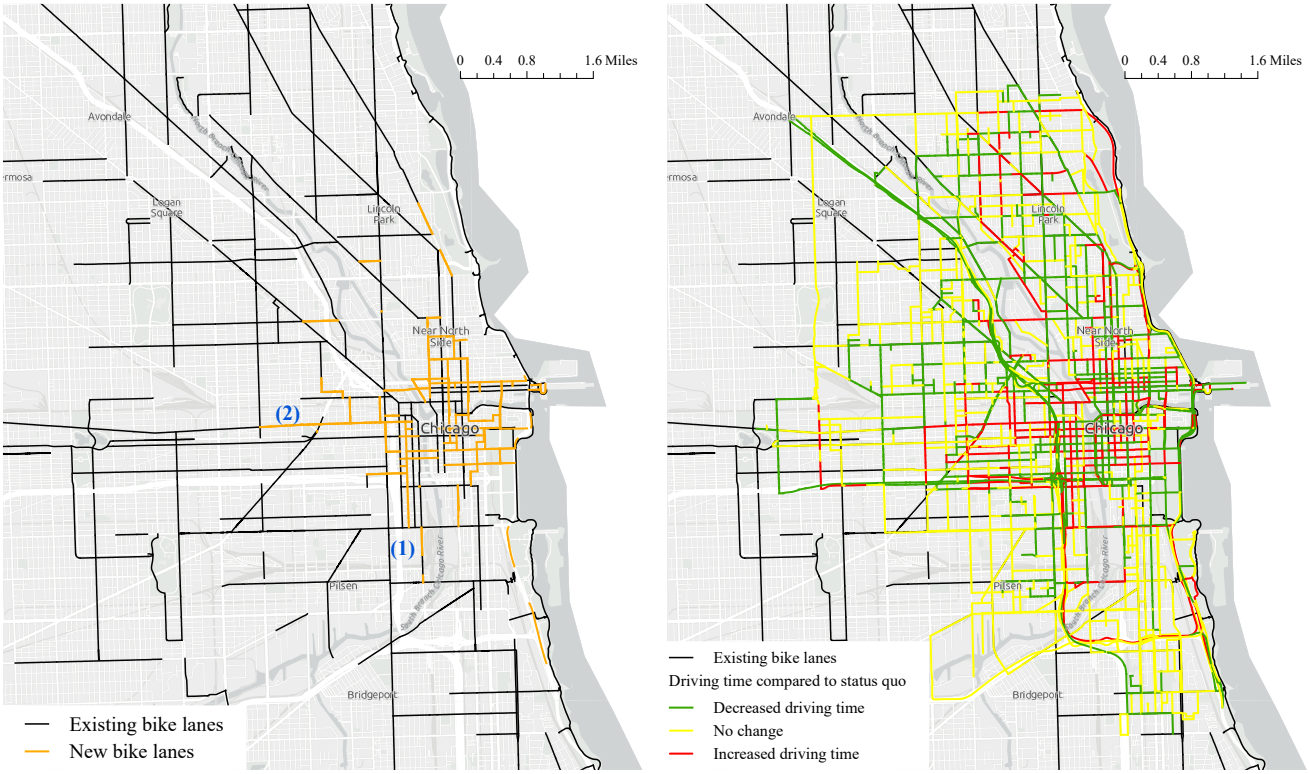
Consistent with the model BL-A, we evaluate congestion using the worst-case increase in driving times over all driving paths:

$$\max_{p \in \mathcal{P}^D} \left\{ \frac{t_p^D(\mathbf{x}^*) - t_p^D(\tilde{\mathbf{x}})}{t_p^D(\tilde{\mathbf{x}})} \right\}. \tag{43}$$

Our choice of the congestion constraint (35) (and the corresponding performance metric (43)) 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 E.

As an illustrative example, Figure 3 visualizes the model's recommended bike lane plan and the predicted impacts on driving times for the case where $(B, \tau) = (25, 10\%)$. Figure 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 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 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, 82.1% have longer driving times and 7.0% have shorter driving times (no change is observed in the remaining 10.9% 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 capacity effect". On the other hand, building bike lanes increases cycling utility and thus boosts cycling adoption, which reduces driving demand and thus alleviates congestion, which we refer to as the "demand effect". Figure 3b visualizes the segments where the road capacity effect dominates the demand effect (red) and vice versa (green).



(a) Bike lane expansion.

(b) Change in driving times.

Figure 3 **Recommended bike lane expansion and congestion impacts in downtown Chicago for** $(B, \tau) = (25, 10\%)$**.**

We also observe spillover effects of new bike lanes to the entire network. Specifically, among the road segments on which no new bike lanes are built, 13.4% have longer driving times and 40.1% have shorter driving times. Because the demand effect spills over to the rest of the network, the reduction in driving times is expected. The congestion-increasing road capacity 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 summarizes model performance under each of the 12 combinations of (B, τ). In general, we predict notable increases to cycling ridership even when the congestion tolerance and the budget are minimal – for example, under $(B, \tau) = (10, 5\%)$, cycling shares increase by 1.3% (from 3.6% to 4.9%). Further, our results suggest diminishing returns on increasing the congestion tolerance τ – increasing τ from 5% to 10% boosts cycling ridership by 0.6% to 1.2%, depending on the bike lane mileage budget B, wheres increasing τ from 10% to 15% is associated with at most a 0.6% absolute increase in cycling ridership. We observe a similar behavior as B increases, with most of the gains in cycling ridership occurring within the first 50 miles of bike lane expansion.

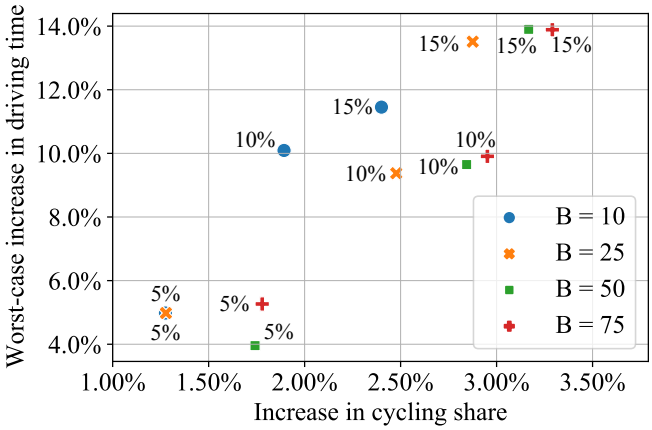


Figure 4 **Increase in cycling ridership and worst-case driving travel times under 12 combinations of (B, τ).**

5.5. Comparison with Alternative Bike Lane Planning Methods

To further evaluate the bike lane plans recommended by our model, we benchmark their performance against three alternative methods for bike lane planning. We briefly describe each method below and provide their precise steps in Appendix E.

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, $t_s^D(\tilde{\mathbf{x}})$, which are returned by the estimation method described in §3. We then solve a variant of the path selection model BL-A with these fixed driving times.

Greedy heuristic. We also present a greedy heuristic for solving BL-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 main steps. First, we sort the OD pairs 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 heuristic 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.

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).

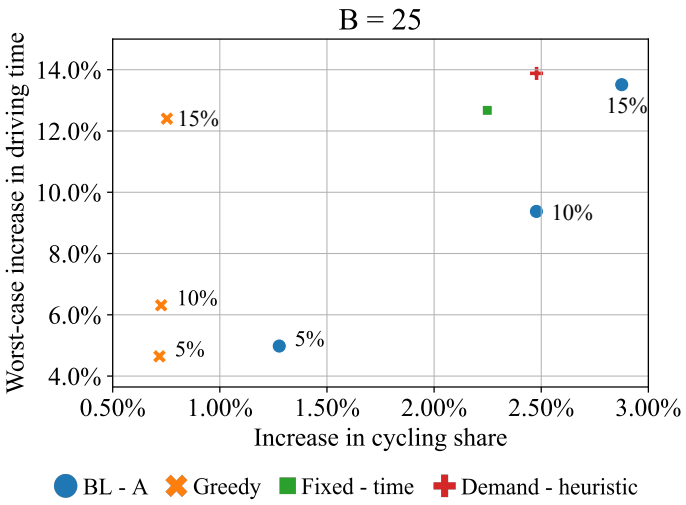


Figure 5 **Increase in ridership and worst-case driving time for bike lane path selection model BL-A and three benchmark methods.**

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 BL-A with the three benchmark methods. For brevity, we present the detailed results for $B = 25$ miles in Figure 5, and defer the cases for $B \in \{10, 50, 75\}$ to Appendix E.2. Note that the fixed-time model and demand heuristic are not parameterized by τ, so they each correspond to only one point in Figure 5. We interpret these results by first comparing our model's prescriptions (blue '•') 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 yields the highest increase in cycling mode share (2.5%), although the resulting increase in congestion is relatively high (13.9% increase in worst-case driving time). In particular, the demand heuristic is dominated by the recommendations from BL-A when $\tau = 10\%$ or $\tau = 15\%$. With respect to the fixed-time model (green '■'), it increases the cycling mode share by 2.2% (from 3.6% to 5.8%), yet also significantly increases congestion (12.7%). To that end, the fixed-time model is also dominated by the bike lane expansion recommended by our model under $\tau = 10\%$. This difference underscores the importance of endogenizing congestion effects in bike lane planning, which permits city planners to modulate driving times throughout the network. Finally, we examine the greedy approach (orange '×') that considers congestion effects 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 increase in the cycling mode share under our method is more than twice that of the greedy heuristic (2.9% vs. 0.8%), while leading to comparable increases in driving times (13.5% vs 12.4%).

As an additional point of reference, we also compared our model's prescriptions with the bike lanes actually installed in Chicago following our study period of 2018. From November 2018 to

November 2021, the City of Chicago added additional 14.48 miles of bike lanes in the downtown Chicago study region (Chicago Data Portal 2023). According to our model, this expansion in the bike lane network corresponds to a predicted 0.4% increase in the cycling mode share and an 8.2% maximum increase in the driving time over all driving paths. By comparison, setting a bike lane budget of $B = 14.48$ miles and a congestion tolerance of $\tau = 5\%$ in our prescriptive model results in an increase of 1.1% in the cycling mode share and a 4.3% maximum increase in the driving time. See Appendix E.3 for details and a brief discussion. Taken together, these results demonstrate the value of endogenizing congestion and ridership effects, as well as taking a holistic and network-based perspective to urban bike lane planning.

6. 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 quantify the potential 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 3.6% to 6.1%, while increasing driving times by at most 9.4% 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 notion 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 the one presented here 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.

Our work 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 the road networks, 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. Although we incorporate factors including points of interest and elevation change along the cycling routes in cycling utility, other factors that we do not consider include safety (in particular, vehicle traffic near bike lanes) and continuity of the bike lane network.

We conclude by noting potential directions for future work. First, although we focused on maximizing cycling ridership, 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 performing counterfactual analyses to support other types of infrastructure planning where congestion impacts are a salient concern, such as the addition of dedicated bus or carpool lanes.

References

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.

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* .

Arora, Kashish, Fanyin Zheng, Karan Girotra. 2022. Private vs. pooled transportation: Customer preference and design of green transport policy .

Aswani, Anil, Zuo-Jun Shen, Auyon Siddiq. 2018. Inverse optimization with noisy data. *Operations Research* **66**(3) 870–892.

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.

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

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.

Bertsimas, Dimitris, John N Tsitsiklis. 1997. *Introduction to linear optimization*, vol. 6. Athena Scientific.

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.

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. 2022. An inverse optimization approach to measuring clinical pathway concordance. *Management Science* **68**(3) 1882–1903.

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. 2021. Inverse optimization: Theory and applications. *arXiv preprint arXiv:2109.03920* .

Charalambos, D, Border Aliprantis. 2013. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer.

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 Data Portal. 2023. Bike routes: City of chicago: Data portal. URL `https://data.cityofchicago.org/Transportation/Bike-Routes/3w5d-sru8`.

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`.

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**.

City of Chicago Planning and Policy Division. 2003. Chicago Central Area Plan: The Physical and Economic Assessment. URL `https://www.chicago.gov/content/dam/city/depts/zlup/Planning_and_Policy/Publications/Central_Area_Plan_DRAFT/03_Central_Area_Plan_Chapter2a.pdf`.

Çolak, Serdar, Antonio Lima, Marta C González. 2016. Understanding congested travel in urban areas. *Nature Communications* **7**(1) 1–8.

Cominetti, Roberto, Valerio Dose, Marco Scarsini. 2021. The price of anarchy in routing games as a function of the demand. *Mathematical Programming* 1–28.

Cooper, Lee. 1993. *Chapter 6 Market-share models*, vol. 5. 259–314. doi:10.1016/S0927-0507(05)80029-5.

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.

Dafermos, Stella C, Frederick T Sparrow. 1969. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards B* **73**(2) 91–118.

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.

Davis, Peter. 2006. Spatial competition in retail markets: movie theaters. *The RAND Journal of Economics* **37**(4) 964–982.

Dempe, Stephan. 2002. *Foundations of bilevel programming*. Springer Science & Business Media.

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`.

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. 2006. Federal Highway Administration University Course on Bicycle and Pedestrian Transportation. URL `https://www.fhwa.dot.gov/publications/research/safety/pedbike/05085/chapt15.cfm`.

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.

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.

Google Maps. 2021. The Directions API overview. URL `https://developers.google.com/maps/documentation/directions/overview`.

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.

Hausman, Jerry A. 1996. Valuation of new goods under perfect and imperfect competition. *The economics of new goods*. University of Chicago Press, 207–248.

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.

Hensher, David A, William H Greene. 2003. The mixed logit model: the state of practice. *Transportation* **30** 133–176.

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-&amp;-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/`.

Kabra, Ashish, Elena Belavina, Karan Girotra. 2020. Bike-share systems: Accessibility and availability. *Management Science* **66**(9) 3803–3824.

Khany, Sam. 2022. Personal communication. *Senior Transportation Planner, City of Vancouver* .

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, Young-Jae, Vukan R Vuchic. 2005. Transit network design with variable demand. *Journal of Transportation Engineering* **131**(1) 1–10.

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.

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* .

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.

Magnanti, Thomas L, Richard T Wong. 1984. Network design and transportation planning: Models and algorithms. *Transportation Science* **18**(1) 1–55.

Malhotra, Naresh K. 1984. The use of linear logit models in marketing research. *Journal of Marketing Research* **21**(1) 20–31.

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 .

National Household Travel Survey. 2017. URL `https://nhts.ornl.gov/`.

Nevo, Aviv. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.

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.

Prashker, Joseph N, Shlomo Bekhor. 2000. Some observations on stochastic user equilibrium and system optimum of traffic assignment. *Transportation Research Part B: Methodological* **34**(4) 277–291.

Prashker, Joseph N, Shlomo Bekhor. 2004. Route choice models used in the stochastic user equilibrium problem: a review. *Transport reviews* **24**(4) 437–463.

Reich, Gregor, Kenneth L Judd. 2020. Efficient likelihood ratio confidence intervals using constrained optimization. *Available at SSRN 3455484* .

Roughgarden, Tim, Éva Tardos. 2002. How bad is selfish routing? *Journal of the ACM (JACM)* **49**(2) 236–259.

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*.

Stock, James H, Motohiro Yogo. 2002. Testing for weak instruments in linear IV regression.

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.

Su, Che-Lin, Kenneth L Judd. 2012. Constrained optimization approaches to estimation of structural models. *Econometrica* **80**(5) 2213–2230.

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.

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 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/`.

Theil, Henri. 1969. A multinomial extension of the linear logit model. *International economic review* **10**(3) 251–259.

Thomas, Roy. 1991. *Traffic assignment techniques*. Avebury Publishing Company.

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`.

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. 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. 2018. 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`.

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* .

Wu, Chien-Fu. 1981. Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics* **9**(3) 501–513.

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.

Yildirimoglu, Mehmet, Osman Kahraman. 2018. Searching for empirical evidence on traffic equilibrium. *PloS one* **13**(5) e0196997.

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.

Appendix

A. Additional Details on Estimation of Congestion Parameter $\boldsymbol{\theta}$

This section presents additional details on the estimation of the congestion parameter. §A.1 provides the conditions for consistency of the estimates. §A.2 provides more details on the data used to estimate and validate the congestion parameter. §A.3 introduces the estimation procedure described in in §5.2. In §A.4, we tune the penalty parameter λ. Finally, we compare our method with two machine learning models in terms of predicting OD driving time in §A.5.

A.1. Consistency of Congestion Parameter Estimates

Here we provide conditions under which the estimation problem (13) is consistent, i.e., the sequence of estimates $\hat{\boldsymbol{\theta}}_n$ converges in probability to the true parameter $\boldsymbol{\theta}_0$. The result follows closely from Theorem 2 of Wu (1981).

Theorem 2. *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 (iii) $\lim_{n\to\infty} \Pr(\sum_{s=1}^{n} (v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0))^2 > 0) = 1$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ as $n \longrightarrow \infty$.*

The proof of Theorem 2 is contained in Appendix H along with the proofs of results from the main body. At the outset, it should be noted that Theorem 2 describes an idealized setting, with conditions that are analogous to those used in classical results regarding consistency of non-linear least squares regression (e.g., Jennrich (1969), Wu (1981)). Condition (i) specifies a data-generating model based on a true, unknown parameter $\boldsymbol{\theta}_0$. Note that condition (i) assumes the observed vehicle flows $\mathbf{z}$ are zero-mean, noisy observations of the true equilibrium flows, which are assumed to arise under Wardrop equilibrium. In practice, the observed vehicle flows $\mathbf{z}$ may depend on other factors as well, including drivers' idiosyncratic preferences over driving routes or unobserved characteristics of drivers or routes. Our model assumes drivers select routes exclusively based on minimizing travel time and no other factors; violations of this assumption would lead to biased estimates of $\boldsymbol{\theta}_0$.

Condition (iii) is an identifiability condition, which ensures that the ground truth parameter $\boldsymbol{\theta}_0$ can be pinpointed from data. Intuitively, the condition states that as the size of the network grows large, any two distinct values of $\boldsymbol{\theta}$ should lead to two distinct equilibrium 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 estimation procedure.

The intuition behind condition (iii) is as follows: For very large networks, perturbing $\boldsymbol{\theta}$ should shift the equilibrium vehicle flow $v_s^*(\boldsymbol{\theta})$ in at least one road segment s. The corollary below provides a slightly stronger but more interpretable condition under which the model is identifiable.

Corollary 1. *Let $\mathcal{X}$ be the set of networks such that for each $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, there exists an OD pair $w \in \mathcal{W}$ with paths p_1 and p_2 that have strictly positive flows in equilibrium, and where*

$$\sum_{s \in \mathcal{S}_{p_1}^D \setminus \mathcal{S}_{p_2}^D} ((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{q}_s) v_s^*(\boldsymbol{\theta}_0) \neq \sum_{s \in \mathcal{S}_{p_2}^D \setminus \mathcal{S}_{p_1}^D} ((\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{q}_s) v_s^*(\boldsymbol{\theta}_0). \tag{44}$$

Then condition (iii) from Theorem 2 holds under any random network process $\{\mathcal{G}_n\}_{n\geq 1}$ *such that* $\lim_{n\to\infty}\Pr(\mathcal{G}_n \in \mathcal{X}) = 1$.

The proof is also contained in Appendix H. At a high level, the intuition behind (44) in Corollary 1 is that for each possible value of $\boldsymbol{\theta}$ where $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, there should exist at least one OD pair $w \in \mathcal{W}$ such that changing $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}$ results in a change in travel time on one path that is not exactly cancelled out by a change in travel time on another path.

Statistical consistency in inverse optimization was first addressed by Aswani et al. (2018). The main difference with our setting 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 2 and Corollary 1 differ slightly from those given in Aswani et al. (2018).

Lastly, the assumption that Θ is finite (condition (*ii*) in Theorem 2) can be weakened to compactness of Θ by suitably modifying the identifiability condition (*iii*).

Corollary 2. *Suppose condition* (*i*) *in Theorem 2 holds and the parameter set* Θ *is compact. If for any* $\delta > 0$, $\lim_{n\to\infty}\Pr\left(\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0))^2 > 0\right) = 1$, *then* $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$ *as* $n \longrightarrow \infty$.

A.2. Data and Summary Statistics

To make full use of the available data, we estimate $\boldsymbol{\theta}$ using data from the entire City of Chicago (full area in Figure 1). Table 5 presents details on the size of the Chicago network. Table 6 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 $\sum_{s\in\mathcal{S}_p^D} T_s$. Figure 6 depicts the empirical distribution of these variables. 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 $\boldsymbol{\theta}$ due to how these parameters enter into the model.

Table 5 Size of Chicago network used to estimate congestion parameter $\boldsymbol{\theta}$.

	Road segments $\lvert\mathcal{S}\rvert$	Driving paths $\lvert\mathcal{P}^D\rvert$	OD pairs $\lvert\mathcal{W}\rvert$
Size	31,815	399,356	146,847

To validate model fit, we use detailed data on 1,408,833 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. We drop records with trip times in the upper and lower 2.5% quantiles to remove extreme observations, which result in taxi travel times in 3,698 OD pairs. A summary of taxi trip times is presented in Table 7. Results from the taxi trip data validation are given in Figure 2 in §5.2.

Table 6 Summary of data for estimating congestion parameter θ.

	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

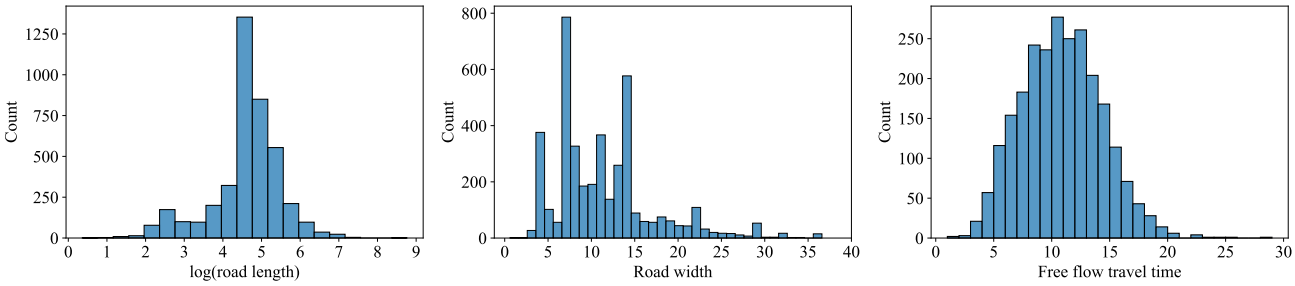


Figure 6 Empirical distributions of road segment lengths (l_s), total segment widths ($n_s \cdot w_s$), and path-level free flow travel times ($T_p = \sum_{s \in \mathcal{S}_p} T_s$).

Table 7 Summary of OD-level taxi trip times.

	Units	Mean	SD	Min	25%	50%	75%	Max
Taxi trip durations (taxi_w)	minutes	17.1	8.1	2.0	11.1	16.0	21.5	48.5

A.3. Estimation via Block Coordinate Descent (BCD) Algorithm

To facilitate solving WLS-A, throughout this section we eliminate the variable ϵ, and focus on the equivalent problem

$$\begin{aligned} &\underset{\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{v},\mathbf{b}}{\text{minimize}} \ \|\mathbf{v}-\mathbf{z}\|_2^2 + \lambda \cdot \left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b}\right) \\ &\text{subject to } (14c)-(14f). \end{aligned}$$

Let $d_i^{(o)}$ be the driving demand originating from each census tract $i \in \mathcal{I}$. Then any vehicle traffic pattern $(\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v})$ should satisfy

$$\sum_{\substack{j \in \mathcal{I} \\ j \neq i}} \sum_{(i,j) \in \mathcal{W}} d_w^D = d_i^{(o)}, \quad i \in \mathcal{I}. \tag{46}$$

Define $\Lambda = \{(\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v}) \,|\, (\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v})$ satisfies (46) 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 $(\mathbf{d}^D, \boldsymbol{\phi}, \mathbf{v}) \in \Lambda$.

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 WLS-A for jointly estimating the driving demands $\mathbf{d}^D$ and the congestion parameter $\boldsymbol{\theta}$ is

$$\underset{\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\theta},\mathbf{b}}{\text{minimize}} \ \|\mathbf{v}-\mathbf{z}\|_2^2 + \lambda \cdot \left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b}\right) \tag{47a}$$

$$\text{subject to} \sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w^D\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right), \quad p\in\mathcal{P}, \tag{47b}$$

$$(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}) \in \Lambda, \tag{47c}$$

$$\boldsymbol{\theta}^\top \mathbf{q}_s(\tilde{\mathbf{x}}) \geq \delta, \quad s\in\mathcal{S}, \tag{47d}$$

$$\boldsymbol{\theta} \in \Theta. \tag{47e}$$

Formulation (47) is multi-convex in the sense that fixing $(\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v})$ 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 $(\boldsymbol{\theta}^k,\mathbf{b}^k)$ be the variable values after their k^{th} update. The BCD method then consists of iteratively solving the following two subproblems:

$$\begin{bmatrix}\left(\mathbf{d}^D\right)^k\\ \left(\boldsymbol{\phi}\right)^k\\ \left(\mathbf{v}\right)^k\end{bmatrix} = \underset{\mathbf{d}^D,\boldsymbol{\phi},\mathbf{v}}{\operatorname{argmin}} \ \|\mathbf{v}-\mathbf{z}\|_2^2 + \lambda\cdot\left(\left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}^{k-1}\right)\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b}^{k-1}\right) + \frac{1}{2\mu_1^k}\left\|\begin{bmatrix}\mathbf{d}^D\\ \boldsymbol{\phi}\\ \mathbf{v}\end{bmatrix} - \begin{bmatrix}\left(\mathbf{d}^D\right)^{k-1}\\ \left(\boldsymbol{\phi}\right)^{k-1}\\ \left(\mathbf{v}\right)^{k-1}\end{bmatrix}\right\|^2 \tag{48a}$$

$$\text{subject to} \sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w^D\}} b_w^{k-1} \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}^{k-1}\right), \quad p\in\mathcal{P} \tag{48b}$$

which is a quadratic program (QP), and

$$\begin{bmatrix}\boldsymbol{\theta}^k\\ \boldsymbol{b}^k\end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{b}}{\operatorname{argmin}} \ \|(\mathbf{v})^k-\mathbf{z}\|_2^2 + \lambda\cdot\left(\left(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)\big|_{\boldsymbol{\phi}=(\boldsymbol{\phi})^k}\right)^\top (\boldsymbol{\phi})^k - \mathbf{b}^\top \left(\mathbf{d}^D\right)^k\right) + \frac{1}{2\mu_2^k}\left\|\begin{bmatrix}\boldsymbol{\theta}\\ \mathbf{b}\end{bmatrix} - \begin{bmatrix}\boldsymbol{\theta}^{k-1}\\ \mathbf{b}^{k-1}\end{bmatrix}\right\|^2 \tag{49a}$$

$$\text{subject to} \sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)\big|_{\boldsymbol{\phi}=(\boldsymbol{\phi})^k}, \quad p\in\mathcal{P} \tag{49b}$$

$$\boldsymbol{\theta}^\top \mathbf{q}_s(\tilde{\mathbf{x}}) \geq \delta, \quad s\in\mathcal{S}, \tag{49c}$$

$$\boldsymbol{\theta} \in \Theta, \tag{49d}$$

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 $G_k = \|\mathbf{v}-\mathbf{z}\|_2^2 + \lambda\cdot(\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi},\boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b})$ be the loss after solving (49) 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 (47) when $\lambda = 0$ and $\boldsymbol{\theta}^0$ is held fixed (in which case (47) 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.

A.4. 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,[22] and (3) out-of-sample error[23]. The reason for evaluating model fit according to these three criteria is to balance the fit of the 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:

$$\text{Flow error } = \frac{\|\hat{\mathbf{v}} - \mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2}, \tag{50a}$$

$$\text{Wardrop error} = \frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \frac{\max_{\{p \in \mathcal{P}_w | \hat{\phi}_p > 0\}} t_p^D(\tilde{\mathbf{x}}, \hat{\mathbf{v}}) - \min_{p \in \mathcal{P}_w} t_p^D(\tilde{\mathbf{x}}, \hat{\mathbf{v}})}{\min_{p \in \mathcal{P}_w} t_p^D(\tilde{\mathbf{x}}, \hat{\mathbf{v}})}, \tag{50b}$$

$$\text{Out-of-sample error} = \frac{1}{\sum_{w \in \mathcal{W}^{\text{taxi}}} |N_w^{\text{taxi}}|} \sum_{w \in \mathcal{W}^{\text{taxi}}} \left|N_w^{\text{taxi}}\right| \frac{|t_w^D(\tilde{\mathbf{x}}, \hat{\mathbf{v}}) - \text{taxi}_w|}{\text{taxi}_w}. \tag{50c}$$

In the expressions above, $(\hat{\boldsymbol{\phi}}, \hat{\mathbf{v}})$ 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 §5.1 and §A.2), which is not used during estimation.

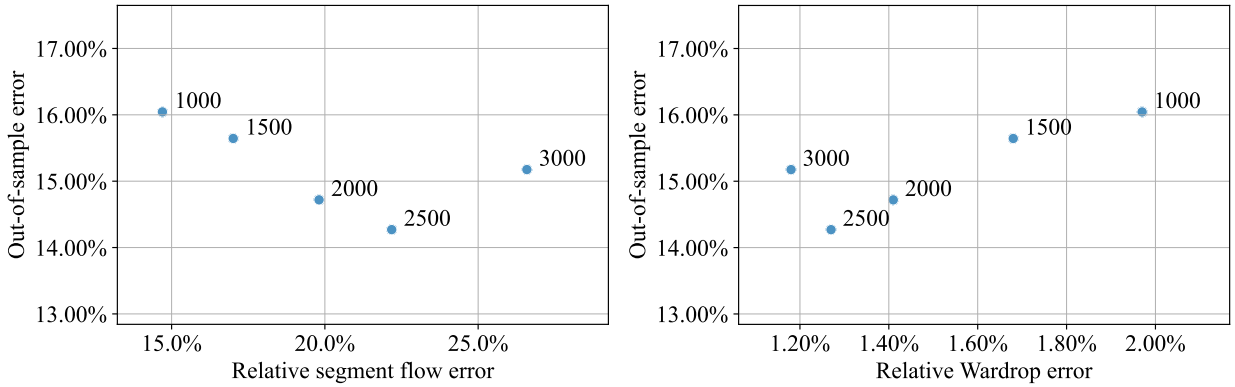
Figure 7 visualizes the performance of each value of λ according to the three error metrics. Although there is no value that dominates, we adopt $\lambda = 2500$ for our §5 empirical study, which is the only value of λ that underperforms the other values on only one error metric, and also has the lowest out-of-sample error. Under $\lambda = 2500$, the flow error is 22.2%, the Wardrop error is 1.3%, and the out-of-sample error is 14.3%.

For completeness, Table 8 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 (50c), i.e.,

$$\frac{1}{|\mathcal{W}^{\text{taxi}}|} \sum_{w \in \mathcal{W}^{\text{taxi}}} \frac{|t_w^D(\tilde{\mathbf{x}}, \hat{\mathbf{v}}) - \text{taxi}_w|}{\text{taxi}_w}. \tag{51}$$

[22] 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.

[23] For our measure of out-of-sample error, we use the average discrepancy between the observed average taxi trip times and predicted OD driving times, weighted by the number of taxi trips per OD pair



(a) Out-of-sample vs. flow error.

(b) Out-of-sample error vs. Wardrop error.

Figure 7 Estimation errors for varying values of penalty parameter λ.

The errors in Table 8 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 8 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 as the value added by using a model of traffic congestion in addition to free-flow travel times when making predictions.

Table 8 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%	18.9%	31.6%

A.5. Comparison with Linear Regression and Regression Trees on OD Driving Time Predictions

In this section, we predict travel times on the taxi trip data using two alternative methods, namely, linear regression and regression tree. The purpose of this comparison is to validate the accuracy of the predictions generated by the traffic congestion model described in §2.2, which is based on the assumption of Wardrop equilibrium.

To conduct a fair comparison, all three prediction methods use the same set of input features based on 3,698 OD pairs from 1,408,833 taxi trip records: The observed road segment lengths l_s, the total road width $n_s \cdot w_s$, the free flow driving times T_s, and the observed traffic flows z_s. We note two challenges in this prediction task that require simplifying assumptions. First, not all road segments have observed traffic flows; to address this, we impute the missing values by averaging the traffic flows of the road segments connected to the focal road segment. Second, because the observations on driving time (i.e., taxi trip duration) are at the OD-level, predicting the corresponding OD travel times requires specifying the driving path taken by each

taxi trip, which is unobserved in the data. For simplicity, we assume drivers take the most recommended driving path as per the Google Directions API. Identifying the driving path (i.e., set of connected road segments $\mathcal{S}_p^D$) for each OD pair then allows us to determine which segment-level feature data correspond to each OD pair.

For our comparison, we first fit the following linear regression model to the taxi trip data:

$$\text{taxi}_p = \sum_{s \in \mathcal{S}_p^D} \left[\theta_0 + \theta_1 \cdot T_s + \theta_2 \cdot l_s + \theta_3 \cdot n_s \cdot w_s + \theta_4 \cdot \frac{l_s}{n_s \cdot w_s} + \left(\theta_5 + \theta_6 \cdot l_s + \theta_7 \cdot n_s \cdot w_s + \theta_8 \cdot \frac{l_s}{n_s \cdot w_s} \right) z_s \right] \quad (52)$$

On the left hand side, taxi_p is the average duration of taxi trips (taken during the morning rush hours of 6am to 10am on weekdays in 2018) between the OD pair for which the most recommended path is p. The right hand side includes an intercept term and seven different covariates constructed from the four input features described above; note that this specification is more general than our driving travel time function described in (36)-(38). Fitting the model using 10-fold cross validation generates an average prediction error (given in (51)) of 17.1%. Next, we train a fine-tuned regression tree, again with 10-fold cross validation using the same input features as in (52), which produces an average prediction error of 16.9%.

Although our prediction method produces a slightly higher prediction error of 18.9% (see Table 8), in contrast to the linear regression and regression tree methods, it is notably *not* trained on the taxi trip durations, taxi_p. Our model effectively replaces the average trip durations taxi_p with the road network structure and commuter demand data, and instead determines the driving time between OD pair w, t_w^D, endogenously in equilibrium. The comparable errors across all three methods suggests the traffic congestion model produces sensible predictions of travel times on the network.

We highlight a few advantages of our method compared to standard prediction techniques like linear regression and regression trees. Most importantly, because we explicitly model congestion in the road network, our approach is more suitable for evaluating counterfactuals, and in particular can be used to assess the impact of bike lane network expansion on driving travel times. Additionally, our method can generate reasonable travel time predictions in settings where detailed OD trip data are not available, but commuter demand and vehicle flows data are. Lastly, as a result of imposing a network structure on the data and enforcing Wardrop equilibrium, our method does not require knowledge of the driving path chosen by commuters and can handle missing vehicle flows on some road segments, which must otherwise be imputed. However, if high-resolution mobility data is available (e.g., GPS coordinates of individual commuters), then standard prediction methods like linear regression and regression trees may generate more accurate travel time predictions. We also note that our approach is more computationally demanding because of the need to solve the non-convex optimization problem WLS-A, given in (16).

B. Map of Divvy Bike Share Stations for Mode Choice Estimation

As described in §5.1, 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 8 depicts locations of all Divvy stations in Chicago, as well as the study region for estimating the mode choice parameters, which approximately corresponds to downtown Chicago. The road network corresponding to the study region is summarized in Table 9.

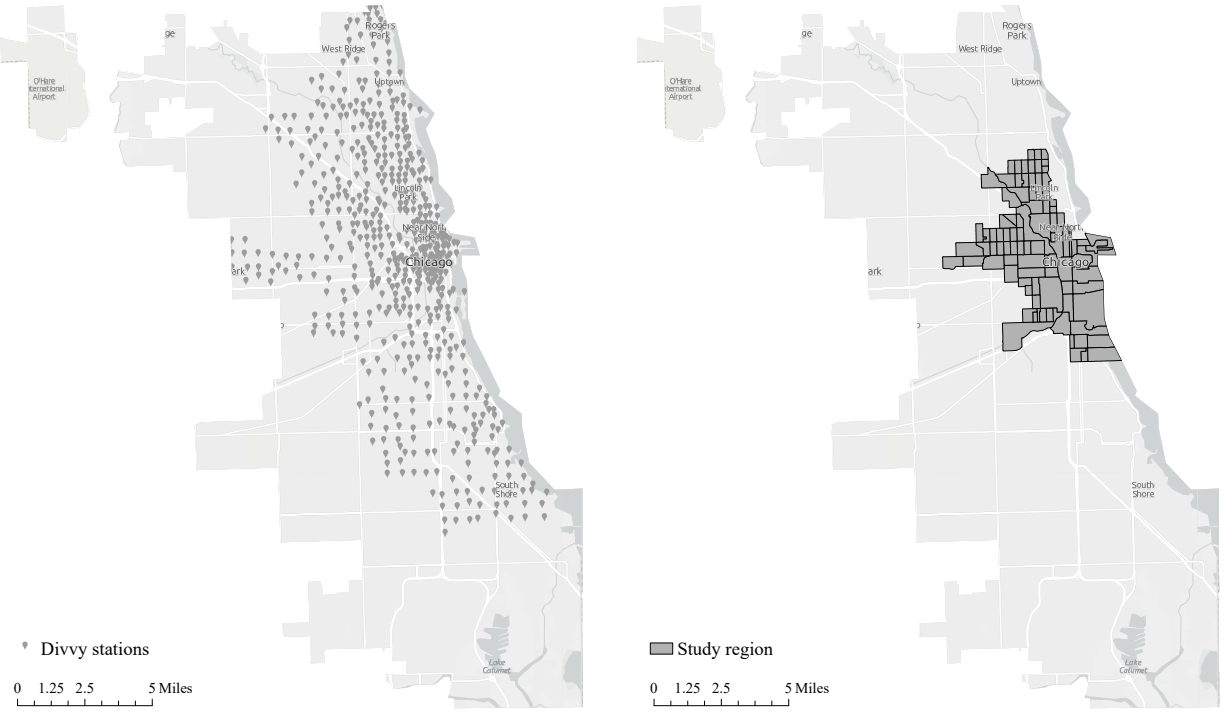


Figure 8 Locations of Divvy stations and study region for mode choice estimation.

Table 9 Size of downtown Chicago network used to estimate mode choice parameters β.

	Road segments $\lvert\mathcal{S}\rvert$	Driving paths $\lvert\mathcal{P}^D\rvert$	OD pairs $\lvert\mathcal{W}\rvert$
Size	6,101	14,632	5,272

C. Robustness Checks: Alternative Instruments for Mode Choice Estimation

In this section, we verify robustness of our mode choice model estimates by considering alternative instrumental variables. First, for the BLP-like instrument described in §5.3, we implement several different choices of the distance interval $[a, b]$. Then, we consider two additional instruments: A "spatial" instrument in the spirit of Hausman-type instruments (Hausman 1996, Nevo 2001), which uses the endogenous variables of nearby OD pairs, and a network-based instrument in the style of He et al. (2021) and Arora et al. (2022), which is similar to the BLP-like instrument but more explicitly leverages the underlying network structure. For conciseness, we present the estimates for the two key parameters of interest, which are the coefficients for bike lane coverage parameter (β_1^C) and driving travel time (β_1^D). In summary, we find our estimates to be highly stable across all implemented instruments.

For the BLP-like instrument, we vary the distance interval $[a, b]$ by combining each value of $a \in \{500, 1000, 1500\}$ with each value of $b \in \{7000, 8000\}$, for a total of six different distance intervals. The estimates for β_1^C and β_1^D are shown in columns (1)-(6) in Table 10. Note that column (2) provides the most

conservative (i.e., smallest) estimate of the effect of bike lane coverage, which is the value used when solving the bike lane optimization problem in our empirical study §5.4.

The spatial instruments are the average bike lane coverage and driving time of a subset of OD pairs other than the focal OD pair. Following our setup for the BLP-like instruments in §5.3, for each OD pair w we define a set of other OD pairs

$$\mathcal{Z}^{(a,b)}_{w=\{i,j\}} = \{w' = (i', j') \in \mathcal{W} | \text{dis}\,(i', i) \in [a, b]\,, \text{dis}\,(j', j) \in [a, b]\}\,. \tag{53}$$

Then for each OD pair w', the instruments are defined as

$$Z^{(a,b)}_{w'} = \left[\frac{1}{\left|\mathcal{Z}^{(a,b)}_{w'=\{i',j'\}}\right|} \sum_{w \in \mathcal{Z}^{(a,b)}_{w'=\{i',j'\}}} \rho_w,\ \frac{1}{\left|\mathcal{Z}^{(a,b)}_{w'=\{i',j'\}}\right|} \sum_{w \in \mathcal{Z}^{(a,b)}_{w'=\{i',j'\}}} t^D_w \right] \tag{54}$$

For the distance interval $[a, b]$, we set $a = 1500$ and $b = 7500$, which corresponds to the largest value of a and the average value of b used in the BLP-like instrument. Choosing a to be large ensures that the instrumental OD pairs are moderately distant from the focal OD pair, which is intended to promote the exclusion criterion.

The validity of our spatial instrument, which can be interpreted as a Hausman-type instrument if one views each OD pair as a "market", follows from arguments similar to those used in Nevo (2001). In our setting, relevance follows from the fact that bike lanes and driving travel times are likely to be spatially correlated due to neighborhood-specific factors. The exclusion criterion is satisfied under the assumption that unobserved factors that influence cycling or driving demand are independent across OD pairs that are moderately separated by distance. It is worth noting that the use of this spatial instrument is not completely free from validity concerns – for instance, in estimating the driving travel time coefficient β^D_1, there may exist complex and high-level network effects that impact traffic congestion, which challenges the exclusion criterion. In employing Hausman-type instruments, Nevo (2001) recognizes potential challenges with their validity, and those caveats extend to our study as well. Nonetheless, we find the stability in estimates observed in Table 10 to be encouraging.

The third instrument is a network-based instrument in the spirit of He et al. (2021) and Arora et al. (2022), which are similar to BLP-like instruments but explicitly utilize the underlying network structure (of bike stations and ride-hailing pick-up/drop-off locations, respectively). To explain the variations in bike lane coverage and driving time between OD pair $w = (i, j)$, we look for exogenous variables that are correlated with commute demand at census tracts i and j. Analogous to He et al. (2021) and Arora et al. (2022), we construct instrumental variables for both the origin i and destination j; below, we describe how the origin instrumental variables are constructed and note that the destination instruments follow similarly.

To construct the origin instrumental variables, we look for a set of exogenous characteristics over census tract $h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}$ that satisfies the following two criteria: (1) h is *connected* to the origin i of the focal OD pair in the sense that there is strictly positive commute demand between h and i, and (2)

h is sufficiently distant from the destination j of the focal OD pair. Formally, we define $\mathcal{H}^{\text{origin},b}_{w=(i,j)} = \left\{h \in \mathcal{I} | d_{w=(h,i)} + d_{w=(i,h)} > 0, \text{dis}\,(h,j) > b\right\}$. Then the origin instruments are defined as

$$Z^{\text{origin},b}_{w} = \begin{bmatrix} \frac{1}{\left|\mathcal{H}^{\text{origin},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}} \text{pd}_h \\ \frac{1}{\left|\mathcal{H}^{\text{origin},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}} \mathbf{npi}_h \\ \frac{1}{\left|\mathcal{H}^{\text{origin},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}} (\mathbf{1}^T \cdot \mathbf{npi}_h) \times \text{pd}_h \\ \frac{1}{\left|\mathcal{H}^{\text{origin},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}} \text{hi}_h \end{bmatrix}$$

where pd_h, $\mathbf{npi}_h$, and hi_h denote the population density, number of each type of point of interest, and median household income, and $(\mathbf{1}^T \cdot \mathbf{npi}_h) \times \text{pd}_h$ is the interaction between the total number of points of interest of all categories, $(\mathbf{1}^T \cdot \mathbf{npi}_h)$ and population density in the census tract $h \in \mathcal{H}^{\text{origin},b}_{w=(i,j)}$, respectively. The destination instruments are then

$$Z^{\text{dest},b}_{w} = \begin{bmatrix} \frac{1}{\left|\mathcal{H}^{\text{dest},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{dest},b}_{w=(i,j)}} \text{pd}_h \\ \frac{1}{\left|\mathcal{H}^{\text{dest},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{dest},b}_{w=(i,j)}} \mathbf{pi}_h \\ \frac{1}{\left|\mathcal{H}^{\text{dest},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{dest},b}_{w=(i,j)}} (\mathbf{1}^T \cdot \mathbf{pi}_h) \times \text{pd}_h \\ \frac{1}{\left|\mathcal{H}^{\text{dest},b}_{w=(i,j)}\right|} \sum_{h \in \mathcal{H}^{\text{dest},b}_{w=(i,j)}} \text{hi}_h \end{bmatrix}$$

where the set $\mathcal{H}^{\text{dest},b}_{w=(i,j)} = \left\{h \in \mathcal{I} | d_{w=(h,j)} + d_{w=(j,h)} > 0, \text{dis}\,(h,i) > b\right\}$. The validity of these network-based instruments parallel the reasoning provided in He et al. (2021) and Arora et al. (2022), where the connectedness and distance requirements (1) and (2) are intended to satisfy relevance and exclusion, respectively.

Table 10 Robustness results for all instrumental variable specifications.

	(1)	(2, adopted)	(3)	(4)	(5)	(6)	Spatial	Network
bike lane coverage (β^C_1)	2.87*** (0.60)	1.88*** (0.64)	4.35*** (0.65)	2.91*** (0.63)	2.27*** (0.64)	1.96*** (0.62)	2.73* (1.65)	2.74*** (0.83)
driving time (β^D_1)	-0.29*** (0.03)	-0.21*** (0.04)	-0.28*** (0.03)	-0.19*** (0.04)	-0.26*** (0.04)	-0.18*** (0.04)	-0.36*** (0.01)	-0.19*** (0.04)
Cragg–Donald statistic	26.04	21.63	24.45	25.27	23.41	22.78	72.51	15.40

Note: $^{**}p < 0.05$; $^{***}p < 0.01$. Columns (1) $\sim$ (6) represent the estimations using BLP-like instruments based on different combinations of $(a,b) = (500,7000), (500,8000), (1000,7000), (1000,8000), (1500,7000), (1500,8000)$. Columns "Spatial" and "Network" are the estimation based on the spatial instruments with $(a,b) = (1500,7500)$ and network-based instruments with $b = 8000$.

As shown in Table 10, all instruments produce comparable estimates for the coefficients for bike lane coverage (β^C_1) and driving time (β^D_1) in the mode choice model, and we reject the weak instrument null hypothesis, where the relevant critical value is 20.7 (for the Network-based instrument, we reject at a relative bias of 10% under the critical value of 11.05). For completeness, we present estimates for all parameters in Table 12. Next, we conduct a sensitivity analysis to evaluate the influence of each estimate on the corresponding optimal bike lane network as it relates to cycling ridership and traffic congestion. Specifically, for each IV estimate shown in Table 10, we re-solve the corresponding BL-A problem to generate the optimal bike lane design under a bike lane budget of $B = 25$ miles and congestion tolerance $\tau = 10\%$. We then assess each

bike lane design using the associated cycling mode share increase (defined in (42)) and worst-case increase in driving time (defined in (43)), which is obtained by solving the exact user equilibrium problem (24).

The performance metrics of all optimal bike lane plans are presented in Table 11. Note the main set of results presented in §5 uses the estimates in column (2), because it corresponds to the most conservative estimate of the impact of bike lanes on cycling demand. Our sensitivity analysis reveals that all of the instrumental variable estimates produce similar values for the impact of the optimal bike lane design on cycling adoption and traffic congestion, which is unsurprising given the robustness of the estimates shown in Table 10.

Table 11 Impact of optimal bike lane design based on different IV estimates.

	(1)	(2, adopted)	(3)	(4)	(5)	(6)	Spatial	Network
Increase in cycling share	2.81%	2.48%	2.93%	2.59%	2.69%	2.63%	2.76%	2.71%
Worst-case increase in driving time	10.58%	9.37%	11.56%	11.50%	11.05%	7.74%	11.24%	7.14%

See the notes for Table 10.

Table 12 Estimates for all specifications of the commuter mode choice model.

		(1)	(2)	(3)	(4)	(5)	(6)	Spatial	Netwrok	OLS
Cycling	intercept	0.96	2.36	2.01	2.86	3.36	3.93	1.76	2.07	2.16
		(2.57)	(2.43)	(2.47)	(2.43)	(2.45)	(2.46)	(2.50)	(2.44)	(1.71)
	cycling time	-0.24***	-0.22***	-0.25***	-0.22***	-0.23***	-0.21***	-0.25***	-0.22***	-0.21***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
	average elevation	-3.01***	-2.66***	-3.41***	-2.74***	-3.18***	-2.84***	-2.72***	-2.68***	-2.10***
	grade	(0.80)	(0.79)	(0.82)	(0.80)	(0.79)	(0.79)	(0.81)	(0.81)	(0.66)
	bike lane coverage	2.87***	1.88***	4.35***	2.91***	2.27***	1.96***	2.74*	2.74***	-0.14
		(0.60)	(0.64)	(0.65)	(0.63)	(0.64)	(0.62)	(1.65)	(0.83)	(0.15)
Driving	intercept	4.08***	3.97**	4.32***	4.11***	4.32***	4.29***	4.95***	3.87**	4.39***
		(1.52)	(1.55)	(1.51)	(1.57)	(1.51)	(1.58)	(1.44)	(1.57)	(1.69)
	vehicle ownership	0.58***	0.60***	0.62***	0.57***	0.74***	0.68***	0.52**	0.62***	0.49
		(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)	(0.23)	(0.22)	(0.31)
	driving time	-0.29***	-0.21***	-0.28***	-0.19***	-0.26***	-0.18***	-0.36***	-0.19***	-0.25***
		(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.01)	(0.04)	(0.01)
	household income	0.02	0.03**	0.02*	0.03**	0.03**	0.03**	0.01	0.03**	0.03
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Outside option	transit time	-0.02***	-0.03***	-0.02***	-0.03***	-0.02***	-0.03***	-0.01***	-0.03***	-0.02***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.00)
	transit stations	0.004***	0.004***	0.003**	0.004***	0.004***	0.004***	0.005***	0.004***	0.005***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cycling controls	movie	-0.64***	-0.59***	-0.69***	-0.62***	-0.59***	-0.57***	-0.60***	-0.62***	-0.51***
	theater	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)	(0.08)
	park	-0.93	-1.01	-1.02	-1.19	-1.52	-1.65	-0.74	-0.79	-0.90
		(1.04)	(1.03)	(1.06)	(1.05)	(1.05)	(1.05)	(1.04)	(1.03)	(0.62)
	restaurant	-1.00	-2.01	-2.01	-2.53	-2.34	-2.66	-1.98	-1.95	-1.72
		(2.36)	(2.21)	(2.24)	(2.20)	(2.22)	(2.23)	(2.29)	(2.22)	(1.59)
	shopping	0.12	0.14	0.07	0.10	0.13	0.12	0.11	0.15	0.19**
	mall	(0.11)	(0.11)	(0.12)	(0.11)	(0.11)	(0.11)	(0.12)	(0.11)	(0.09)
	supermarket	-0.37***	-0.30***	-0.38***	-0.36***	-0.34***	-0.33***	-0.32***	-0.39***	-0.27***
		(0.11)	(0.10)	(0.11)	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)	(0.09)
	bakery	0.08	0.19	0.02	0.17	-0.06	0.07	0.08	-0.07	0.13
		(0.34)	(0.34)	(0.35)	(0.34)	(0.34)	(0.34)	(0.34)	(0.34)	(0.26)
	bar	1.95***	1.78***	1.82***	1.77***	2.00***	1.81***	1.68***	1.77***	1.79***
		(0.52)	(0.52)	(0.52)	(0.52)	(0.52)	(0.52)	(0.52)	(0.51)	(0.38)
	university	0.70***	0.66***	0.66***	0.65***	0.70***	0.67***	0.70***	0.70***	0.73***
		(0.16)	(0.16)	(0.17)	(0.16)	(0.16)	(0.16)	(0.17)	(0.16)	(0.13)
	hospital	-0.39***	-0.36**	-0.44***	-0.39***	-0.39***	-0.39***	-0.34**	-0.38***	-0.26**
		(0.14)	(0.14)	(0.15)	(0.14)	(0.14)	(0.14)	(0.15)	(0.14)	(0.11)
	library	-0.12	-0.11	-0.05	-0.08	-0.13	-0.16	-0.14	-0.13	-0.17**
		(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.09)
Driving controls	movie	-0.36***	-0.39***	-0.36***	-0.41***	-0.37***	-0.42***	-0.30***	-0.40***	-0.36***
	theater	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
	park	-0.53	-0.71	-0.54	-0.77*	-0.56	-0.80*	-0.40	-0.72	-0.67
		(0.47)	(0.47)	(0.47)	(0.47)	(0.47)	(0.47)	(0.46)	(0.47)	(0.62)
	restaurant	0.37	0.14	0.03	-0.12	-0.17	-0.36	-0.12	-0.04	-0.12
		(1.45)	(1.48)	(1.44)	(1.49)	(1.44)	(1.51)	(1.37)	(1.49)	(1.59)
	shopping	-0.19***	-0.23***	-0.21***	-0.22***	-0.21***	-0.21***	-0.18***	-0.21***	-0.19**
	mall	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)
	supermarket	-0.12*	-0.08	-0.10	-0.09	-0.09	-0.09	-0.09	-0.09	-0.08
		(0.07)	(0.06)	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.09)
	bakery	-0.10	-0.31*	-0.09	-0.29	-0.13	-0.27	0.05	-0.24	-0.14
		(0.18)	(0.19)	(0.18)	(0.18)	(0.18)	(0.19)	(0.18)	(0.18)	(0.26)
	bar	0.97***	0.84***	0.89***	0.80***	0.87***	0.73***	0.99***	0.77***	0.83**
		(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.29)	(0.28)	(0.38)
	university	-0.36***	-0.36***	-0.36***	-0.34***	-0.34***	-0.32***	-0.37***	-0.33***	-0.34***
		(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.13)
	hospital	0.10	0.14*	0.11	0.15*	0.13	0.17**	0.08	0.17**	0.13
		(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.11)
	library	-0.40***	-0.36***	-0.39***	-0.35***	-0.39***	-0.38***	-0.39***	-0.35***	-0.35***
		(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.09)

See the notes for Table 10.

D. Mixed-Integer Linear Programming Reformulation of Path Selection Model BL-A

This section develops the mixed-integer linear programming (MILP) reformulation of the linearized bike lane path selection model BL-A given in (32). In preparation, we first re-write the constraints (30d) and (30e) as

$$\xi_s \geq \rho_s^r(x_s) \cdot v_s + \epsilon_s^r(x_s), \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{55a}$$

$$\psi_w^m \geq a_w^r \cdot d_w^m + b_w^r, \quad r \in \mathcal{R}_\psi, w \in \mathcal{W}, m \in \mathcal{M}, \tag{55b}$$

where

$$\rho_s^r(x_s) = \alpha_s(x_s) \cdot v_s^r \quad r \in \mathcal{R}_\xi, s \in \mathcal{S}, \tag{56a}$$

$$\epsilon_s^r(x_s) = -\frac{1}{2}\cdot\alpha_s(x_s)\cdot(v_s^r)^2, \quad r\in\mathcal{R}_\xi, s\in\mathcal{S}, \tag{56b}$$

$$a_w^r = [1+\log(d^r)], \quad r\in\mathcal{R}_\psi, w\in\mathcal{W}, m\in\mathcal{M}, \tag{56c}$$

$$b_w^r = -d^r, \quad r\in\mathcal{R}_\psi, w\in\mathcal{W}, m\in\mathcal{M}. \tag{56d}$$

The lower-level (i.e., user equilibrium) problem in in (32) is a linear program. As a result, by leveraging linear programming duality, the bi-level optimization problem (32) 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:

$$\underset{\substack{\mathbf{x},\mathbf{y},\mathbf{d},\boldsymbol{\phi},\mathbf{v},\boldsymbol{\xi},\\ \boldsymbol{\psi},\boldsymbol{\mu},\boldsymbol{\gamma},\boldsymbol{\lambda},\boldsymbol{\pi},\boldsymbol{\sigma}}}{\text{maximize}} \quad \sum_{w\in\mathcal{W}} d_w^C$$

subject to $(23b)-(23d)$, upper level constraints

$$S_L(\mathbf{x},\mathbf{d},\mathbf{v},\boldsymbol{\psi},\boldsymbol{\xi}) = \sum_{w\in\mathcal{W}}\left(\mu_w d_w + \sum_{r\in\mathcal{R}_\psi}\sum_{m\in\mathcal{M}}\pi_w^{mr} b_w^r\right) + \sum_{r\in\mathcal{R}_\xi}\sum_{s\in\mathcal{S}}\sigma_s^r\epsilon_s^r(x_s) \quad \text{strong duality}$$

primal feasibility:

$$\begin{aligned}
&\xi_s \geq \rho_s^r(x_s)\cdot v_s + \epsilon_s^r(x_s), && r\in\mathcal{R}_\xi, s\in\mathcal{S},\\
&\psi_w^m \geq a_w^r\cdot d_w^m + b_w^r, && r\in\mathcal{R}_\psi, w\in\mathcal{W}, m\in\mathcal{M},\\
&d_w^O + d_w^C + d_w^D = \bar{d}_w, && w\in\mathcal{W},\\
&\textstyle\sum_{p\in\mathcal{P}_w^D}\phi_p = d_w^D, && w\in\mathcal{W},\\
&v_s = \textstyle\sum_{\{p\in\mathcal{P}^D | s\in\mathcal{S}_p^D\}}\phi_p, && s\in\mathcal{S},\\
&\boldsymbol{\phi}\geq\mathbf{0}, &&
\end{aligned}$$

dual feasibility:

$$\begin{aligned}
&\mu_w - \textstyle\sum_{r\in\mathcal{R}_\psi} a_w^r\cdot\pi_w^{mr} \leq -u_w^m, && w\in\mathcal{W}, m\in\{C,O\},\\
&\mu_w + \gamma_w - \textstyle\sum_{r\in\mathcal{R}_\psi} a_w^r\cdot\pi_w^{Dr} \leq -\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top\mathbf{X}^D\right), && w\in\mathcal{W},\\
&-\textstyle\sum_{\{w\in\mathcal{W}|p\in\mathcal{P}_w^D\}}\gamma_w - \sum_{s\in\mathcal{S}_p^D}\lambda_s \leq 0, && p\in\mathcal{P}^D,\\
&\lambda_s - \textstyle\sum_{r\in\mathcal{R}_\xi}\rho_s^r(x_s)\cdot\sigma_s^r \leq -\beta_1^D\cdot T_s, && s\in\mathcal{S},\\
&\textstyle\sum_{r\in\mathcal{R}_\psi}\sigma_s^r = -\beta_1^D, && s\in\mathcal{S},\\
&\textstyle\sum_{n\in\mathcal{N}}\pi_w^{mn} = 1, && w\in\mathcal{W}, m\in\mathcal{M},\\
&\boldsymbol{\pi},\boldsymbol{\sigma}\geq 0. &&
\end{aligned}$$

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

$$\xi_s \geq \rho_s^r(1)\cdot v_s + \epsilon_s^r(1) - M_1\cdot(1-x_s), \quad r\in\mathcal{R}_\xi, s\in\mathcal{S}, \tag{58a}$$

$$\xi_s \geq \rho_s^r(0)\cdot v_s + \epsilon_s^r(0) - M_1\cdot x_s, \quad r\in\mathcal{R}_\xi, s\in\mathcal{S}, \tag{58b}$$

and

$$\lambda_s \leq -\beta_1^D\cdot T_s + \sum_{r\in\mathcal{R}_\xi}\rho_s^r(1)\cdot\sigma_s^r + M_2\cdot(1-x_s), \quad s\in S, \tag{59a}$$

$$\lambda_s \leq -\beta_1^D \cdot T_s + \sum_{r \in \mathcal{R}_\xi} \rho_s^r(0) \cdot \sigma_s^r + M_2 \cdot x_s, \quad s \in S, \tag{59b}$$

where M_1 and M_2 are large constants. For computational efficiency in solving the bike lane path selection model BL-A, we omit the OD pairs representing the smallest 20% of pairs by demand within the study region (shaded area of Figure 8). 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 13.

Table 13 Size of downtown Chicago network used in bike lane optimization model BL-A.

	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

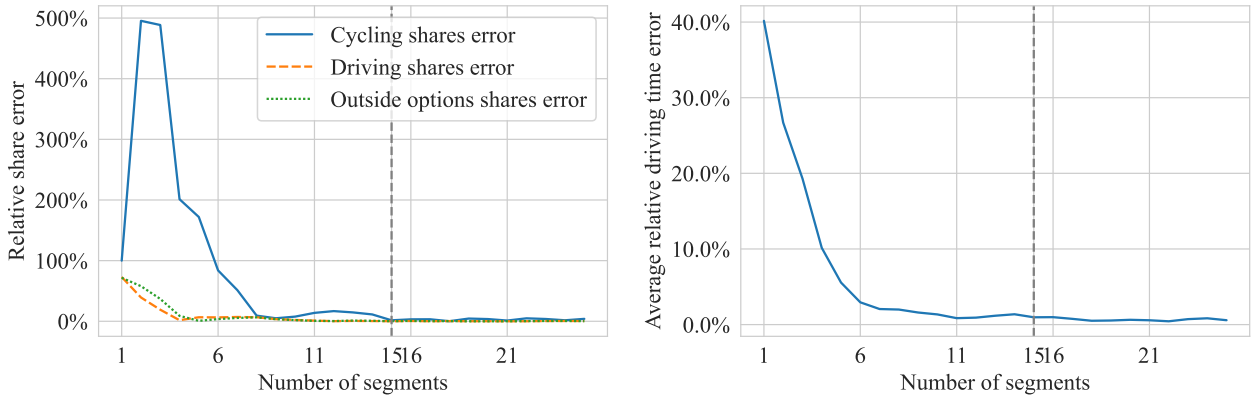
D.1. Sensitivity of Approximation Error to Number of Linear Segments

In our empirical study, we set the number of piecewise linear segments in the approximation BL-A as $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 15$. In this section, we investigate numerically how the number of segments used in the linearization technique affects the approximation error between the exact formulation BL and the approximation BL-A. To do so, we fix $|\mathcal{R}_\xi| = |\mathcal{R}_\psi|$, vary both from 1 to 50, and compare the mode shares ($d_w^m/\tilde{d}_w$) and driving times (t_p^D) computed by the linearized lower-level problem (30) to the true mode shares and driving times from the exact lower-level problem (24). 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 9 shows the relative errors in mode shares and driving times between the exact and linearized user equilibrium problems. When $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 15$, the relative errors for cycling, driving and outside options shares are 3.22%, 0.49% and 0.62%, respectively, and the average relative driving time error across all OD pairs is 0.98%. Note that because the objective function of BL is simply the total cycling demand $\sum_{w \in \mathcal{W}} d_w^C$, the relative error in cycling mode shares of 3.22% indicates a small suboptimality gap in Theorem 1 when $|\mathcal{R}_\xi| = |\mathcal{R}_\psi| = 15$. 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 BL.

D.2. Optimality Gaps at Termination of MILP Formulation

In solving the MILP reformulation of BL-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 14, we present the relative MIP gaps at termination for each of the 12 combinations of (β,τ) used in the empirical study. The gaps are largest when both the budget constraint B and the tolerance τ are small, with a maximum MIP gap of 41.33% 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). Note that not solving BL-A to optimality implies the estimated cycling ridership under the recommended bike lane network presented in §5.4 and §5.5 are conservative estimates of the ridership under an optimal solution to BL-A.



(a) Mode share errors.

(b) Path driving time errors.

Figure 9 Approximation errors due to linearization of user equilibrium problem (24)**.**

Table 14 Relative optimality gaps at termination when solving BL-A as a mixed-integer linear program.

	$B = 10$	$B = 25$	$B = 50$	$B = 75$
$\tau = 5\%$	41.33%	41.25%	34.44%	33.63%
$\tau = 10\%$	27.48%	18.19%	6.50%	4.79%
$\tau = 15\%$	18.24%	8.69%	3.46%	0.66%

E. Comparison with Alternative Bike Lane Planning Methods

This section provides additional technical details and numerical results for the benchmark bike lane planning methods discussed in §5. §E.1 presents the detailed steps for each of the benchmark methods, and §E.2 summarizes results. In §E.3, we use our estimated model to evaluate the performance of the bike lanes actually installed by Chicago after our study period (i.e., from 2018-2021).

E.1. Description of Benchmark Methods

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 BL-A, where the segment travel times t_s^D are fixed to their values under the status quo bike lane network $\tilde{\mathbf{x}}$. Under this assumption, the total driving utility is

$$\sum_{w \in \mathcal{W}} \left(\int_0^{d_w^D(\mathbf{x})} u_w^D(\mathbf{x}) \cdot dd \right) = \sum_{w \in \mathcal{W}} \left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D \right) \cdot d_w^D(\mathbf{x}) + \beta_1^D \sum_{s \in \mathcal{S}} t_s^D(\tilde{\mathbf{x}}) \cdot v_s. \tag{60}$$

We then obtain the following fixed-time analog of $S(\mathbf{x}, \mathbf{d}, \mathbf{v})$:

$$S^{\mathrm{FT}}(\mathbf{x}, \mathbf{d}, \mathbf{v}) = \sum_{w \in \mathcal{W}} - \left(u_w^O d_w^O + u_w^C(\mathbf{x}) \cdot d_w^C + \beta_0^D d_w^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D d_w^D \right) - \beta_1^D \sum_{s \in \mathcal{S}} t_s^D(\tilde{\mathbf{x}}) \cdot v_s + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} d_w^m \cdot \log(d_w^m). \tag{61}$$

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

$$\underset{\mathbf{x}, \mathbf{y}, \mathbf{d}, \boldsymbol{\phi}, \mathbf{v}}{\text{maximize}} \quad \sum_{w \in \mathcal{W}} d_w^C \tag{62a}$$

$$\text{subject to } (23b) - (23d) \tag{62b}$$

$$(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) = \underset{\mathbf{d},(\phi,\mathbf{v})\in\Omega(\mathbf{d}^D)}{\operatorname{argmin}} S^{\mathrm{FT}}(\mathbf{x}, \mathbf{d}, \mathbf{v}). \tag{62c}$$

As a result of the fixed-time assumption, congestion tolerance constraints such as (35) are irrelevant in this model, and (62) does not permit control over driving times. The model can be reformulated and linearized in the same manner as described in §4.2, which leads to a MILP that can be solved using commercial optimization software. The termination criteria are the same as in §D.2: the solver is terminated when the optimality gap (MIP gap) is within 0.1% or the solution time reaches 6 hours. Table 15 reports the optimality gaps at termination for different values of B. We observe that the optimality gaps for the fixed-time model (62) are comparable to the gaps for BL-A presented in Table 14.

Table 15 Relative optimality gaps at termination when solving fixed-time model as a MILP.

B	10	25	50	75
Gap	25.97%	23.23%	10.05%	8.07%

Greedy heuristic. Next, we present the steps for the greedy heuristic discussed in §5.5. 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 $\tilde{t}_p^D$ 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

$$\tau_w = \max_{p\in\mathcal{P}^D} \left\{ \frac{t_p^w - \tilde{t}_p^D}{\tilde{t}_p^D} \right\} \tag{63}$$

to be the worst-case relative increase in driving times over all paths if bike lane are 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 (24). Next, define

$$B_w = \sum_{s\in S_w^C} \ell_s(1 - \tilde{x}_s), \tag{64}$$

which represents the consumption of the budget B due to the addition of bike lanes to path p_w^C. Lastly, define

$$\delta_w = \frac{1}{B_w} \sum_{w'\in\mathcal{W}} \left(d_{w'}^{C,w} - \tilde{d}_{w'}^C \right). \tag{65}$$

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 $\max_{p \in \mathcal{P}^D} \{t_p^D(\bar{\mathbf{x}}) / \tilde{t}_p^D\} > 1 + \tau$, update $\sigma \leftarrow \sigma - \gamma$ and go to Step 1; else, set $\mathbf{x}^G = \bar{\mathbf{x}}$ and terminate.

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 $\sum_{s \in \mathcal{S}^H} \ell_s (1 - \tilde{x}_s) \leq B$.
3. Set $x_s^H = 1$ for all $s \in \mathcal{S}^H \bigcup \tilde{S}$.

E.2. Comparison with Benchmark Methods on Cycling Ridership and Congestion

Figure 10 depicts the performance of BL-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 §5.5 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 BL-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

$$\frac{\sum_{w \in \mathcal{W}} d_w^D(\mathbf{x}^*) \cdot t_w^D(\mathbf{x}^*) - \sum_{w \in \mathcal{W}} d_w^D(\tilde{\mathbf{x}}) \cdot t_w^D(\tilde{\mathbf{x}})}{\sum_{w \in \mathcal{W}} d_w^D(\tilde{\mathbf{x}}) \cdot t_w^D(\tilde{\mathbf{x}})}. \tag{66}$$

Figure 11 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 capacity effect discussed in §5.4 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 capacity effect. Further, when the bike lane budget B is small ($B \in \{10, 25, 50\}$), the recommended bike lane plan from BL-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 are 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 prescriptive model is particularly valuable when the bike lane mileage budget is modest.

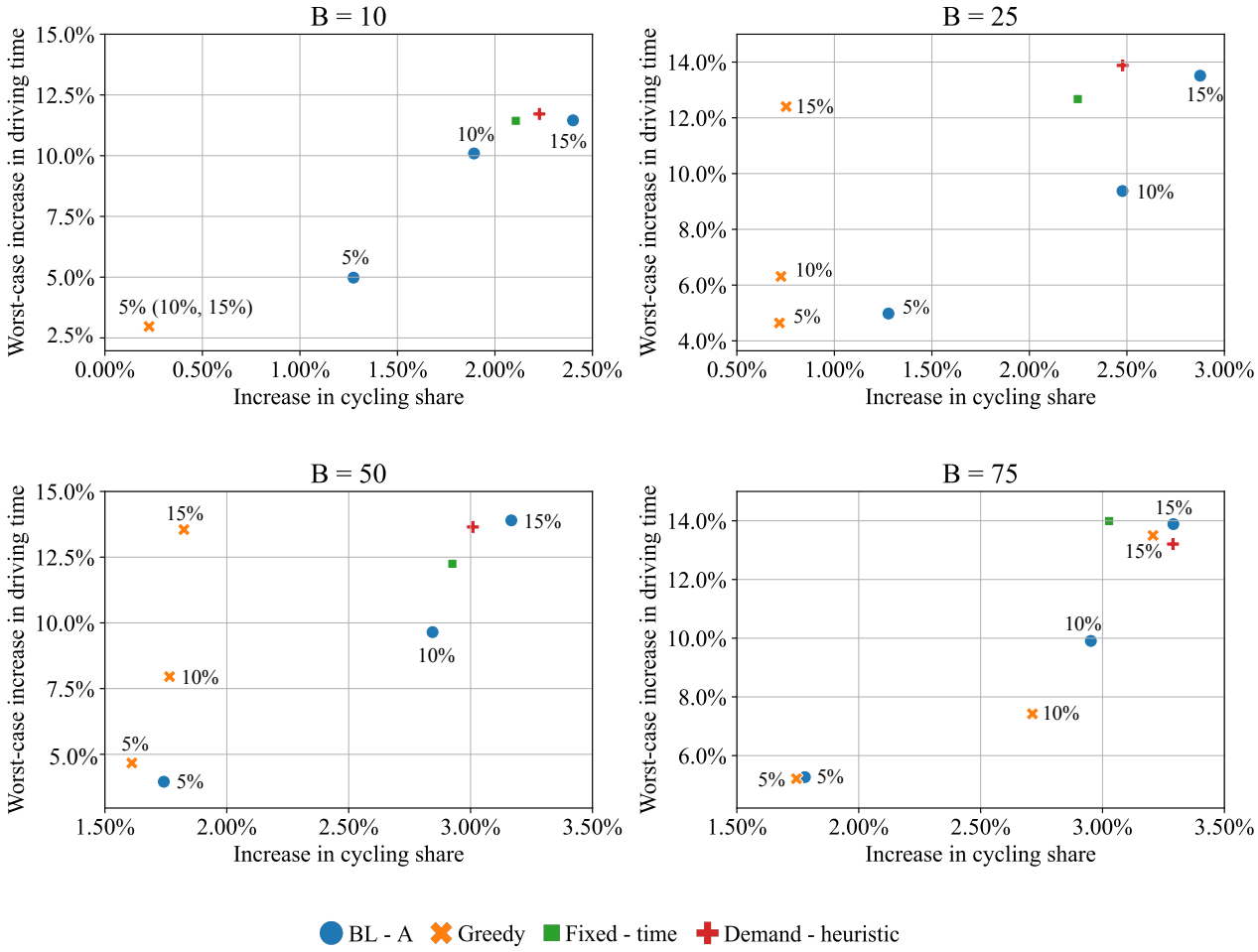


Figure 10 **Comparison of BL-A and benchmark methods on worst-case driving time and cycling ridership.**

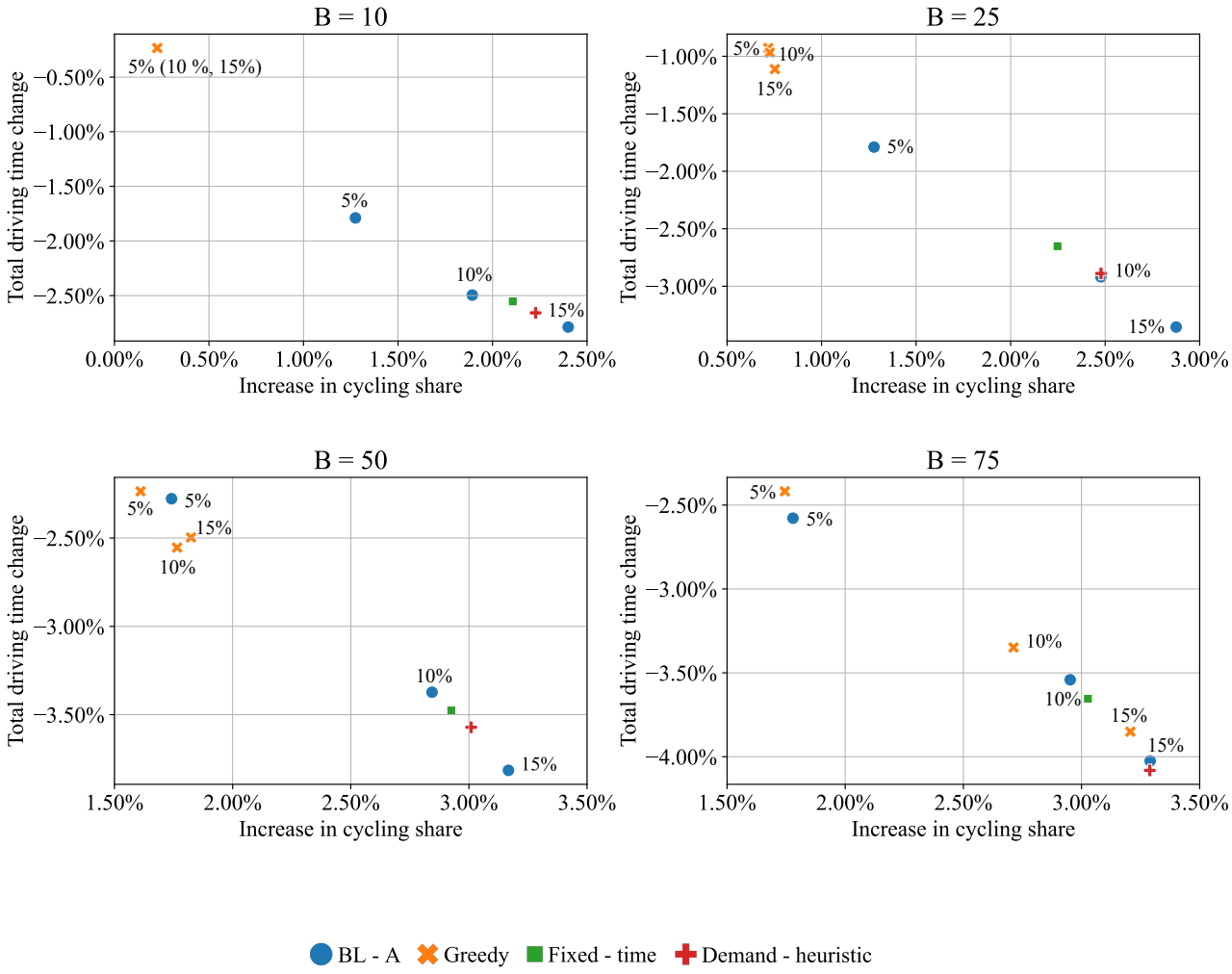


Figure 11 **Comparison of BL-A and benchmark methods on system-wide driving time and cycling ridership.**

E.3. Comparison with Chicago's Bike Lane Expansion Plan

Here we evaluate the performance (as predicted by our estimated model) of the new bike lanes constructed in Chicago following our study period of 2018. We obtained data on the bike lanes added from November 2018 to November 2021 from the Chicago Open Data Portal (Chicago Data Portal 2023). These new bike lanes represent an additional 14.48 miles in our study region of downtown Chicago, as illustrated in Figure 12. We match these new bike lanes to our road network and pass the complete bike lane network through our estimated model (i.e., by solving the associated user equilibrium problem (24)) to obtain predictions of the impacts on cycling adoption and congestion. Based on our model, these new bike lanes lead to an increase in the cycling mode share of 0.38% and a maximum driving time increase of 8.23% over all OD pairs.

For comparison purposes, we also re-solve the greedy heuristic described in §5.5 and Appendix E and our optimization problem BL-A using a budget of $B = 14.48$ miles and a congestion tolerance of 10%, to examine how the predictions from each of the three bike lane plans differ. The findings are summarized in Table 16. Remarkably, we observe that the expansion of bike lanes in Chicago yields similar results to those achieved by the greedy heuristic: The worst-case increase in driving time between the OD pairs is 8.23%, slightly higher than the 7.99% recorded by the greedy heuristic, with a difference in cycling shares of 0.03%. Moreover, in line with the observations we discussed in Section 5.5, BL-A surpasses both Chicago's actual implementation and the greedy heuristic with respect to cycling share. In addition, our method with the specified parameters of $B = 14.48$ miles and $\tau = 5\%$ dominates the other two approaches.

While the results in Table 16 suggest the prescriptions from our model might stimulate higher cycling ridership than Chicago's current expansion plan, there are some important caveats. First, city planners may have considered other objectives in addition to ridership and congestion, which would result in differences in the recommended locations for bike lane installation. Similarly, city planners likely face numerous constraints that we do not have visibility into when identifying bike lane locations, including land use considerations, feasibility of construction, and pressure from residents. Lastly, the values in Table 16 are based on predictions from our estimated congestion and mode choice models, and thus reflect the assumptions built into those models. Nonetheless, these results suggest that data-driven methods have the potential to significantly benefit city planners by identifying effective candidate areas for the installation of bike lanes.

Table 16 Comparison between actual bike lane expansion in Chicago, greedy heuristic, and bike lane optimization model BL-A. For greedy and BL-A, parentheses indicate bike lane budget (in miles) and congestion tolerance.

	New bike lanes (2018–2021)	Greedy (14.48, 10%)	BL-A (14.48, 10%)	BL-A (14.48, 5%)
Worst-case increase in driving time	8.23%	7.99%	8.57%	4.26%
Increase in cycling share	0.38%	0.41%	2.32%	1.12%

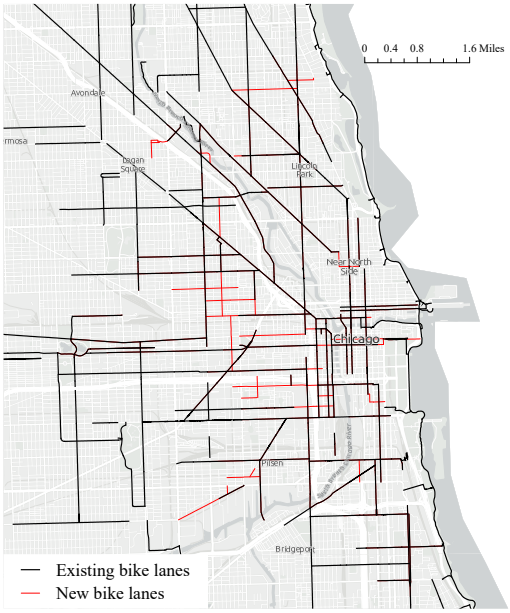


Figure 12 **Chicago's bike lane network up to 2018 and recently installed bike lanes between 2018–2021.**

F. Validating Estimation Method with Synthetic Data

This section describes additional numerical experiments for validating our estimation procedure. We use synthetic road network data designed to have similar properties to the real Chicago network studied in §5. The purpose of these experiments is three-fold: (1) to test the estimation method's ability to learn the true parameter $\boldsymbol{\theta}$, (2) to empirically verify the statistical consistency of the estimates (i.e., that estimation error decreases in network size), and (3) to empirically verify the validity of separating the estimation procedure for $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ into two stages.

F.1. Data and Setup

We consider three different sizes of networks by varying the number of edges $|\mathcal{S}| \in \{100, 500, 1000\}$, and 50 simulation trials for each network size. For each network size and simulation trial, the underlying data-generation process is as follows.

Network structure. We first construct an undirected road network whose node degree distribution follows the empirical node degree distribution of the Chicago network. To force the number of edges to be $|\mathcal{S}|$, we fix the total node degree in the network to be $2|\mathcal{S}|$. We then set the number of OD pairs $|\mathcal{W}|$ as $\frac{1}{3}|\mathcal{S}|$, which approximately matches the ratio of OD pairs to road segments in the Chicago network.

Network features. We generate network features including road lengths l_s, $s \in \mathcal{S}$, road widths $n_s \cdot w_s$, $s \in \mathcal{S}$, coverage of bike lanes ρ_w, $w \in \mathcal{W}$, and total commute demands d_w, $w \in \mathcal{W}$ by sampling from their corresponding empirical distributions observed in the Chicago network. Between each OD pair $w \in \mathcal{W}$, we designate the three shortest paths (in length) as the candidate driving paths set $\mathcal{P}^D$, and the shortest path as the candidate cycling path p_w^C. The free flow driving time t_s^D on each road segment is obtained by setting the driving speed to be 40 km/h, i.e., $T_s = \frac{l_s}{40(km/h)}$ for all $s \in \mathcal{S}$. Between each OD pair $w \in \mathcal{W}$, cycling time

is set as equal to double the free-flow driving time on the shortest path, i.e., $t_w^C = 2\sum_{s\in p_w^C} T_s$, and the outside option travel time is set equal to $t_w^O = \sum_{s\in p_w^C} T_s + 2u$, where u is a random variable uniformly distributed between 0 and 1.

Driving travel times and mode choice model. Following our modeling framework from §2, we assume commuters' mode choice follows a multinomial logit model and drivers' path choices satisfy Wardrop equilibrium. Specifically, the utilities of driving, cycling, and the outside option on OD pair w are given by

$$u_w^D(\boldsymbol{\theta}) = \beta_0^D + \beta_1^D \cdot t_w^D(\boldsymbol{\theta}) + \epsilon_w^D, \tag{67a}$$

$$u_w^C = \beta_0^C + \beta_1^C \cdot t_w^C + \beta_2^C \cdot \rho_w + \epsilon_w^D \tag{67b}$$

$$u_w^O = \beta_1^O \cdot t_w^O + \epsilon_w^O, \tag{67c}$$

where t_w^D, t_w^C, and t_w^O are the travel time for driving, cycling and the outside option, respectively, ρ_w is the bike lane coverage on the cycling path between OD pair w, and ϵ_w^D, ϵ_w^C and ϵ_w^O are idiosyncratic errors. The demand for each mode $m \in \{D, C, O\}$ is then given by

$$d_w^m = d_w \frac{e^{u_w^m}}{e^{u_w^D} + e^{u_w^C} + e^{u_w^O}}. \tag{68}$$

Following §2, the driving time on road segment s is assumed to take the form

$$t_s^D(\boldsymbol{\theta}) = \left(\theta_0 + \theta_1 \cdot \ell_s + \theta_2 n_s \cdot w_s + \theta_3 \left(\frac{\ell_s}{n_s \cdot w_s}\right)\right) v_s + T_s. \tag{69}$$

Ground truth. We first generate random values for the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ as ground truth. We generate a four-dimensional parameter vector $\boldsymbol{\theta}$ by randomly generating each component from the uniform distribution with support $[0, 10^{-3}]$ to ensure positiveness of $t_s^D(\boldsymbol{\theta})$. For $\boldsymbol{\beta}$, we let $[\beta_0^D, \beta_0^C, \beta_1^D, \beta_1^C, \beta_1^O, \beta_2^C] = [u_1, -2.5u_2, -u_3, -u_4, -u_5, 2u_6]$, where $\{u_i : i = 1, 2, \ldots, 6\}$ are drawn from i.i.d. uniform distributions with support $[0, 1]$. This approach to randomly generating $\boldsymbol{\beta}$ ensures that the resulting mode shares resemble the empirical mode shares in Chicago. We anticipate our results would be qualitatively similar under a different procedure for generating $\boldsymbol{\beta}$. Then the corresponding equilibrium demands for driving $\mathbf{d}^{D*}$, cycling $\mathbf{d}^{C*}$ and outside option $\mathbf{d}^{O*}$, as well as the path flows $\boldsymbol{\phi}^*(\boldsymbol{\theta})$ and segment flows $\mathbf{v}^*(\boldsymbol{\theta})$ are all obtained by solving the exact user equilibrium problem (24).

Vehicle flow observations. The observed traffic flow vector $\mathbf{z}$ is generated as $z_s = \max\{v_s^*(\boldsymbol{\theta}) + u_s^v, 0\}$ for all $s \in \mathcal{S}$, where the u_s^v's are drawn in three different ways: (1) i.i.d. from the standard Normal distribution $\mathcal{N}(0, 1)$, (2) i.i.d. from normal distribution with mean 0 and standard deviation 5, denoted $\mathcal{N}(0, 5)$, and (3) from a multivariable Normal distribution with 0 mean, and covariance matrix $\Sigma = H^T H$, where H is a $|\mathcal{S}| \times |\mathcal{S}|$ matrix with each element drawn i.i.d. from the standard Normal distribution.

For illustration purposes, Table 17 presents some summary statistics from a single simulation trial for the equilibrium traffic flows v_s^*, road length l_s, total road width $n_s \cdot w_s$ and the binary bike lane indicator variable x_s.

Table 17 Summary statistics of road segment features in experiments with synthetic data.

count	variable	mean	min	median	max
500	v_s^*	12.75	0.00	9.96	58.15
	l_s	159.24	0.05	148.34	1727.47
	$n_s \cdot w_s$	10.38	3.02	9.05	38.25
	x_s	0.11	0.00	0.00	1.00
1000	v_s^*	29.00	0.00	25.34	114.69
	l_s	158.62	0.46	144.06	1355.00
	$n_s \cdot w_s$	10.28	3.03	8.97	38.71
	x_s	0.09	0.00	0.00	1.00
5000	v_s^*	182.50	0.00	168.81	649.24
	l_s	161.57	0.01	149.52	3968.41
	$n_s \cdot w_s$	10.40	3.01	9.01	44.13
	x_s	0.10	0.00	0.00	1.00

F.2. Estimation

For each of the 50 simulation trials and for each network size $|\mathcal{S}| \in \{500, 1000, 5000\}$, we randomly generated ground-truth values of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ according to the procedure described above, and evaluate the accuracy of the estimates returned by the two-step method described in §3. In particular, for each of the 50 trials, the first step involves estimating $\boldsymbol{\theta}$ from the observe traffic flows $\mathbf{z}$, driving demand $\mathbf{d}^{D*}$, and network features. We denote the estimate by $\hat{\boldsymbol{\theta}}$ and the associated vector of driving time predictions as $\hat{\mathbf{t}}^D = [\hat{t}_w^D]_{w \in \mathcal{W}}$. In the second step, the predicted driving time $\hat{\mathbf{t}}^D$ is plugged into the mode choice model (67), from which we generate an estimate of $\boldsymbol{\beta}$, denoted $\hat{\boldsymbol{\beta}}$.

To estimate $\boldsymbol{\theta}$ in each trial, we first solve the WLS-A problem by applying the BCD algorithm described in §A.3, starting with an initialized solution $\boldsymbol{\theta} = [0, 0, 0, 0]$ and re-solving for each value of λ in $\{50, 100, 250, 500\}$. 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 trial, we select λ as the value that minimizes a simple sum of the three errors defined in (50):

$$\underbrace{\frac{\|\hat{\mathbf{v}} - \mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2}}_{\text{Flow error}} + \underbrace{\frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \frac{\max_{\{p \in \mathcal{P}_w | \hat{\phi}_p > 0\}} \hat{t}_p^D - \min_{p \in \mathcal{P}_w} \hat{t}_p^D}{\min_{p \in \mathcal{P}_w} \hat{t}_p^D}}_{\text{Wardrop error}} + \underbrace{\frac{1}{|\mathcal{W}|} \sum_{w \in \mathcal{W}} \frac{\left|\hat{t}_w^D - t_w^{D*}\right|}{t_w^{D*}}}_{\text{Out-of-sample error}}. \tag{70}$$

In the expression above, the flow error measures the discrepancy between predicted traffic flows $\hat{\mathbf{v}}$ and the observations $\mathbf{z}$; Wardrop error describes the deviation of the prediction from exact Wardrop equilibrium, and the out-of-sample error measures the difference between the predicted driving times $\hat{t}_w^D$, $w \in \mathcal{W}$ and the ground truth driving times t_w^{D*}, $w \in \mathcal{W}$.

We evaluate the performance of the estimation of $\boldsymbol{\theta}$ by computing two quantities: (1) the out-of-sample error in predicting driving times, corresponding to the third term in (50), and (2) the errors in estimating $\boldsymbol{\theta}$, defined as

$$\text{Err}(\boldsymbol{\theta}) = \frac{\left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_2}{\|\boldsymbol{\theta}\|_2}. \tag{71}$$

By plugging the predicted OD travel times $\hat{t}_w^D$ into (67) and applying the linear logarithm transformation in (18), we can then estimate the MNL model by linear regression. We then evaluate the errors in estimating $\boldsymbol{\beta}$ as

$$\text{Err}(\boldsymbol{\beta}) = \frac{\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\|_2}{\|\boldsymbol{\beta}\|_2}. \tag{72}$$

F.3. Results: Estimation and Prediction Errors for Varying Network Sizes

Table 18 reports average estimation errors for $\boldsymbol{\theta}$ over 50 trials, along with the OD driving time prediction errors. The distribution of both errors are visualized in Figure 13. We offer a few remarks on the results. First, when there is less noise in the traffic flow observations (i.e. $\mathcal{N}(0,1)$ vs $\mathcal{N}(0,5)$), the estimation error on both $\boldsymbol{\theta}$ and OD driving time prediction are much lower, as expected; however, the error in estimating $\boldsymbol{\theta}$ in the road network with 5000 segments is still only 7.1% in the $\mathcal{N}(0,5)$ case. We also observe that the error in $\boldsymbol{\theta}$ decreases from 32.1% to 0.9% when the network size grows from 500 to 5000 in the $\mathcal{N}(0,1)$ case; the error in predicting OD driving time also decreases from 4.8% to 0.6%. We observe a similar tendency in the $\mathcal{N}(0,5)$ and $\mathcal{N}(\mathbf{0},\Sigma)$ cases. These decreases in the errors align with the consistency result presented in Theorem 2. In addition, although we assume i.i.d. error in Theorem 2, in the $\mathcal{N}(\mathbf{0},\Sigma)$ case where errors are correlated, the estimation errors continue to decrease in the network size, analogous to the i.i.d. case, and drop to below 1% for $|\mathcal{S}| = 5000$.

Further, because in our §5 empirical study we only observe the driving demand *originating* from each census tract, we ran another set of experiments by using (46) to impute demand at each destination census tract jointly while estimating $\boldsymbol{\theta}$. Using $\mathcal{N}(0,1)$ distributed flow errors and 5000 road segments, the average errors over 50 trials in estimating $\boldsymbol{\theta}$ and predicting OD driving times were 0.87% and 0.59%, respectively; these errors were 0.86% and 0.57% in the setting where destination demands were directly observable. This remarkable similarity in errors suggests that the vehicle flow data and network structure collectively carry significant information about destination demands, meaning imputing them does not substantially increase errors.

We note here that our experimental setting differs from what is assumed by Theorem 2 in two ways: 1) as discussed in §3, for tractability we solve the approximate estimation problem WLS-A instead of the exact estimation problem WLS, and 2) the BCD algorithm is not guaranteed to return a globally optimal solution to either WLS-A or WLS because they are non-convex. Nevertheless, the numerical results in Table 18 and Figure 13 suggest that our estimation procedure is reasonably accurate at recovering the ground-truth congestion parameters, even when the assumptions of Theorem 2 are not strictly satisfied. In summary, these results suggest that solving the approximate estimation problem WLS-A using the block coordinate descent algorithm described in §A.3 produces sound estimates on realistically sized networks.

Next, Table 19 presents the average errors in estimating $\boldsymbol{\beta}$ in the mode choice model, $\text{Err}(\boldsymbol{\beta})$, over 50 trials. The distribution of errors is similarly depicted in Figure 14. We observe that the estimation error in $\boldsymbol{\beta}$ also decreases in the network size, which is due to the OD pair driving time predictions becoming more accurate as the network grows. In addition, we note that when traffic flows are observed with errors that

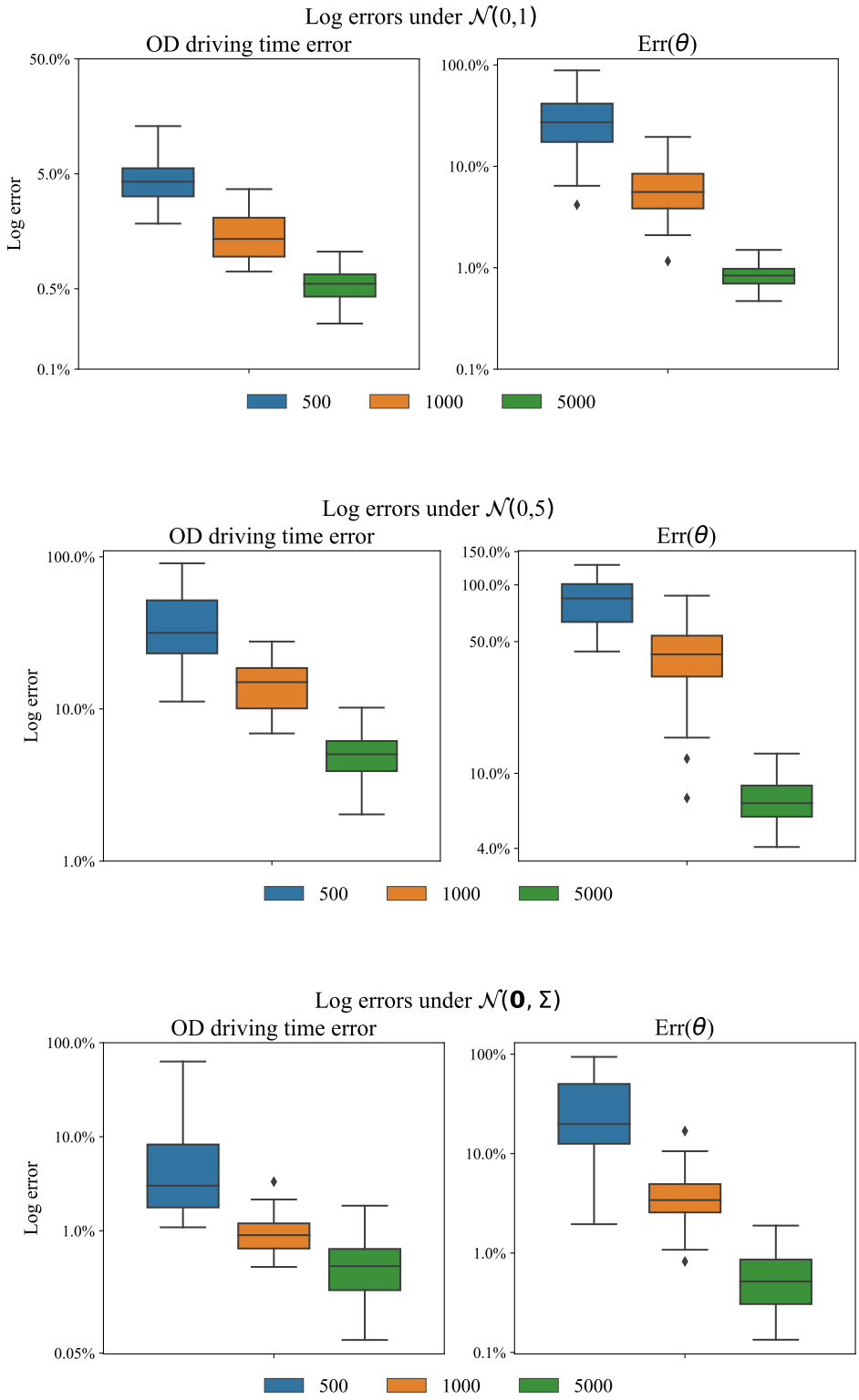


Figure 13 **Distribution of parameter estimation and travel time prediction errors under varying traffic flow perturbations and network sizes (50 trials).**

Table 18 Average parameter estimation and travel time prediction errors under varying traffic flow perturbations and network sizes (50 trials).

		Traffic flow perturbation		
	$\lvert\mathcal{S}\rvert$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,5)$	$\mathcal{N}(\mathbf{0},\Sigma)$
$\text{Err}(\boldsymbol{\theta})$	500	32.13%	84.41%	32.28%
	1000	6.75%	43.21%	4.14%
	5000	0.86%	7.14%	0.65%
OD driving time	500	4.82%	38.28%	10.65%
	1000	1.59%	14.73%	0.99%
	5000	0.57%	5.10%	0.51%

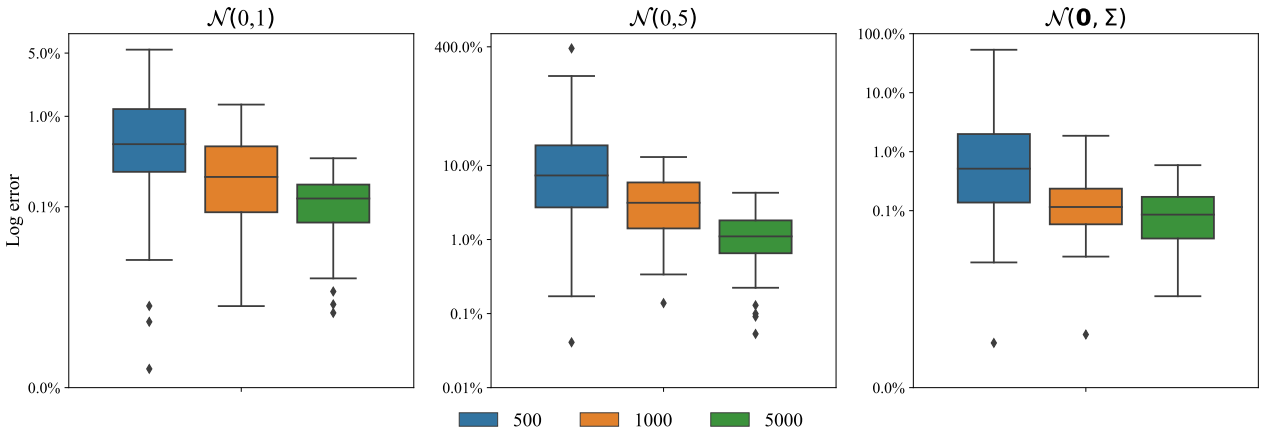


Figure 14 Distribution of mode choice model estimation errors under varying traffic flow perturbations and network sizes (50 trials).

follow $\mathcal{N}(0,5)$, the estimation error $\text{Err}(\boldsymbol{\beta})$ is 3.95% and 1.31% when the number of road segments is 1000 and 5000, respectively.

Lastly, we note that the behavior of the errors shown in Figures 13 and 14 also support our approach of separating the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ into two steps. An important caveat when interpreting these results is that the reported errors correspond to the specific data generation procedure described above, and that alternative procedures may yield different behavior in the errors.

Table 19 Average mode choice model estimation errors ($\text{Err}(\boldsymbol{\beta})$) over 50 trials.

$\lvert\mathcal{S}\rvert$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,5)$	$\mathcal{N}(\mathbf{0},\Sigma)$
500	0.88%	26.33%	4.21%
1000	0.33%	3.95%	0.22%
5000	0.13%	1.31%	0.13%

G. Stochastic User Equilibrium Model

In our model, vehicle flows are assumed to abide by a deterministic user equilibrium (i.e., Wardrop equilibrium), wherein all users choose the path with the shortest driving time. A well-known alternative is the

stochastic user equilibrium (SUE), wherein users have idiosyncratic preferences over different driving paths (Prashker and Bekhor 2004), and path choices also follow an MNL model (i.e., in addition to mode choices). In particular, path flows in the SUE are given by

$$\phi_p = d_w^D \frac{e^{-\zeta \cdot \mathrm{tc}_p}}{\sum_{p' \in \mathcal{P}_w^D} e^{-\zeta \cdot \mathrm{tc}_{p'}}}, p \in \mathcal{P}_w^D, w \in \mathcal{W} \tag{73}$$

where tc_p represents a general travel cost (or disutility) of path p, which itself may depend on the driving time $t_p^D(\boldsymbol{\theta})$ in addition to other route features. The constant ζ reflects the relative importance of travel cost to the random utility term. Note that the resulting traffic assignment, after also considering mode choices, echoes the structure of a nested logit model. Following a parallel argument to Proposition 2, it can be shown that the SUE traffic flows correspond to the set of optimal solutions to the following convex program.

To estimate a SUE-based model from the same data used in our empirical study, one could consider the following estimation problem:

$$\begin{aligned}
\min_{\boldsymbol{\phi}, \mathbf{v}, \boldsymbol{\theta}, \zeta} \ & \|\mathbf{v} - \mathbf{z}\|_2^2 \\
\text{s.t. } v_s = & \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p, \quad s \in \mathcal{S}^D, \\
& \phi_p = d_w^D \frac{e^{-\zeta \cdot \mathrm{tc}_p}}{\sum_{p' \in \mathcal{P}_w^D} e^{-\zeta \cdot \mathrm{tc}_{p'}}}, \quad p \in \mathcal{P}_w^D,\ w \in \mathcal{W}, \\
& (\boldsymbol{\phi}, \mathbf{v}) \geq \mathbf{0}.
\end{aligned}$$

Although we do not implement the SUE-based estimator above, we comment on a few potential challenges in doing so. The first is the complexity of the estimation problem. The formulation above does not benefit from the multi-convex structure of the estimator in §3; as a consequence, different computational methods may be needed to obtain sound estimates. In addition, the identifiability conditions are likely to be substantially different than those used in our setting (see Appendix A.1). For example, if we specify tc_p, in the simplest case, as the driving time on path p, i.e., $\mathrm{tc}_p = t_p^D(\boldsymbol{\theta}) = \sum_{s \in \mathcal{S}_p} (T_s + \alpha_s(\boldsymbol{\theta}) \cdot v_s)$, the estimator contains product terms of the form $\zeta \cdot \theta$, making identification of both ζ and $\boldsymbol{\theta}$ non-trivial. Lastly, embedding the SUE model within our bike lane optimization model – where driving demand d_w^D is also endogenous – may undermine tractability due to the nested logit structure.

A key advantage of a SUE-based formulation is that one can naturally incorporate path-specific attributes into drivers' path choice problem (in addition to idiosyncratic preferences over paths), allowing for a richer model of traffic dynamics than the classic Wardrop equilibrium. Nevertheless, the deterministic user equilibrium and SUE have been observed to yield similar flow patterns in congested networks (Thomas 1991, Prashker and Bekhor 2000), suggesting our main results are unlikely to be drastically different under an SUE model.

H. Proofs

Proof of Proposition 1. 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:

$$\underset{\boldsymbol{\phi}, \mathbf{v}}{\text{minimize}} \ \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot v_s^2 + T_s \cdot v_s \right) \tag{75a}$$

$$\text{subject to} \ \sum_{p \in \mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{75b}$$

$$v_s = \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p, \quad s \in \mathcal{S}, \tag{75c}$$

$$\boldsymbol{\phi}, \mathbf{v} \geq \mathbf{0}. \tag{75d}$$

This equivalence follows from a well-known result from Dafermos and Sparrow (1969) and its proof is omitted for conciseness. Next, we re-write formulation (75) by using a variable change based on the equality in (75c). Substituting $\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p$ for v_s allows us to take the constraint (75c) into the objective (75a), which yields

$$g(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot \left[\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p \right]^2 + T_s \cdot \left[\sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p \right] \right) \tag{76}$$

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

$$\underset{\boldsymbol{\phi}}{\text{minimize}} \ g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right) \tag{77a}$$

$$\text{subject to} \ \sum_{p \in \mathcal{P}_w^D} \phi_p = d_w^D, \quad w \in \mathcal{W}, \tag{77b}$$

$$\boldsymbol{\phi} \geq \mathbf{0}. \tag{77c}$$

Next, expressing the optimality conditions as a variational inequality, a candidate solution $\boldsymbol{\phi}$ is optimal to (77) if and only if the following inequalities hold:

$$\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)^\top \left(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi} \right) \geq 0, \tag{78a}$$

$$\sum_{p \in \mathcal{P}_w^D} \tilde{\phi}_p = d_w^D, \quad w \in \mathcal{W}, \tag{78b}$$

$$\tilde{\boldsymbol{\phi}} \geq \mathbf{0}, \tag{78c}$$

where ∇ is the gradient operator, and the p^{th} element of $\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)$ is

$$\frac{\partial g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)}{\partial \phi_p} = \sum_{s \in \mathcal{S}_p^D} \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot v_s + \sum_{s \in \mathcal{S}_p^D} T_s. \tag{79}$$

Following Theorem 1 of Aghassi et al. (2006), it can be shown using linear programming duality that a candidate solution $\boldsymbol{\phi}$ solves (78) if and only if there exists a vector $\mathbf{b} \in \mathbb{R}^{|\mathcal{W}|}$ such that

$$\sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w^D\}} b_w \leq \nabla_{\phi_p} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right), \quad p \in \mathcal{P}, \tag{80a}$$

$$\nabla_{\boldsymbol{\phi}} g\left(\boldsymbol{\phi}, \boldsymbol{\theta}\right)^\top \boldsymbol{\phi} - \left(\mathbf{d}^D\right)^\top \mathbf{b} = 0. \tag{80b}$$

Therefore, $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$ if and only if $(\boldsymbol{\phi}, \mathbf{v}, \mathbf{b})$ satisfies (75b), (75c), and (80). The result follows because WLS also minimizes $\|\mathbf{v} - \mathbf{z}\|_2^2$ over Θ in addition to satisfying $(\boldsymbol{\phi}, \mathbf{v}) \in \Psi(\mathbf{d}^D, \boldsymbol{\theta})$. □

Proof of Proposition 2. 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 $(\mathbf{d}, \mathbf{v})$ follows from standard arguments using strict convexity of the optimization problem (24). 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 (24). Using $v_s = \sum_{\left\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\right\}} \phi_p,\ s \in \mathcal{S}$ from (1), the user equilibrium problem (24) can be rewritten as

$$\underset{\boldsymbol{\phi} \geq 0, \mathbf{d}}{\text{minimize}}\ S^\dagger(\mathbf{d}, \boldsymbol{\phi}) \tag{81a}$$

$$\text{subject to } \tilde{d}_w = \sum_{m \in \mathcal{M}} d_w^m, \quad w \in \mathcal{W}, \tag{81b}$$

$$d_w^D - \sum_{p \in \mathcal{P}_w} \phi_p^D = 0, \quad w \in \mathcal{W}, \tag{81c}$$

where

$$S^\dagger(\mathbf{d}, \boldsymbol{\phi}) = \sum_{w \in \mathcal{W}} -\left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D d_w^D\right) - \beta_1^D \sum_{s \in \mathcal{S}} \left(\int_0^{\sum_{\left\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\right\}} \phi_p} t_s^D(v) \cdot dv \right) \tag{82a}$$

$$+ \sum_{w \in \mathcal{W}} \left(d_w^D \log\left(d_w^D\right) + d_w^C \log\left(d_w^C\right) + d_w^O \log\left(d_w^O\right)\right). \tag{82b}$$

Let λ_p, μ_w and γ_w be the Lagrangian multipliers for constraints $\phi_p \geq 0, p \in \mathcal{P}^D$, (81b) and (81c), respectively. We can then write the Lagrangian as

$$\mathcal{L}(\mathbf{d}, \boldsymbol{\phi}) = S^\dagger(\mathbf{d}, \boldsymbol{\phi}) - \sum_{p \in \mathcal{P}^D} \lambda_p \phi_p + \sum_{w \in \mathcal{W}} \left(\mu_w \left(d_w^O + d_w^C + d_w^D - \tilde{d}_w \right) + \gamma_w \left(d_w^D - \sum_{p \in \mathcal{P}_w^D} \phi_p^D \right) \right). \tag{83}$$

It is straightforward to verify that $S^\dagger(\mathbf{d}, \boldsymbol{\phi})$ 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 (24) if and only if there exist $\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\gamma}$ that satisfy the Karush-Kuhn Tucker conditions of (24), i.e., the primal feasibility constraints $\phi_p \geq 0$, $p \in \mathcal{P}^D$, (81b) and (81c), 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

$$\frac{\partial \mathcal{L}}{\partial d_w^O} = -u_w^O + \log d_w^O + 1 + \mu_w = 0, \quad w \in \mathcal{W} \tag{84a}$$

$$\frac{\partial \mathcal{L}}{\partial d_w^C} = -u_w^C + \log d_w^C + 1 + \mu_w = 0, \quad w \in \mathcal{W}, \tag{84b}$$

$$\frac{\partial \mathcal{L}}{\partial d_w^D} = -\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D\right) + \log d_w^D + 1 + \mu_w + \gamma_w = 0, \quad w \in \mathcal{W}, \tag{84c}$$

$$\frac{\partial \mathcal{L}}{\partial \phi_p} = -\beta_1^D \sum_{s \in \mathcal{S}_p^D} t_s^D(v_s) - \sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w^D\}} \gamma_w = \lambda_p, \quad p \in \mathcal{P}^D. \tag{84d}$$

Using the definition given in (5) that $t_p^D(\mathbf{v}) = \sum_{s \in \mathcal{S}_p^D} t_s^D(v_s)$, and using the dual feasibility and complementary slackness conditions, it is straightforward to show that (84d) is equivalent to

$$\frac{\partial \mathcal{L}}{\partial \phi_p} = -\beta_1^D \cdot t_p^D(\mathbf{v}) - \sum_{\{w \in \mathcal{W} | p \in \mathcal{P}_w^D\}} \gamma_w \begin{cases} = 0, & \text{if } \phi_p > 0, \\ \geq 0, & \text{if } \phi_p = 0, \end{cases} \quad p \in \mathcal{P}^D. \tag{85}$$

Next, using (6), it can be shown that $\gamma_w = -\beta_1^D \cdot t_w^D$ for all $w \in \mathcal{W}$ satisfies (85), assuming $\beta_1^D < 0$. Then by the definition of driving utility u_w^D in (39a), it follows that $u_w^D = \beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D - \gamma_w$. Next, note that solving (84a)–(84c) and (81b) yields the unique solution

$$d_w^m = \tilde{d}_w \frac{e^{u_w^m}}{e^{u_w^C} + e^{\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D - \gamma_w\right)} + e^{u_w^O}}, \quad m \in \{C, O\}, w \in \mathcal{W}, \tag{86a}$$

$$d_w^D = \tilde{d}_w \frac{e^{\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D - \gamma_w\right)}}{e^{u_w^C} + e^{\left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D - \gamma_w\right)} + e^{u_w^O}}, \quad w \in \mathcal{W} \tag{86b}$$

which implies

$$d_w^m = \tilde{d}_w \frac{e^{u_w^m}}{e^{u_w^C} + e^{u_w^D} + e^{u_w^O}}, \quad m \in \mathcal{M}, w \in \mathcal{W}. \tag{87}$$

Note that $(\mathbf{d}^*, \boldsymbol{\phi}^*)$ is optimal to (81) if and only if $(\mathbf{d}^*, \boldsymbol{\phi}^*, \mathbf{v}^*)$ is optimal to (24), where $v_s^* = \sum_{\{p \in \mathcal{P}^D | s \in \mathcal{S}_p^D\}} \phi_p^*$, for all $s \in \mathcal{S}$ from (1b). Then observing the equivalence between (85) and (6), and between (87) and (10), we conclude that $(\mathbf{d}^*, \boldsymbol{\phi}^*, \mathbf{v}^*)$ is an optimal solution to problem (24) if and only if $(\mathbf{d}, \boldsymbol{\phi}, \mathbf{v}) \in \Gamma(\mathbf{x})$. □

Proof of Theorem 1. We first introduce some additional notation. Let

$$\boldsymbol{\gamma} = \begin{bmatrix} \mathbf{d} \\ \mathbf{v} \end{bmatrix} \tag{88}$$

be the stacked vector of $\mathbf{d}$ and $\mathbf{v}$. Further, for any $\mathbf{x}$, let

$$\boldsymbol{\gamma}^*(\mathbf{x}) = \begin{bmatrix} \mathbf{d}^*(\mathbf{x}) \\ \mathbf{v}^*(\mathbf{x}) \end{bmatrix}, \quad \tilde{\boldsymbol{\gamma}}(\mathbf{x}) = \begin{bmatrix} \tilde{\mathbf{d}}(\mathbf{x}) \\ \tilde{\mathbf{v}}(\mathbf{x}) \end{bmatrix}, \tag{89}$$

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 (24) and the linearized user equilibrium problem (30), 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

$$e = \mu_1 \sqrt{|\mathcal{W}|} \left(\mu_2 \frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi| - 1} + \mu_3 \frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi| - 1} \right), \tag{90}$$

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 $\mathbf{x}$, $R(\mathbf{x}, \boldsymbol{\gamma})$ 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 $\mathbf{x}$, $|\hat{C}(\mathbf{x}, \boldsymbol{\gamma}^*(\mathbf{x})) - \hat{C}(\mathbf{x}, \tilde{\boldsymbol{\gamma}}(\mathbf{x}))| \le e$. Lastly, we use the results from Step 1 and Step 2 to prove the main result. Step 1. By definition of $R(\mathbf{x}, \boldsymbol{\gamma})$, we have

$$R(\mathbf{x}, \boldsymbol{\gamma}) = \nabla_{\boldsymbol{\gamma}} S(\mathbf{x}, \boldsymbol{\gamma}) = \operatorname{diag}\left(\begin{bmatrix} \log(d_w^D) - \left(\beta_0^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D\right) + 1 : w \in \mathcal{W} \\ \log(d_w^C) - u_w^C + 1 : w \in \mathcal{W} \\ \log(d_w^O) - u_w^O + 1 : w \in \mathcal{W} \\ -\beta_1^D (\alpha_s(\mathbf{x}) \cdot v_s + T_s) : s \in \mathcal{S} \end{bmatrix} \right) \tag{91}$$

where $\operatorname{diag}(\mathbf{k})$ is a diagonal matrix whose diagonal elements are vector $\mathbf{k}$. It then follows that

$$\nabla_{\boldsymbol{\gamma}} R(\mathbf{x}, \boldsymbol{\gamma}) = \operatorname{diag}\left(\begin{bmatrix} 1/d_w^m : w \in \mathcal{W}, m \in \mathcal{M} \\ -\beta_1^D \alpha_s(\mathbf{x}) : s \in \mathcal{S} \end{bmatrix} \right). \tag{92}$$

Note that for any fixed $\mathbf{x}$, because $d_w^m \le \bar{d}$ $\nabla_{\boldsymbol{\gamma}} R(\mathbf{x}, \boldsymbol{\gamma})$ 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 (27), $\hat{S}(\mathbf{x}, \boldsymbol{\gamma})$ can be written as

$$\hat{S}\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \underbrace{-\left(\beta_1^D \sum_{s \in \mathcal{S}} \left(T_s \cdot v_s\right) + \beta_1^D \sum_{s \in \mathcal{S}} \xi_s\left(v_s\right)\right)}_{S_1(\mathbf{x}, \boldsymbol{\gamma})} + \underbrace{\sum_{w \in \mathcal{W}} -\left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D d_w^D\right) + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \psi\left(d_w^m\right)}_{S_2(\mathbf{x}, \boldsymbol{\gamma})} \tag{93}$$

Let $R_1\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \nabla_{\boldsymbol{\gamma}} S_1\left(\mathbf{x}, \boldsymbol{\gamma}\right)$ and $R_2\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \nabla_{\boldsymbol{\gamma}} S_2\left(\mathbf{x}, \boldsymbol{\gamma}\right)$. 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 $\tilde{R}\left(\mathbf{x}, \boldsymbol{\gamma}\right) = \nabla_{\boldsymbol{\gamma}} \tilde{S}\left(\mathbf{x}, \boldsymbol{\gamma}\right)$. Because $\hat{S}\left(\mathbf{x}, \boldsymbol{\gamma}\right)$ and $\tilde{S}\left(\mathbf{x}, \boldsymbol{\gamma}\right)$ are both convex in $\boldsymbol{\gamma}$, it follows that

$$\left\langle -R\left(\mathbf{x}, \boldsymbol{\gamma}^*\right), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\rangle \geq 0, \tag{94a}$$

$$\left\langle \tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\rangle \geq 0, \tag{94b}$$

where $\langle \cdot \rangle$ is the Euclidean inner product operator. By summing the two inequalities above, we obtain

$$\left\langle \tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \boldsymbol{\gamma}^*\right), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\rangle \geq 0. \tag{95}$$

Next, by Step 1 and the definition of strong monotonicity, it follows that

$$\left\langle R\left(\mathbf{x}, \boldsymbol{\gamma}^*\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\rangle \geq \frac{1}{\mu_1} \left\| \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\|^2. \tag{96}$$

Combining (95) and (96) yields

$$\left\langle \tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right), \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\rangle \geq \frac{1}{\mu_1} \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|^2. \tag{97}$$

Applying the Cauchy-Schwarz inequality, we have

$$\mu_1 \| \tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) \| \geq \left\| \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\|. \tag{98}$$

Next, using the definition of $\hat{C}(\mathbf{x}, \boldsymbol{\gamma})$ and the Cauchy–Schwarz inequality, we have

$$\left| \hat{C}(\mathbf{x}, \boldsymbol{\gamma}^*) - \hat{C}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) \right| = \left| \sum_{w \in \mathcal{W}} \left(d_w^{C*}(\mathbf{x}) - \tilde{d}_w^C(\mathbf{x})\right) \right| \tag{99a}$$

$$\leq \sqrt{|\mathcal{W}|} \cdot \|\boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}}\|. \tag{99b}$$

It follows from (98) and (99b) and that

$$\left| C\left(\mathbf{x}, \boldsymbol{\gamma}^*\right) - C\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) \right| \leq \sqrt{|\mathcal{W}|} \cdot \left\| \boldsymbol{\gamma}^* - \tilde{\boldsymbol{\gamma}} \right\| \tag{100a}$$

$$\leq \mu_1 \sqrt{|\mathcal{W}|} \cdot \| \tilde{R}\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) - R\left(\mathbf{x}, \tilde{\boldsymbol{\gamma}}\right) \|. \tag{100b}$$

Next, we bound the right hand side of (99b). 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

$$\tilde{S}_1(\mathbf{x}, \boldsymbol{\gamma}) = -\left(\beta_1^D \sum_{s \in \mathcal{S}} \left(t_s^D \cdot v_s\right) + \beta_1^D \sum_{s \in \mathcal{S}} \tilde{\xi}_s\left(v_s\right)\right), \tag{101a}$$

$$\tilde{S}_2(\mathbf{x}, \boldsymbol{\gamma}) = \sum_{w \in \mathcal{W}} -\left(u_w^O d_w^O + u_w^C d_w^C + \beta_0^D d_w^D + (\tilde{\boldsymbol{\beta}}^D)^\top \mathbf{X}_w^D d_w^D\right) + \sum_{w \in \mathcal{W}} \sum_{m \in \mathcal{M}} \tilde{\psi}\left(d_w^m\right). \tag{101b}$$

By differentiating, we obtain

$$\tilde{R}_1(\mathbf{x},\boldsymbol{\gamma}) - R_1(\mathbf{x},\boldsymbol{\gamma}) = \operatorname{diag}\left(\left[\begin{array}{c} \mathbf{0}_{3|\mathcal{W}|} \\ -\beta_1^D\left(\frac{\partial\tilde{\xi}_s(v_s)}{\partial v_s} - \alpha_s(\mathbf{x})\cdot v_s\right) : s\in\mathcal{S}\end{array}\right]\right), \tag{102a}$$

$$\tilde{R}_2(\mathbf{x},\boldsymbol{\gamma}) - R_2(\mathbf{x},\boldsymbol{\gamma}) = \operatorname{diag}\left(\left[\begin{array}{c} \frac{\partial\tilde{\psi}(d_w^m)}{\partial d_w^m} - (\log(d_w^m)+1) : w\in\mathcal{W}, m\in\mathcal{M} \\ \mathbf{0}_{|\mathcal{S}|}\end{array}\right]\right), \tag{102b}$$

where $\mathbf{0}_a$ is the $a\times 1$ vector of zeros. By construction, the linear segments in the piecewise linear function $\tilde{\xi}_s(\cdot)$ have slopes that are equidistant in the interval $\alpha_s(\mathbf{x})\cdot[\underline{v},\bar{v}]$, and the difference between two consecutive slope values is $-\beta_1^D\alpha_s(\mathbf{x})\cdot(\bar{v}-\underline{v})/(|\mathcal{R}_\xi|-1)$. It follows that

$$\left|-\beta_1^D\left(\frac{\partial\tilde{\xi}_s(v_s)}{\partial v_s} - \alpha_s(\mathbf{x})\cdot v_s\right)\right| \le \mu_2\cdot\frac{1}{|\mathcal{R}_\xi|-1} \tag{103}$$

holds for all $s\in\mathcal{S}$, which implies

$$\|\tilde{R}_1(\mathbf{x},\boldsymbol{\gamma}) - R_1(\mathbf{x},\boldsymbol{\gamma})\| = \sqrt{\sum_{s\in\mathcal{S}}\left(-\beta_1^D\left(\frac{\partial\tilde{\xi}_s(v_s)}{\partial v_s} - \alpha_s(\mathbf{x})\cdot v_s\right)\right)^2} \tag{104a}$$

$$\le \mu_2\cdot\frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi|-1}. \tag{104b}$$

Similarly, for any $w\in\mathcal{W}$ and $m\in\mathcal{M}$, because the linear segments in the piecewise linear function $\tilde{\psi}(\cdot)$ 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 $(\log(\bar{d})-\log(\underline{d}))/(|\mathcal{R}_\psi|-1)$. It follows that

$$\frac{\partial\tilde{\psi}(d_w^m)}{\partial d_w^m} - (\log(d_w^m)+1) \le \mu_3\cdot\frac{1}{|\mathcal{R}_\psi|-1} \tag{105}$$

holds for every $w\in\mathcal{W}$ and $m\in\mathcal{M}$, which implies

$$\left\|\tilde{R}_2(\mathbf{x},\boldsymbol{\gamma}) - R_2(\mathbf{x},\boldsymbol{\gamma})\right\| = \sqrt{\sum_{m\in\mathcal{M}}\sum_{w\in\mathcal{W}}\left(\frac{\partial\tilde{\psi}(d_w^m)}{\partial d_w^m} - (\log(d_w^m)+1)\right)^2} \tag{106a}$$

$$\le \mu_3\cdot\frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi|-1}. \tag{106b}$$

Combining inequalities (100b), (104b) and (106b), we obtain

$$|C(\mathbf{x},\boldsymbol{\gamma}^*) - C(\mathbf{x},\tilde{\boldsymbol{\gamma}})| \le \mu_1\sqrt{|\mathcal{W}|}\left\|\tilde{R}(\mathbf{x},\tilde{\boldsymbol{\gamma}}) - R(\mathbf{x},\tilde{\boldsymbol{\gamma}})\right\| \tag{107a}$$

$$\le \mu_1\sqrt{|\mathcal{W}|}\left(\left\|\tilde{R}_1(\mathbf{x},\tilde{\boldsymbol{\gamma}}) - R_1(\mathbf{x},\tilde{\boldsymbol{\gamma}})\right\| + \left\|\tilde{R}_2(\mathbf{x},\tilde{\boldsymbol{\gamma}}) - R_2(\mathbf{x},\tilde{\boldsymbol{\gamma}})\right\|\right) \tag{107b}$$

$$\le \mu_1\sqrt{|\mathcal{W}|}\left(\mu_2\cdot\frac{\sqrt{|\mathcal{S}|}}{|\mathcal{R}_\xi|-1} + \mu_3\cdot\frac{\sqrt{3|\mathcal{W}|}}{|\mathcal{R}_\psi|-1}\right) \tag{107c}$$

$$= e, \tag{107d}$$

as desired. Step 2. We now show $0\le C(\mathbf{x}^*,\mathbf{d}^*(\mathbf{x}^*),\mathbf{v}^*(\mathbf{x}^*)) - C(\tilde{\mathbf{x}},\mathbf{d}^*(\tilde{\mathbf{x}}),\mathbf{v}^*(\tilde{\mathbf{x}})) \le 2e$. Note that from (107d) and the definitions of $\boldsymbol{\gamma}^*$ and $\tilde{\boldsymbol{\gamma}}$, we have

$$C(\tilde{\mathbf{x}},\mathbf{d}^*(\tilde{\mathbf{x}}),\mathbf{v}^*(\tilde{\mathbf{x}})) + e \ge C(\tilde{\mathbf{x}},\tilde{\mathbf{d}}(\tilde{\mathbf{x}}),\tilde{\mathbf{v}}(\tilde{\mathbf{x}})), \tag{108a}$$

$$C(\mathbf{x}^*,\mathbf{d}^*(\mathbf{x}^*),\mathbf{v}^*(\mathbf{x}^*)) - e \le C(\mathbf{x}^*,\tilde{\mathbf{d}}(\mathbf{x}^*),\tilde{\mathbf{v}}(\mathbf{x}^*)). \tag{108b}$$

Next, because $\mathbf{x}^*$ and $\tilde{\mathbf{x}}$ are the optimal solutions to BL and BL-A and $C(\mathbf{x}, \mathbf{d}, \mathbf{v})$ is the objective function of both BL and BL-A, it follows that

$$C(\mathbf{x}^*, \tilde{\mathbf{d}}(\mathbf{x}^*), \tilde{\mathbf{v}}(\mathbf{x}^*)) \leq C(\tilde{\mathbf{x}}, \tilde{\mathbf{d}}(\tilde{\mathbf{x}}), \tilde{\mathbf{v}}(\tilde{\mathbf{x}})), \tag{109a}$$

$$C(\tilde{\mathbf{x}}, \mathbf{d}^*(\tilde{\mathbf{x}}), \mathbf{v}^*(\tilde{\mathbf{x}})) \leq C(\mathbf{x}^*, \mathbf{d}^*(\mathbf{x}^*), \mathbf{v}^*(\mathbf{x}^*)). \tag{109b}$$

Combining (108) and (109), we have

$$0 \leq C(\mathbf{x}^*, \mathbf{d}^*(\mathbf{x}^*), \mathbf{v}^*(\mathbf{x}^*)) - C(\tilde{\mathbf{x}}, \mathbf{d}^*(\tilde{\mathbf{x}}), \mathbf{v}^*(\tilde{\mathbf{x}})) \leq 2e, \tag{110}$$

which concludes the proof. □

Proof of Theorem 2. The proof proceeds in two steps. First, we show that for any $\mathbf{x}$, $\mathbf{v}^*(\mathbf{x}, \boldsymbol{\theta})$ 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. First, the equilibrium segment flows $\mathbf{v}^*(\boldsymbol{\theta})$ are unique by a straightforward extension of a well-known result on Wardrop equilibria (Hall 1978). Thus, $\mathbf{v}^*(\boldsymbol{\theta})$ is a vector-valued function of $\boldsymbol{\theta}$. Next, we prove continuity of $\mathbf{v}^*(\boldsymbol{\theta})$ by following the argument for Proposition 3.1 in Cominetti et al. (2021). First, define $\tau_s(\boldsymbol{\theta}) := t_s^D(\mathbf{v}^*(\boldsymbol{\theta}), \boldsymbol{\theta})$ 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^D, respectively. Next, from the proof of Proposition 1, note that $\mathbf{v}^*(\boldsymbol{\theta})$ is attained at an optimal solution to the convex optimization problem (75). Let the objective function of (75) be defined as:

$$f(\cdot) := \sum_{s \in \mathcal{S}} \left(\frac{1}{2} \cdot \alpha_s(\mathbf{x}, \boldsymbol{\theta}) \cdot v_s^2 + T_s \cdot v_s \right) \tag{111}$$

Next, following Fukushima (1984), for any fixed $\boldsymbol{\theta}$, $\boldsymbol{\tau}(\boldsymbol{\theta})$ is the unique solution to the following strictly convex dual program:

$$\min_{\mathbf{t}} f^*(\mathbf{t}, \boldsymbol{\theta}) - \sum_{w \in \mathcal{W}} \left(d_w^D \cdot \min_{p \in \mathcal{P}_w} \sum_{s \in \mathcal{S}_p^D} t_s^D(\boldsymbol{\theta}) \right) \tag{112}$$

where $f^*(\cdot)$ is the conjugate of $f(\cdot)$ and is strictly convex. By Berge's maximum theorem, because (112) 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 $\boldsymbol{\theta}$. Step 2. Note $L_n(\boldsymbol{\theta}) = \|\mathbf{v}^*(\boldsymbol{\theta}) - \mathbf{z}\|_2^2$. Then because Θ is finite, for each $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ we have

$$L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_0) = \|\mathbf{v}^*(\boldsymbol{\theta}) - \mathbf{z}\|_2^2 - \|\mathbf{v}^*(\boldsymbol{\theta}_0) - \mathbf{z}\|_2^2 \tag{113a}$$

$$= \sum_{s=1}^{n} (v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0) - \varepsilon_s)^2 - \sum_{s=1}^{n} (v_s^*(\boldsymbol{\theta}_0) - v_s^*(\boldsymbol{\theta}_0) - \varepsilon_s)^2 \tag{113b}$$

$$= \sum_{s=1}^{n} (v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0))^2 - 2 \sum_{s=1}^{n} \varepsilon_s (v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0)), \tag{113c}$$

where the final equality 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 $\sum_{s=1}^{n} \varepsilon_s (v_s^*(\boldsymbol{\theta}) - v_s^*(\boldsymbol{\theta}_0)) \longrightarrow 0$ in probability as $n \longrightarrow \infty$. Further,

by condition (iii), $\lim_{n\to\infty}\Pr(\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0))^2>0)=1$, which implies $\lim_{n\to\infty}\Pr(L_n(\boldsymbol{\theta})-L_n(\boldsymbol{\theta}_0)>0)=1$. Note $L_n(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ because $\mathbf{v}^*(\boldsymbol{\theta})$ is continuous by Step 1. The result then follows by Lemma 1 of Wu (1981). □

Proof of Corollary 1. We show that if $\mathcal{G}_n\in\mathcal{X}$, then $\sum_{s=1}^{n}|v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0)|>0$. Let $\mathcal{W}^+\subseteq\mathcal{W}$ be the subset of OD pairs such that for all $\boldsymbol{\theta}\in\Theta$, each $w\in\mathcal{W}^+$ has at least two paths with positive flows in equilibrium, i.e., $\phi_{p_1}^*>0$ and $\phi_{p_2}^*>0$ for some $p_1,p_2\in\mathcal{P}_w^D$, where $(\boldsymbol{\phi}^*(\boldsymbol{\theta}),\mathbf{v}^*(\boldsymbol{\theta}))$ satisfy the equilibrium conditions (7). Note $\mathcal{W}^+$ is non-empty because $\mathcal{G}_n\in\mathcal{X}$ by assumption. Next, recall $\mathcal{S}_p^D$ is the set of segments on path p. By the equilibrium travel time conditions (6), for each $w\in\mathcal{W}^+$ we have $t_{p_1}^D=t_{p_2}^D$, or equivalently,

$$\sum_{s\in\mathcal{S}_{p_1}^D}(\alpha_s(\boldsymbol{\theta})\cdot v_s^*(\boldsymbol{\theta})+T_s)=\sum_{s\in\mathcal{S}_{p_2}^D}(\alpha_s(\boldsymbol{\theta})\cdot v_s^*(\boldsymbol{\theta})+T_s). \tag{114}$$

Next, note $\alpha_s(\boldsymbol{\theta})=\boldsymbol{\theta}^\top\mathbf{q}_s$, and let $\mathbf{A}(\boldsymbol{\theta})$ be a $|\mathcal{W}^+|\times n$ matrix where the entry in (w,s) is given by

$$A_{ws}(\boldsymbol{\theta})=\begin{cases}\boldsymbol{\theta}^\top\mathbf{q}_s, & \text{if}\quad s\in\mathcal{S}_{p_1}^D\setminus\mathcal{S}_{p_2}^D,\\ -\boldsymbol{\theta}^\top\mathbf{q}_s, & \text{if}\quad s\in\mathcal{S}_{p_2}^D\setminus\mathcal{S}_{p_1}^D,\\ 0, & \text{otherwise,}\end{cases} \tag{115}$$

where $p_1,p_2\in\mathcal{P}_w$ and $w\in|\mathcal{W}^+|$. Define $\mathbf{T}$ as the vector with length $|\mathcal{W}^+|$, where the w^{th} entry is $t_w=\sum_{s\in\mathcal{S}_{p_2}^D}T_s-\sum_{s\in\mathcal{S}_{p_1}^D}T_s$ for $p_1,p_2\in\mathcal{P}_w$. Following (114) and the definition of $\mathbf{A}(\boldsymbol{\theta})$, the segment flows $\mathbf{v}^*(\boldsymbol{\theta})$ satisfy $\mathbf{A}(\boldsymbol{\theta})\mathbf{v}^*(\boldsymbol{\theta})=\mathbf{T}$ for each $\boldsymbol{\theta}\in\Theta$. Next, suppose by way of contradiction that $\mathcal{G}_n\in\mathcal{X}$ and there exists $\tilde{\boldsymbol{\theta}}\in\Theta$ such that $\tilde{\boldsymbol{\theta}}\neq\boldsymbol{\theta}_0$ and $\sum_{s=1}^{n}|v_s^*(\tilde{\boldsymbol{\theta}})-v_s^*(\boldsymbol{\theta}_0)|=0$, or equivalently, $\mathbf{v}^*(\tilde{\boldsymbol{\theta}})=\mathbf{v}^*(\boldsymbol{\theta}_0)$. Because $\mathbf{A}(\tilde{\boldsymbol{\theta}})\mathbf{v}^*(\tilde{\boldsymbol{\theta}})=\mathbf{T}$, $\mathbf{A}(\boldsymbol{\theta}_0)\mathbf{v}^*(\boldsymbol{\theta}_0)=\mathbf{T}$, and $\mathbf{v}^*(\tilde{\boldsymbol{\theta}})=\mathbf{v}^*(\boldsymbol{\theta}_0)$, it follows that $(\mathbf{A}(\tilde{\boldsymbol{\theta}})-\mathbf{A}(\boldsymbol{\theta}_0))\mathbf{v}^*(\boldsymbol{\theta}_0)=\mathbf{0}$. Therefore, because $v_s^*(\boldsymbol{\theta}_0)\geq 0$ for all $s\in\mathcal{S}$, for all $w\in\mathcal{W}^+$ we have $\sum_{s=1}^{n}(A_{ws}(\tilde{\boldsymbol{\theta}})-A_{ws}(\boldsymbol{\theta}_0))v_s^*(\boldsymbol{\theta}_0)=0$. By definition of $\mathbf{A}$, this can be written equivalently as

$$\sum_{s\in\mathcal{S}_{p_1}^D\setminus\mathcal{S}_{p_2}^D}((\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^\top\mathbf{q}_s)v_s^*(\boldsymbol{\theta}_0)+\sum_{s\in\mathcal{S}_{p_2}^D\setminus\mathcal{S}_{p_1}^D}((\boldsymbol{\theta}_0-\tilde{\boldsymbol{\theta}})^\top\mathbf{q}_s)v_s^*(\boldsymbol{\theta}_0)=0.$$

However, because $\mathcal{G}_n\in\mathcal{X}$, there must exist $w\in\mathcal{W}^+$ such that (44) holds, which contradicts the expression above. We conclude that $\mathcal{G}_n\in\mathcal{X}$ implies $\mathbf{v}^*(\boldsymbol{\theta})\neq\mathbf{v}^*(\boldsymbol{\theta}_0)$ and thus $\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}))^2>0$ for all $\boldsymbol{\theta}\neq\boldsymbol{\theta}_0$, as desired. □

Proof of Corollary 2. The proof proceeds similarly to the proof of Theorem 2. For each $\delta>0$, we have

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}(L_n(\boldsymbol{\theta})-L_n(\boldsymbol{\theta}_0))=\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}\left(\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0))^2-2\sum_{s=1}^{n}\varepsilon_s(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0))\right) \tag{116a}$$

$$\geq\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0))^2-\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}2\sum_{s=1}^{n}\varepsilon_s(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0)) \tag{116b}$$

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 $\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}(\sum_{s=1}^{n}\varepsilon_s(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0)))\longrightarrow 0$ in probability as $n\longrightarrow\infty$. Further, because $\lim_{n\to\infty}\Pr(\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}\sum_{s=1}^{n}(v_s^*(\boldsymbol{\theta})-v_s^*(\boldsymbol{\theta}_0))^2>0)=1$ by assumption, we conclude $\lim_{n\to\infty}\Pr(\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\delta}(L_n(\boldsymbol{\theta})-L_n(\boldsymbol{\theta}_0))>0)=1$. Note $L_n(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$ because $\mathbf{v}^*(\boldsymbol{\theta})$ is continuous by Step 1 of the proof of Theorem 2. The result follows by Lemma 1 of Wu (1981). □