Final Report:

A general traffic equilibrium framework with ride-hailing services that considers flow-dependent waiting time and public transit

May 2024

A Research Report from the Pacific Southwest Region University Transportation Center

Jong-Shi Pang, University of Southern California

Maged Dessouky, University of Southern California

Wei Gu, University of Southern California

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. PSR 22-15 TO 066	**2. Government Accession No.** N/A	**3. Recipient's Catalog No.** N/A
4. Title and Subtitle A general traffic equilibrium framework with ride-hailing services that considers flow-dependent waiting time and public transit		**5. Report Date** May 2024 **6. Performing Organization Code** N/A
7. Author(s) Jong-Shi Pang, Ph.D. Maged Dessouky, Ph.D. Wei Gu		**8. Performing Organization Report No.** PSR 22-15 TO 066
9. Performing Organization Name and Address University of Southern California METRANS Transportation Consortium University Park Campus, RGL 216 Los Angeles, CA 90089-0626		**10. Work Unit No.** N/A **11. Contract or Grant No.** Caltrans 65A0674
12. Sponsoring Agency Name and Address Caltrans DRISI 1727 30th Street, MS 82 Sacramento, CA 95816		**13. Type of Report and Period Covered** Research Report (June 2023 - May 2024) **14. Sponsoring Agency Code** USDOT OST-R

15. Supplementary Notes
Project page: https://www.metrans.org/research/a-general-traffic-equilibrium-framework-with-ride-sourcing-services-that-considers-flow-dependent-waiting-time-and-public-transit

16. Abstract
We develop a general traffic equilibrium model that considers ride-hailing services provided by Transportation Network Companies (TNCs, e.g., Uber and Lyft) and customer waiting. The equilibrium model integrates three interacting sub-models, including TNCs' decision on dispatching ride-hailing vehicles, travelers' choice of which mode to use, and queueing dynamics that explicitly capture customers' waiting costs. The TNCs' choices and customers' choices together form a generalized Nash equilibrium, coupled with the queueing system of customer waiting. We provide the conditions under which there exists an equilibrium solution. Then the proposed model is validated in the Sioux-Falls networks. Numerical experiments show that compared with our method, an existing method and a linear waiting cost function tend to underestimate the waiting cost of ride-hailing customers. As a result, they overestimate the mode share of ride-hailing customers and the vehicle miles traveled in the system.

17. Key Words Traffic equilibrium, Ride-hailing, Customer waiting, Generalized Nash, Queueing	**18. Distribution Statement** No restrictions.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 49	**22. Price** N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Contents

About the Pacific Southwest Region University Transportation Center

The Pacific Southwest Region University Transportation Center (UTC) is the Region 9 University Transportation Center funded under the US Department of Transportation's University Transportation Centers Program. Established in 2016, the Pacific Southwest Region UTC (PSR) is led by the University of Southern California and includes seven partners: Long Beach State University; University of California, Davis; University of California, Irvine; University of California, Los Angeles; University of Hawaii; Northern Arizona University; Pima Community College.

The Pacific Southwest Region UTC conducts an integrated, multidisciplinary program of research, education and technology transfer aimed at improving the mobility of people and goods throughout the region. Our program is organized around four themes: 1) technology to address transportation problems and improve mobility; 2) improving mobility for vulnerable populations; 3) Improving resilience and protecting the environment; and 4) managing mobility in high growth areas.

U.S. Department of Transportation (USDOT) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

California Department of Transportation (CALTRANS) Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the United States Department of Transportation's University Transportation Centers program, in the interest of information exchange. The U.S. Government and the State of California assumes no liability for the contents or use thereof. Nor does the content necessarily reflect the official views or policies of the U.S. Government and the State of California. This report does not constitute a standard, specification, or regulation. This report does not constitute an endorsement by the California Department of Transportation (Caltrans) of any product described herein.

Disclosure

Principal Investigator, Co-Principal Investigators, others, conducted this research titled, "A general traffic equilibrium framework with ride sourcing services that considers flow-dependent waiting time and public transit" at the Daniel J. Epstein Department of Industrial & System Engineering, Viterbi School of Engineering, University of Southern California. The research took place from 06/01/2023 to 05/31/2024 and was funded by a grant from the United States, Department of Transportation, in the amount of $xxx,xxx. The research was conducted as part of the Pacific Southwest Region University Transportation Center research program.

Acknowledgements

We acknowledge the Pacific Southwest Region University Transportation Center for their financial support of this research.

Abstract

We develop a general traffic equilibrium model that considers ride-hailing services provided by Transportation Network Companies (TNCs, e.g., Uber and Lyft) and customer waiting. The equilibrium model integrates three interacting sub-models, including TNCs' decision on dispatching ride-hailing vehicles, travelers' choice of which mode to use, and queueing dynamics that explicitly capture customers' waiting costs. The TNCs' choices and customers' choices together form a generalized Nash equilibrium, coupled with the queueing system of customer waiting. We provide the conditions under which there exists an equilibrium solution. Then the proposed model is validated in the Sioux-Falls networks. Numerical experiments show that compared with our method, an existing method and a linear waiting cost function tend to underestimate the waiting cost of ride-hailing customers. As a result, they overestimate the mode share of ride-hailing customers and the vehicle miles traveled in the system.

Executive Summary

The rapid rise of ride-hailing services provided by Transportation Network Companies (TNCs) such as Uber, Lyft, DiDi, Grab, and Ola are transforming the travel behavior of individuals and urban mobility patterns. While these emerging shared mobility services enriched the user experience by providing more mobility options, they also raise challenges for transportation planners in terms of analyzing the system: (1) how to model precisely the waiting cost of ride-hailing customers and the deadheading miles incurred by ride-hailing vehicles; (2) how to capture the complex interactions between TNCs, travelers, and ride-hailing customer waiting.

In response to these challenges, we propose a general equilibrium model that includes ride-hailing services and customer waiting in urban transportation networks. We develop three modules of our proposed model: (1) TNC choice module: each TNC decides the dispatching policy of ride-hailing drivers that maximizes its total profit, subject to the supply of ride-hailing drivers, the flow conservation equations, and the fleet size constraint of each TNC; (2) Traveler choice module: rational travelers choose the mode (i.e., ride-hailing, solo driving, or public transit) that minimizes their disutility. The waiting cost of ride-hailing customers is part of their disutility; (3) Customer waiting module: we consider the ride-hailing system as a queueing system, in which the ride-hailing customers arrive in the system, and wait for the ride-hailing drivers to serve them. The closed-form waiting cost functions are derived, as functions in terms of TNCs' dispatching flows and travelers' choice flows. The three modules interact with each other: the TNC choice module and the traveler choice module together form a generalized Nash equilibrium, coupled with the queueing dynamics of the customer waiting module.

Then we analyze the mathematical properties of the overall equilibrium model. We provide the conditions under which an equilibrium solution exists for our model. Furthermore, in order to solve the equilibrium problem, we derive the equivalent mixed equivalent complementarity formulation of the proposed model.

Finally, our proposed model is validated using the Sioux-Falls network. The method of Ban et al. (2019) and a linear waiting cost function are used to compare with our model. Numerical results show that without modeling the waiting cost explicitly, Ban et al. (2019) underestimates the waiting cost, and as a result, overestimates the mode share of ride-hailing travelers and VMT in the system. For example, Ban et al. (2019) overestimates VMT by 63.8% compared with our model. Similarly, if we use a linear waiting cost function, it will underestimate the waiting cost by 29.9% compared with our method. Consequently, a linear waiting cost function would output 50% more mode share of ride-hailing travelers and overestimate the VMT by 40.7%. With the proposed general modeling framework, transportation planners can better understand ride-hailing customers' waiting costs and deadheading miles induced by ride-hailing vehicles. This could help policymakers develop appropriate incentives that capture the features of ride-hailing services to reduce congestion in the transportation system.

1 Introduction

Ride-hailing services, provided by Transportation Network Companies (TNCs) such as Uber, Lyft, DiDi, Grab, Ola, etc., have achieved substantial growth and received much attention in recent years. For example, Uber drivers completed 9.4 billion trips in 2023, which is 23% more than that of 2022 (Iqbal, 2024). In spite of the proliferation of ride-hailing services and the convenience they bring to travelers, recent evidence shows that the expansion of the ride-hailing market deteriorates traffic conditions in some major cities (Schaller, 2018; Erhardt et al., 2019). For instance, the growth of ride-hailing services leads to a 36% increase of vehicle miles traveled (VMT) in New York City between 2013 and 2017 (Schaller, 2018), and contributes to 47% of increase in VMT in San Francisco from 2010 to 2016 (Castiglione et al., 2018).

One of the main reasons that ride-hailing may cause more traffic congestion is that, after each driver drops off a customer, (s)he may cruise on the street until matched to a new customer, and then travels from his/her location to the customer's current location for pick-up. This feature distinguishes ride-hailing services from (A) the traditional street-hailing taxis (e.g., Yang and Wong, 1998; Wong et al., 2008), in which the drivers will roam the streets until they meet the next customer; (B) the emerging ridesharing services (e.g., Xu et al., 2015; Li et al., 2020), in which drivers share the vehicles as well as the travel costs with passengers who have similar itineraries and time schedules. The trips before ride-hailing drivers to pick up customers lead to deadhead miles (i.e., empty miles). When drivers are scarce compared with demand in the ride-hailing market, the TNCs may send drivers unnecessarily far away to pick up customers. There exists an inefficient equilibrium in steady-state operations, which is referred to as the "wild goose chase" (WGC) phenomenon and verified both theoretically and empirically by Castillo et al. (2017, 2023). Furthermore, to balance supply and demand in the ride-hailing market, some TNCs provide incentives to motivate drivers to pick up customers in possibly distant locations with low driver density (Yan et al., 2019; Ma et al., 2022). Consequently, the deadheading miles could be large. This has been shown in empirical studies, e.g., Henao and Marshall (2019) reveals that the deadheading miles of ride-hailing contribute to over 40% of the overall vehicle miles traveled in a city.

From the customers' perspective, large deadhead miles could lead to long waiting for the ride-hailing vehicles to arrive, as shown by recent evidence in metropolitan areas such as Los Angeles (Allyn, 2021), Boston (Rosenberg, 2021), and Washington DC (Spiegel, 2024). Long waiting of ride-hailing customers happens especially due to the shortage of drivers or the inefficient dispatching policy of TNCs (Castro et al., 2021). An important feature of the waiting costs in the ride-hailing market is that they change with the flow of drivers and choice of customers, due to various reasons such as change of supply and demand or different traffic conditions. In turn, the waiting cost influences the decision of customers. If the waiting cost is too large, a customer may cancel the trip or stick with traditional transportation modes such as solo driving or public transit. As an alternative mode choice, ride-hailing makes the behavior of travelers much more complicated due to its interactions with the traditional modes.

With the significance of the deadheading miles and customer waiting caused by ride-hailing services, it is important for transportation planners to capture the deadheading miles and quantify the corresponding waiting cost of customers accurately in their decision-making. There is a clear need to quantify and understand the impact of waiting costs on customers, and the influence of deadhead miles on the transportation system. However, prior research models on ride-hailing systems have not captured the complexity between travelers' and TNCs' choices, and ride-hailing customers' waiting costs, together with the complicated interactions between ride-hailing, solo driving, and public transit.

To fill this research gap, in this study, we develop a general traffic equilibrium framework with ride-hailing services and customer waiting, which includes three interacting modules: (A) TNC choice module: each TNC decides the dispatching flow of ride-hailing vehicles to maximize its own profit; (B) Traveler choice module: each traveler decides the mode (ride-hailing, solo driving, or public transit) that minimizes the travel cost; (C) Customer waiting module: the dynamics of ride-hailing customers' waiting cost in terms of ride-hailing vehicles' dispatching flows and travelers' choice flows. The TNC choice module and the traveler choice module together form a generalized Nash equilibrium, coupled with a customer waiting module formulated as a queueing system. We show there exists a solution for the proposed general traffic equilibrium framework with ride-hailing services and customer waiting, and verify the proposed model on the Sioux-Falls network.

The remainder of this report is organized as follows. In Section 2, we review related literature. Section 3 describes the problem setting, and Section 4 introduces the bipartite network structure on which our model is built. The three interacting modules of our model, namely the TNC choice module, the traveler choice module, and the customer waiting module, are presented in Sections 5, 6, and 7, respectively. Section 8 summarizes the previous section to propose the overall equilibrium model and relevant mathematical analysis. Numerical experimental results are given in Section 9 to illustrate our model. Section 10 concludes this report and points out some possible directions for future research.

2 Literature Review

Ride-hailing systems have been studied extensively in recent years (Wang and Yang, 2019; Yan et al., 2019). In this section, we review the literature that is most related to our research: equilibrium analysis in ride-hailing systems and queueing analysis in ride-hailing systems.

2.1 Equilibrium Analysis in Ride-hailing Systems

Since there are different parties in the ride-hailing systems and they have different objectives (e.g., customers want to minimize their costs while the TNCs aim to maximize their profit), naturally there exist equilibria in the system. One research direction is to analyze the market equilibria between supply and demand. Under the scenario of a mixed ride-hailing and taxi market, He and Shen (2015) established a spatial equilibrium model to balance supply and demand, and at the same time, evaluated travelers' possible adoption of the emerging ride-hailing service; Qian and Ukkusuri (2017) investigated the equilibrium of the competitive market by modeling it as a multiple-leader-follower game: customers are the leaders who aim to minimize their cost, while drivers are the followers seeking to maximize their profit. Zha et al. (2018) provided a matching model in which customers could be matched to a distant vehicle and then analyzed the market equilibrium with spatial pricing and a TNC that maximizes the revenue. Ke et al. (2020a) explored the effects of key decision variables of a ride-hailing platform (such as price and vehicle fleet size) on its revenue and social welfare. Ke et al. (2020b) proposed a ride-hailing market equilibrium model with congestion effects, with a macroscopic fundamental diagram to characterize traffic congestion. Zhang and Nie (2021) put forward a matching-based market equilibrium model to explore the influence of regulation on both ride-hailing and ridesharing services.

By extending the traditional traffic equilibrium problem (Wardrop, 1952; Sheffi, 1985; Patriksson, 2015), another research direction is to study traffic equilibria with ride-hailing services. Ban et al. (2019) proposed the first traffic equilibrium model with ride-hailing, in which the TNCs decide the optimal dispatch of drivers to maximize the profit, and travelers choose the mode (ride-hailing or solo driving) that minimizes the travel cost. Di and Ban (2019) modeled the interaction between ride-hailing and ridesharing services. Xu et al. (2021) extend the modeling framework of Yang and Wong (1998) to capture the interactions between idle and occupied ride-hailing vehicles, as well as the matching between customers and idle drivers between different locations. Chen and Di (2023) put forward a traffic equilibrium model with ride-pool services, namely multiple customers could share the same ride-hailing vehicle.

In the traffic equilibrium problem with ride-hailing services, the choice of travelers will be influenced by the waiting cost of being a ride-hailing customer after requesting the service. At the same time, both travelers' mode choice and TNCs' choice of vehicle dispatching will impact ride-hailing customers' waiting costs. Our contribution to this set of literature is to first model explicitly the waiting cost of ride-hailing customers and capture its interactions with travelers' and TNCs' choices in traffic equilibrium with ride-hailing services.

2.2 Queueing Analysis in Ride-hailing Systems

In the literature, ride-hailing services are often analyzed as a queueing system, which captures the dynamics between customers' arrival and drivers' service. Based on analytical queueing networks, Daganzo and Ouyang (2019) showed that an inefficient equilibrium (namely the "wild goose chase" phenomenon) could exist for multiple types of demand-responsive transportation services including ride-hailing, and Ouyang and Yang (2023) derived approximate formulas for expected system performance under steady state and proposed a strategy based on vehicle swaps to mitigate the inefficient equilibrium. Xu et al. (2020) constructed a double-ended queueing model to analyze the supply curve of ride-hailing systems that captures the relationship between the throughput and cost of customers. Feng et al. (2021) compared the average waiting times of customers under different matching mechanisms via a stylized model of a circular road. Feng et al. (2022) partitioned the entire ride-hailing system into various small blocks, and queueing-based matching was implemented in each block separately to determine the block size.

To optimize the matching process in ride-hailing systems, Braverman et al. (2019) modeled a closed queueing network to optimize the dispatch of empty vehicles. The convergence of the queueing network to a fluid limit was established in a large market regime. Özkan and Ward (2020) formulated a continuous linear program to optimize the matching policies and considered how long customers and drivers are willing to wait. Asymptotic optimality was proved in a large market regime. Sun et al. (2020) used an approximate queueing analysis for the assignment of the TNC and the selection of drivers between different matching policies. With an approximation of pickup time in terms of the number of waiting customers and idle drivers, Wang et al. (2024) analyzed the queueing dynamics between customers and drivers to capture impatient passengers, given that the TNC can set the threshold on the expected pickup time to control the matching process.

Some papers studied the optimal pricing for customers and optimal wage for drivers. Taylor (2018) analyzed how customers' sensitivity to delays and drivers' independence could influence TNC's optimal pricing and wage. Bai et al. (2019) derived the optimal pricing and wage that optimizes TNC's profit while a queueing-based steady-state waiting time was included in the customer's utility function. Li et al. (2019) proposed a queueing theoretic model to capture the ride-hailing market equilibrium between customer's payment, driver's wage, and incentives for both sides. Furthermore, they evaluated the impact of different regulations on TNCs. Considering the heterogeneity of both drivers and customers, Zhong et al. (2023) adopted a two-stage queueing game that jointly analyzes the behaviors of drivers and customers and derives the TNC's optimal decisions on pricing and wage.

If we consider ride-hailing services as a queueing system, travelers' mode choice (ride-hailing, solo driving, or public transit) and TNCs' choice of vehicle dispatching and matching influence the arrival rate and service rate of the queue, respectively. While the queueing dynamics will, in turn, impact travelers' choice since it is related to the waiting cost of ride-hailing customers. To the best of our knowledge, there is no research that captures the queueing dynamics of the ride-hailing system, together with the equilibrium between travelers' and TNCs' choices. We provide the first attempt in this direction.

3 Problem Description

We consider a transportation system with emerging ride-hailing services and the waiting of ride-hailing customers in an urban network. With the ride-hailing services providing more flexibility for travelers, in this system, travelers are rational cost minimizers who can choose to either use the ride-hailing service from one of the Transportation Network Companies (TNCs, e.g., Uber and Lyft), or stick to one of the traditional modes of transport: solo driving and public transit. At the same time, the TNCs want to decide the best policy for picking up their customers in order to maximize total revenue, subject to fleet size constraints and travelers' choices (i.e., how many travelers are choosing each TNC). The decisions of TNCs and travelers will influence the supply and demand in the ride-hailing market, respectively, and as a result, have an effect on the waiting cost of ride-hailing customers. This waiting cost, in turn, will impact travelers' mode choice between ride-hailing, solo driving, and public transit since it is part of the generalized cost for ride-hailing customers. The scenario that we consider in this study is illustrated in Fig. 1. There are three components in our model: (A) TNCs' choice of how to pick up their customers; (B) Travelers' choice of which travel mode to use (solo driving, one of the TNCs, or public transit); (C) Waiting of ride-hailing customers that captures the queueing dynamics between the dispatching of drivers and the choice of travelers.

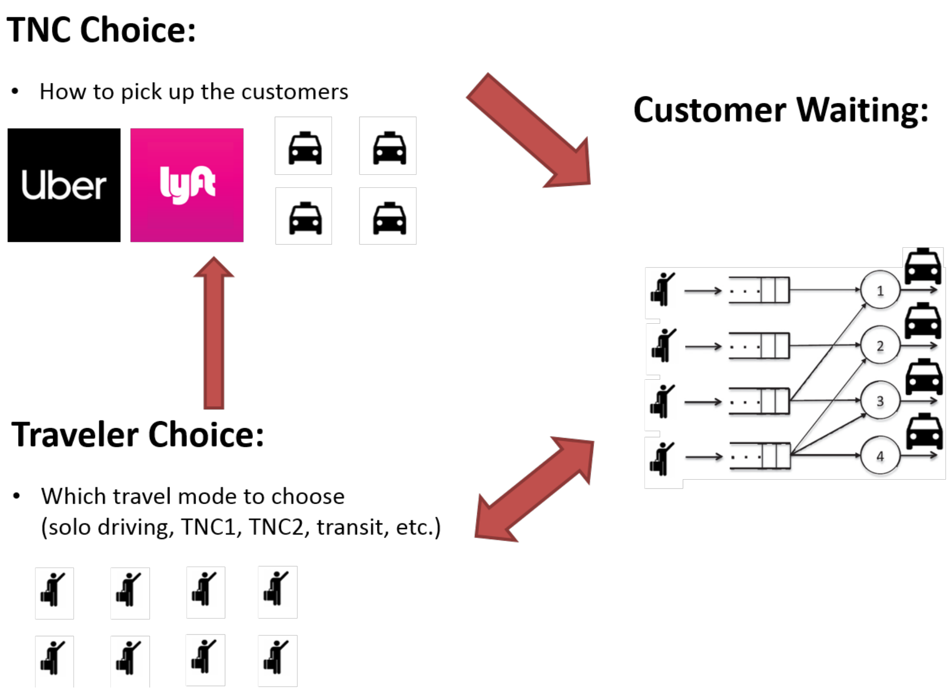

Figure 1: Graphical Representation of the Problem.

In order to balance model realism and mathematical tractability, the following assumptions are made in our model:

- The model describes certain stable equilibrium and traffic patterns to serve transportation planners and policymakers;
- Travel times on the network are known and fixed;
- The fleet size of each TNC is predetermined;
- The drivers from the same TNC are cooperative towards the TNC's benefit. This assumption can be relaxed to capture strategic/selfish drivers by considering each driver as their own TNCs, but the computation complexity of the model would significantly increase.

4 Bipartite Network Structure

To keep the model computationally tractable to solve, we assume the travel time from origin to destination is fixed and given and vice versa. Making use of of this assumption, we next show how to construct a bipartite network that contains the origins and destinations of the original transportation network as nodes and also incorporates information on the travel times between origins and destinations as edge lengths. We will see that such a bipartite network is sufficient for this formulation when the travel times are known and given.

Consider a transportation network with a set $\mathcal{O}$ of origins and a set $\mathcal{D}$ of destinations. Without loss of generality, we assume that $\mathcal{O} \cap \mathcal{D} = \emptyset$. Otherwise, we can replicate every node in $\mathcal{O} \cap \mathcal{D}$ into two nodes, with one acting as an origin and the other as a destination, and then add a directed edge of length 0 between them. Given the travel times from the origins to the destinations and the travel times from the destinations to the origins to be known, we can convert the original transportation network into a bipartite network $\mathcal{G} \triangleq (\mathcal{O} \cup \mathcal{D}, \mathcal{E} \cup \widetilde{\mathcal{E}})$, where $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ are two types of edges in the two directions: $\mathcal{E}$ represents the set of edges from an origin $i \in \mathcal{O}$ to a destination $j \in \mathcal{D}$ and $\widetilde{\mathcal{E}}$ is for the set of edges from a destination $j \in \mathcal{D}$ to an origin $i \in \mathcal{O}$. Fig. 2 shows an example of such a bipartite network, where edges in $\mathcal{E}$ and $\widetilde{\mathcal{E}}$ are shown in solid and dashed edges, respectively.

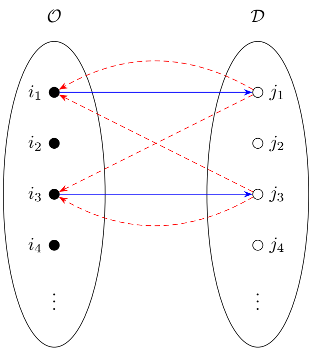

Figure 2: An illustration of the bipartite network.

Without loss of generality, we assume that the original transportation network is strongly connected. Then travelers can go from any origin $i \in \mathcal{O}$ to any destination $j \in \mathcal{D}$, and it is possible for a ride-hailing vehicle from an arbitrary destination $j \in \mathcal{D}$ to pick up a customer at an arbitrary origin $i \in \mathcal{O}$. Thus, $\mathcal{E}$ contains all possible origin-destination (OD) pairs while $\widetilde{\mathcal{E}}$ contains all possible destination-origin pairs, namely, $|\mathcal{E}| = |\mathcal{O}| \times |\mathcal{D}|$ and $|\widetilde{\mathcal{E}}| = |\mathcal{D}| \times |\mathcal{O}|$. Note that by choosing the set $\widetilde{\mathcal{E}}$ properly, the simplified bipartite network can incorporate the case that a platform only informs ride-hailing drivers within a certain distance of a customer for the service request.

The sets, model parameters, and decision variables we use in this study are summarized in Table 1 below.

Table 1. Notations used in this study.

Sets and model parameters	
$\mathcal{O}$, $\mathcal{D}$	Set of travelers' origins and destinations
$\mathcal{E}$, $\widetilde{\mathcal{E}}$	Set of edges in bipartite network from origins to destinations and from destinations to origins, respectively
$\mathcal{M}$	Labels of the TNCs; $\mathcal{M} \triangleq \{1, 2, \ldots, M\}$
$\mathcal{M}_+$	Union of the solo driver label (0), the TNC labels, and the public transit label $(M+1)$; $\mathcal{M}_+ \triangleq \mathcal{M} \cup \{0\} \cup \{M+1\}$
Q_{ij}	Given travel demand rate (trips per unit time) from origin $i \in \mathcal{O}$ to destination $j \in \mathcal{D}$
N^m	The fleet size (namely the total number of vehicles) of TNC $m \in \mathcal{M}$
$\underline{\alpha}_i^m$	Fixed fare charged by TNC $m \in \mathcal{M}$ who picks up customers at origin $i \in \mathcal{O}$
α_1^m, α_2^m	Time and distance based fare rates, respectively, for TNC $m \in \mathcal{M}$
β_1^m, β_2^m	Respective conversion factors from travel time and distance to costs for the drivers of TNC $m \in \mathcal{M}$
γ_1^m, γ_2^m	Value of time for customers of TNC $m \in \mathcal{M}$ while waiting for and traveling in a vehicle, respectively
α_2^0, γ_2^0	Distance based cost rates and value of time while traveling for solo drivers, respectively
t_{ij}	Travel time for all edges $(i,j) \in \mathcal{E} \cup \widetilde{\mathcal{E}}$ in the bipartite network, which is the travel time from i to j in the original transportation network
d_{ij}	Distance of the shortest path for all edges $(i,j) \in \mathcal{E} \cup \widetilde{\mathcal{E}}$ in the bipartite network
Decision variables	
y_{ij}^m	Demand flow of travelers using mode $m \in \mathcal{M}_+$ on edge $(i,j) \in \mathcal{E}$
z_{kij}^m	Service flow of ride-hailing vehicles for TNC $m \in \mathcal{M}$ that are dispatched from destination $k \in \mathcal{D}$ to serve ride-hailing customers on edge $(i,j) \in \mathcal{E}$
w_{ij}^m	Waiting cost of ride-hailing customers using TNC $m \in \mathcal{M}$ on edge $(i,j) \in \mathcal{E}$
$\mathcal{Q}_{ij}^m(\cdot)$	Mapping of waiting cost of ride-hailing customers using TNC $m \in \mathcal{M}$ on edge $(i,j) \in \mathcal{E}$, as a function of demand flows and pickup flows

5 TNC Choice Module

In this study, we consider a market with multiple Transportation Network Companies (TNCs). Each TNC decides the dispatching policy that maximizes its total profit, while the drivers for each TNC are assumed to be cooperative towards that TNC's objective. Let z^m_{kij} be the pickups that originate from destination k (after the previous drop-off) to go pick up the demands of OD pair (i, j). Specifically, in the TNC choice module, each TNC $m \in \mathcal{M}$ aims at finding a feasible pickup flow $z^m \triangleq \left(z^m_{kij}\right)_{(k,i,j)\in\mathcal{D}\times\mathcal{O}\times\mathcal{D}}$ that maximizes its total profit, which is defined as the difference between its total revenue for serving customers and total cost for picking up customers.

A complete service trip for each ride-hailing vehicle is associated with three nodes $k \in \mathcal{D}, i \in \mathcal{O}$ and $j \in \mathcal{D}$. The service trip kij contains two parts: the first part is the pickup trip from the destination $k \in \mathcal{D}$, where the ride-hailing vehicle stays, to the origin $i \in \mathcal{O}$, where a customer requests services; and the second part is the delivery trip that takes the customer from the origin $i \in \mathcal{O}$ to his or her destination $j \in \mathcal{D}$. The pickup trip (k, i) itself does not make any profit, but only incurs the following per-pickup cost that depends on the pickup time t_{ki} and the pickup distance d_{ki}:

$$\underbrace{\beta^m_1 t_{ki}}_{\text{time based cost}} + \underbrace{\beta^m_2 d_{ki}}_{\text{distance based cost}}$$

The delivery trip (i, j) fulfills the customer's service request and generates the following per-customer revenue that includes a fixed passenger delivery revenue, an amount per mile to cover driving costs, and an amount per unit time to cover net driver pay after expenses:

$$\underline{\alpha}^m_i + \underbrace{\alpha^m_1 t_{ij}}_{\text{time based revenue}} + \underbrace{\alpha^m_2 d_{ij}}_{\text{distance based revenue}} - \underbrace{\beta^m_1 t_{ij}}_{\text{time based cost}} - \underbrace{\beta^m_2 d_{ij}}_{\text{distance based cost}}$$

Given travelers' demand flow $\left\{ y^m_{ij} : (i, j) \in \mathcal{E} \right\}$, the total profit for TNC $m \in \mathcal{M}$ is the amount of revenue from delivery trip (i, j) minus the cost generated by the pickup trip (k, i). Thus, for each TNC $m \in \mathcal{M}$, to maximize the overall profit is formulated as below:

$$\underset{z^m_{kij}\geq 0}{\textbf{maximize}} \sum_{(i,j)\in\mathcal{E}} \sum_{k\in\mathcal{D}} \left[\underbrace{\left(\underline{\alpha}^m_i + \alpha^m_1 t_{ij} + \alpha^m_2 d_{ij} - \beta^m_1 t_{ij} - \beta^m_2 d_{ij}\right)}_{\text{revenue for serving customers}} - \underbrace{\left(\beta^m_1 t_{ki} + \beta^m_2 d_{ki}\right)}_{\text{pickup costs}} \right] z^m_{kij}$$

A feasible service flow z^m needs to satisfy customers' demands. In other words, for each OD pair $(i, j) \in \mathcal{E}$ and each TNC $m \in \mathcal{M}$, there should be more ride-hailing vehicles available than customers who want to travel. This constraint can be written as in Eq. (1). We can observe that if no ride-hailing vehicles are dispatched to serve OD pair $(i,j) \in \mathcal{E}$ for TNC $m \in \mathcal{M}$, then no travelers will choose to use TNC m for that OD pair (i,j), namely $y_{ij}^m = 0$.

$$\sum_{k \in \mathcal{D}} z_{kij}^m \geq y_{ij}^m, \quad \forall\, (i,j) \in \mathcal{E} \tag{1}$$

With travelers' demand flow $\{\, y_{ik}^m : (i,k) \in \mathcal{E} \,\}$ known, a feasible service flow z^m has to respect the flow conservation constraints for each destination $k \in \mathcal{D}$ and each TNC $m \in \mathcal{M}$. That is, in a steady-state equilibrium, for each $m \in \mathcal{M}$, the total number of ride-hailing vehicles available in destination k should be equal to the total number of customers who arrive at destination k. The flow conservation constraints can be formulated as in Eq. (2).

$$\sum_{(i,j) \in \mathcal{E}} z_{kij}^m = \sum_{i \in \mathcal{O}} y_{ik}^m, \quad \forall\, k \in \mathcal{D} \tag{2}$$

Last but not least, the total number of ride-hailing vehicles in the system, including the ride-hailing vehicles en route to pick up the customers and the ride-hailing vehicles delivering the customers from their origins to their destinations, should not exceed the fleet size that each TNC $m \in \mathcal{M}$ can provide. Denote N^m as the fleet size of TNC $m \in \mathcal{M}$, we have

$$\underbrace{\sum_{(i,j) \in \mathcal{E}} \sum_{k \in \mathcal{D}} t_{ki}\, z_{kij}^m}_{\#\text{ vehicles to pick up customers}} + \underbrace{\sum_{(i,j) \in \mathcal{E}} t_{ij}\, y_{ij}^m}_{\#\text{ vehicles to deliver customers}} \leq N^m \tag{3}$$

The **TNC choice module** can be summarized as follows. For every $m \in \mathcal{M}$, and with given travelers' demand flow $\{\, y_{ij}^m : (i,j) \in \mathcal{E} \,\}$,

$$\begin{aligned}
\underset{z_{kij}^m \geq 0}{\textbf{maximize}} \quad & \sum_{(i,j) \in \mathcal{E}} \sum_{k \in \mathcal{D}} \left[\underbrace{(\, \underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} - \beta_1^m t_{ij} - \beta_2^m d_{ij} \,)}_{\text{revenue for serving customers}} - \underbrace{(\, \beta_1^m t_{ki} + \beta_2^m d_{ki} \,)}_{\text{pickup costs}} \right] z_{kij}^m \\
\textbf{subject to} \quad & \sum_{k \in \mathcal{D}} z_{kij}^m \geq y_{ij}^m, \quad \forall\, (i,j) \in \mathcal{E} \\
& \sum_{(i,j) \in \mathcal{E}} z_{kij}^m = \sum_{i \in \mathcal{O}} y_{ik}^m, \quad \forall\, k \in \mathcal{D} \\
\textbf{and} \quad & \sum_{(i,j) \in \mathcal{E}} \sum_{k \in \mathcal{D}} t_{ki}\, z_{kij}^m + \sum_{(i,j) \in \mathcal{E}} t_{ij}\, y_{ij}^m \leq N^m.
\end{aligned} \tag{4}$$

Note that the TNC choice module only applies for $m \in \mathcal{M}$ and not for solo driving $(m = 0)$ and public transit ($m = M+1$). Let λ_{ij}^m represent the nonnegative dual variable of the TNC supply constraints (1) for each OD pair $(i,j) \in \mathcal{E}$ and each TNC $m \in \mathcal{M}$, which can be explained as the compensation to ensure that supply will be larger than the demand in the ride-hailing market. Let μ_k^m be the un-signed dual variable with respect to the flow conservation equation (2) for each destination $k \in \mathcal{D}$ and each TNC $m \in \mathcal{M}$, which represents the compensation or surge price that balances the flow of customers and the flow of ride-hailing vehicles. Let s_m be the nonnegative dual variable of the fleet size constraint (3). Notate $R_{kij}^m \triangleq (\underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} - \beta_1^m t_{ij} - \beta_2^m d_{ij}) - (\beta_1^m t_{ki} + \beta_2^m d_{ki})$. Denote $\perp$ as the perpendicularity notation that represents the complementarity slackness between the slack of a constraint and its dual variable, namely $\mathrm{x} \perp \mathrm{y} \iff \mathrm{x}^T\mathrm{y} = 0$. Then the TNC choice module (4) can be reformulated as an equivalent complementarity problem as follows:

$$0 \le z_{kij}^m \perp -R_{kij}^m - \underbrace{(\lambda_{ij}^m + \mu_k^m)}_{\substack{\text{compensation}\\ \text{and surge price}}} + t_{ki}\zeta^m \ge 0, \qquad \forall k \in \mathcal{D}, \forall (i,j) \in \mathcal{E}, m \in \mathcal{M}$$

$$0 \le \underbrace{\lambda_{ij}^m}_{\text{compensation}} \perp \sum_{k \in \mathcal{D}} z_{kij}^m - y_{ij}^m \ge 0, \qquad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M}$$

$$\underbrace{\mu_k^m}_{\substack{\text{compensation}\\ \text{or surge price}}} \text{ free}, \quad \sum_{(i,j) \in \mathcal{E}} z_{kij}^m = \sum_{i \in \mathcal{O}} y_{ik}^m, \qquad \forall k \in \mathcal{D}, \forall m \in \mathcal{M}$$

$$0 \le \zeta^m \perp N^m - \sum_{(i,j) \in \mathcal{E}} \sum_{k \in \mathcal{D}} t_{ki}\, z_{kij}^m - \sum_{(i,j) \in \mathcal{E}} t_{ij}\, y_{ij}^m \ge 0, \qquad \forall m \in \mathcal{M}$$

6 Traveler Choice Module

The customer choice module outputs the demand flow, $y^m \triangleq \left(y_{ij}^m\right)_{(i,j)\in\mathcal{O}\times\mathcal{D}}$, over all travel modes, including solo driving, requesting service from a TNC, and taking public transit. In general, the disutility of a traveler choosing mode $m \in \mathcal{M}_+ \triangleq \mathcal{M} \cup \{0, M+1\}$ (including ride-hailing, solo driving, and public transit) contains two parts: disutility occurs during waiting for the service and during traveling. Let w_{ij}^m be the waiting cost of travelers choosing mode $m \in \mathcal{M}_+$ associated with OD pair $(i,j) \in \mathcal{E}$. For TNC $m \in \mathcal{M}$, the waiting cost w_{ij}^m is a variable depending on the service flow z^m and the demand flow y^m. This will be discussed in more detail in Section 7. For solo driving $m = 0$, we set $w_{ij}^0 = 0$ for all $(i,j) \in \mathcal{E}$. For public transit $m = M+1$, we assume w_{ij}^{M+1} are fixed constants for all $(i,j) \in \mathcal{E}$. The disutility due to waiting, together with the disutility during traveling, will be defined in Section 6.1 in more detail.

Denote V_{ij}^m as the disutility of travelers that need to travel across OD pair $(i,j) \in \mathcal{E}$ and choose mode $m \in \mathcal{M}_+$. For travelers choosing TNC $m \in \mathcal{M}$, the disutility can be written as follows:

$$V_{ij}^m \triangleq \underbrace{\alpha_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij}}_{\text{payment to ride-hailing drivers}} + \underbrace{\gamma_2^m t_{ij}}_{\text{time based disutility}} + \underbrace{w_{ij}^m}_{\text{waiting cost}} \tag{5}$$

For a solo driver, i.e., $m = 0$, the disutility can be represented as:

$$V_{ij}^0 \triangleq \underbrace{\gamma_2^0 t_{ij}}_{\text{time based disutility}} + \underbrace{\alpha_2^0 d_{ij}}_{\text{distance based disutility}} \tag{6}$$

The disutilities are expressed as monetary values. If a traveler uses TNC, the disutility in Eq. (5) contains the fare paid to the ride-hailing drivers, the travel time based cost, and the waiting cost that includes pickup and matching costs. If a traveler chooses to be a solo driver, the disutility in Eq. (6) contains both time-based and distance-based costs.

For travelers taking public transit $m = M+1$, the structure for their disutility, V_{ij}^{M+1}, is more complicated due to possible walking and switching between different bus or subway lines. Nevertheless, similar to the disutility of solo drivers (V_{ij}^0), V_{ij}^{M+1} is simply a constant, which can be computed beforehand. The details for computing V_{ij}^{M+1} is shown in Section 6.1.

Travelers are considered to be rational and will choose the mode (i.e., TNC, solo driving, or public transit) that minimizes their disutility. For each OD pair $(i,j) \in \mathcal{E}$, the summation of the demand flow, y^m, over all travel modes $m \in \mathcal{M}_+$ should be equal to the total demand Q_{ij}. The travelers are not concerned about the constraints related to the operations of TNCs, i.e., Eqs. (1), (2), and (3), and thus do not include these constraints in their optimization problem. The **traveler choice module** can be formulated as below. Given the waiting cost $\left\{ w_{ij}^m : (i,j) \in \mathcal{E} \right\}_{m \in \mathcal{M}_+}$, where $\mathcal{M}_+ \triangleq \mathcal{M} \cup \{0\} \cup \{M+1\}$,

$$\begin{aligned} \underset{y_{ij}^m \geq 0}{\textbf{minimize}} \quad & \underbrace{\sum_{m \in \mathcal{M}} \sum_{(i,j) \in \mathcal{E}} \left(\underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} + \gamma_2^m t_{ij} + \underbrace{w_{ij}^m}_{\text{waiting cost}} \right) y_{ij}^m}_{\text{ride-hailing costs}} \\ & + \underbrace{\sum_{(i,j) \in \mathcal{E}} \left(\alpha_2^0 d_{ij} + \gamma_2^0 t_{ij} \right) y_{ij}^0}_{\text{solo-driving costs}} + \underbrace{\sum_{(i,j) \in \mathcal{E}} V_{ij}^{M+1} y_{ij}^{M+1}}_{\text{public transit costs}} \\ \textbf{subject to} \quad & y_{ij}^0 + y_{ij}^{M+1} + \sum_{m \in \mathcal{M}} y_{ij}^m = Q_{ij}, \quad \forall (i,j) \in \mathcal{E}. \end{aligned} \tag{7}$$

Similar to the TNC choice module, which is separable with respect to each TNC $m \in \mathcal{M}$, the traveler choice module is separable in terms of each OD pair $(i,j) \in \mathcal{E}$. This will be more clear in the equivalent complementarity formulation below. Denote u_{ij} as the least disutility (i.e., the minimum disutility) for OD pair $(i,j) \in \mathcal{E}$ over all travel modes $m \in \mathcal{M}_+$. The optimality conditions of (7) are given by the following complementarity conditions:

$$\begin{aligned} & 0 \leq y_{ij}^m \perp \underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} + \gamma_2^m t_{ij} + \underbrace{w_{ij}^m}_{\text{waiting cost}} - \underbrace{u_{ij}}_{\text{least disutility}} \geq 0, \\ & \qquad \forall (i,j) \in \mathcal{E}, m \in \mathcal{M} \\ & 0 \leq y_{ij}^0 \perp \alpha_2^0 d_{ij} + \gamma_2^0 t_{ij} - \underbrace{u_{ij}}_{\text{least disutility}} \geq 0, \qquad \forall (i,j) \in \mathcal{E} \\ & 0 \leq y_{ij}^{M+1} \perp V_{ij}^{M+1} - \underbrace{u_{ij}}_{\text{least disutility}} \geq 0, \qquad \forall (i,j) \in \mathcal{E} \\ & \underbrace{u_{ij}}_{\text{least disutility}} \text{ free}, \quad y_{ij}^0 + y_{ij}^{M+1} + \sum_{m \in \mathcal{M}} y_{ij}^m = Q_{ij}, \qquad \forall (i,j) \in \mathcal{E} \end{aligned}$$

The complementarity formulation above has a similar interpretation to the classic Wardrop's user equilibrium principle, see, e.g., Section 1.4.5 of Facchinei and Pang (2003). In this context, it means that travelers across OD pair $(i,j) \in \mathcal{E}$ will choose a mode $m \in \mathcal{M}_+$ only if that mode has the minimum disutility among all travel modes along that OD pair. Note that, different from Ban et al. (2019), in this research, we avoid the shared constraints between the TNC choice module and the traveler choice module. As a result, we do not have to introduce the normalized relationship between the multipliers corresponding to the shared constraints.

6.1 Disutility of Public Transit Passengers

The public transit is a network consisting of a number of bus and subway lines. Let $\mathcal{M}_b$ be the collection of labels for all transit lines and ω_m be the headway of transit $m \in \mathcal{M}_b$. We assume that passengers arrive randomly at all public transit stops. Then the passenger waiting time at a stop of transit line $m \in \mathcal{M}_b$ is $\omega_m/2$. It is common in practice for a public transit passenger to switch between different bus or subway lines. Hence, the trip of a public transit passenger contains multiple line segments. The overall disutility is the sum of the disutility (including waiting and traveling) of all the line segments, plus also the disutility of possible walking before the first and after the last stops. For an OD pair (i, j), let S_{ij} be a set of public transit line segments chosen by a passenger, i.e., $S_{ij} = \{(k_1, k_2), (k_2, k_3), \ldots, (k_{\bar{\nu}-1}, k_{\bar{\nu}})\}$, and let $m_\nu \in \mathcal{M}_b$ be the line label of the νth segment $(k_\nu, k_{\nu+1})$. We note that the first transit stop k_1 (and the last stop, respectively) is not necessarily the same as the origin i (and the destination j, respectively), as passengers may need to walk to/from public transient stops in the beginning or at the end of the trip. However, we do require that intermediate segments of S_{ij} to be connected directly such that no walking is needed to switch lines. The overall disutility of a passenger associated with S_{ij} is computed as follows.

$$C(S_{ij}) \triangleq \beta_0 d_{ik_1} + \sum_{(k_\nu, k_{\nu+1}) \in S_{ij}} \left(\gamma_1^{m_\nu} \frac{\omega_{m_\nu}}{2} + \gamma_2^{m_\nu} t_{k_\nu k_{\nu+1}} + \underline{\alpha}_{k_\nu k_{\nu+1}}^{m_\nu} \right) + (|S_{ij}| - 1)\hat{\alpha} + \beta_0 d_{k_{\bar{\nu}} j} \tag{8}$$

where β_0 is the conversion factor from distance to disutility due to walking, $\gamma_1^{m_\nu}$ is the conversion factor from waiting time to disutility, $\gamma_2^{m_\nu}$ is the conversion factor from time to disutility due to traveling, $t_{k_\nu k_{\nu+1}}$ is the travel time from k_ν to $k_{\nu+1}$ taking line m_ν, $\underline{\alpha}_{k_\nu k_{\nu+1}}^{m_\nu}$ is the distance-based fare, and $\hat{\alpha}$ is the disutility of inconvenience caused by a single transferring between lines.

The disutility of travelers that go from origin $i \in \mathcal{O}$ to destination $j \in \mathcal{D}$, and choose public transit $m = M+1$, denoted as V_{ij}^{M+1}, is then defined as follows:

$$V_{ij}^{M+1} \triangleq \min_{S_{ij} \text{ associated with } (i,j)} C(S_{ij}) \tag{9}$$

The above problem is, in fact, a shortest path problem over a cost-weighted network $\hat{\mathcal{G}}$ constructed from the original transportation network $\mathcal{G}_0 = (\mathcal{N}, \mathcal{A})$ and public transit network. $\hat{\mathcal{G}}$ is the direct summation of $|\mathcal{M}_b| + 1$ subnetworks. Associated with every $m \in \mathcal{M}_b$ is a subnetwork $(\mathcal{N}_m, \mathcal{A}_m)$ corresponding to the transit line, where the weight for a link $(k, \ell) \in \mathcal{A}_m$ is $\gamma_2^m t_{kl} + \underline{\alpha}_{k\ell}^m$. The remaining subnetwork is to take into account the walking segments in (8). It has node set $\mathcal{O} \cup \mathcal{D} \cup (\cup_{m \in \mathcal{M}_b} \mathcal{N}_m)$ and between every node $k \in \mathcal{O} \cup \mathcal{D}$ and every transit stop $\ell \in \cup_{m \in \mathcal{M}_b} \mathcal{N}_m$ is a link with a weight: $\beta_0 d_{k\ell} + \gamma_1^m \omega_m / 2$ for $k \in \mathcal{O}$ and $\ell \in \mathcal{N}_m$, and $\beta_0 d_{k\ell}$ for $k \in \mathcal{D}$. In both cases, $d_{k\ell}$ is the shortest walking distance between k and ℓ over network $\mathcal{G}_0$. In the last subnetwork, a node $k \in \mathcal{O} \cup \mathcal{D}$ might be only connected to nearby public transit stops. In order to incorporate the cost of waiting at transit stops and switching transit lines, the same stop in the $|\mathcal{M}_b| + 1$ subnetworks are treated as different nodes in $\hat{\mathcal{G}}$. Links are added between two of these stops at the same location, with weight $\hat{\alpha} + \gamma_1^m \omega_m / 2$ for links to stops of transit line m and weight 0 to stops in the additional subnetwork for walking.

7 Customer Waiting Module

In this section, we provide the customer waiting module, which characterizes the waiting cost w_{ij}^m that appears in the traveler choice module (7). As previously discussed, our model assume that the waiting costs for solo driving and public transit, namely w_{ij}^0 and w_{ij}^{M+1}, are constants (possibly zero) for all $(i,j) \in \mathcal{E}$. We hence focus on the waiting cost for customers that choose each TNC $m \in \mathcal{M}$. For the purpose of waiting cost characterization, the TNCs' service flow $z^m \triangleq \left(z_{kij}^m \right)_{(k,i,j)\in\mathcal{D}\times\mathcal{O}\times\mathcal{D}}$ and the travelers' demand flow $y^m \triangleq \left(y_{ij}^m \right)_{(i,j)\in\mathcal{O}\times\mathcal{D}}$ are taken as given, and satisfy the conditions in the TNC choice module (4) and those in the traveler choice module (7), respectively. The waiting cost of ride-hailing customers depends on both the service flow z^m and the demand flow y^m. Denote $\mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}}, y_{ij}^m \right)$ as a set-valued mapping for every $m \in \mathcal{M}$ and every OD pair $(i,j) \in \mathcal{E}$. Then most generally, the waiting cost of ride-hailing customers can be modelled as

$$w_{ij}^m \in \mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}}, y_{ij}^m \right), \quad \forall (i,j) \in \mathcal{E}, m \in \mathcal{M} \tag{10}$$

In order to estimate the waiting cost of ride-hailing customers, we consider the ride-hailing system as a queueing system: the ride-hailing customers with their origins and destinations arrive in the system, waiting for the ride-hailing drivers to serve them. We consider each OD pair $(i,j) \in \mathcal{E}$ of each TNC $m \in \mathcal{M}$ as a queue. The interarrival times of ride-hailing customers' arrival follow a distribution with mean $1/y_{ij}^m$, while the time intervals of ride-hailing vehicles' service follow a distribution with mean $1/s_{ij}^m$ where $s_{ij}^m = \sum_{k\in\mathcal{D}} z_{kij}^m$. In other words, the mean arrival rate of each queue is y_{ij}^m, and the mean service rate of each queue is s_{ij}^m. For each queueing system associated with OD pair $(i,j) \in \mathcal{E}$ and TNC $m \in \mathcal{M}$, the dependence of mean customer waiting cost w_{ij} on service flow and demand flow occurs on two components: (A) mean *matching cost*: this cost is based on the cost of waiting for matching, i.e., the average cost between customer arrival and the matched ride-hailing vehicles setting out at destinations for picking up. We note that the term *matching cost* is also used in the literature, see, e.g., Zha et al. (2018). To determine the *matching cost*, we model each OD pair as a single-server queue. (B) mean *pickup cost*: this cost is based on the cost of waiting for the vehicles to pick up the customers after the matching is completed, namely the average travel cost of ride-hailing vehicles to come from their previous locations to the customers' origins. To model the *pickup cost*, we use a similar approach as in Ban et al. (2019), Di and Ban (2019), Xu et al. (2021), and Chen and Di (2023).

The service for customers starts at the time when he or she asks for the ride-hailing service and ends at the time when a driver arrives to pick up the customer. Including both the matching and the pickup costs, the waiting cost depends on the service policy chosen by a TNC. Depending on the complexity of the service policy, the queueing model could become complicated. In particular, a desirable service policy for a TNC could often be state-dependent. The design of optimal state-dependent service policy itself is a challenging problem to study. One such example is Banerjee et al. (2022), where the pickup and delivery of customers are assumed to be instantaneous in order to produce meaningful results. In this study, we will derive approximations of customer waiting costs under realistic assumptions.

The waiting cost of ride-hailing customers is made up first of the *matching cost*, which is a function of the time it takes to match ride-hailing customers with ride-hailing drivers. The mean *matching cost* of ride-hailing customers depends on both the TNCs' service flow z^m and the travelers' demand flow y^m. Let γ_1^m be the value of time for customers of TNC $m \in \mathcal{M}$ while waiting for a vehicle. For each queueing system associated with OD pair $(i, j) \in \mathcal{E}$ and TNC $m \in \mathcal{M}$, the mean arrival rate of ride-hailing customers is y_{ij}^m, while the mean service rate of ride-hailing drivers is $s_{ij}^m = \sum_{k \in \mathcal{D}} z_{kij}^m$. With Markovian property, and further assuming that the customers' arrivals are determined by a Poisson process while ride-hailing vehicles' service times follow an exponential distribution (M/M/1), the steady-state mean *matching cost* can be generated using a closed-form solution with respect to y_{ij}^m and s_{ij}^m as follows:

$$\gamma_1^m \frac{y_{ij}^m}{\sum_{k \in \mathcal{D}} z_{kij}^m \left(\sum_{k \in \mathcal{D}} z_{kij}^m - y_{ij}^m \right)}, \quad \forall (i, j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{11}$$

The second part of the waiting cost of ride-hailing customers is based on the time it takes the ride-hailing vehicle to drive to the customer's origin for picking up. In this study, we derive this *pickup cost* using a similar approach as Ban et al. (2019), Di and Ban (2019), Xu et al. (2021), and Chen and Di (2023). That is, the mean *pickup cost* of ride-hailing customers is related to the TNCs' service flow z^m. Specifically, for each OD pair, it can be computed as the average of the travel costs of the ride-hailing vehicles from all possible locations to the origin for picking up the customers. Then for OD pair $(i, j) \in \mathcal{E}$ and TNC $m \in \mathcal{M}$ and with given TNC's service flow $\left\{ z_{kij}^m : (k, i, j) \in \mathcal{D} \times \mathcal{O} \times \mathcal{D} \right\}$, the mean *pickup cost* is formulated as follows:

$$\gamma_1^m \sum_{k \in \mathcal{D}} t_{ki} \left(\frac{z_{kij}^m}{\sum_{k' \in \mathcal{D}} z_{k'ij}^m} \right), \quad \forall i \in \mathcal{O}, \forall m \in \mathcal{M} \tag{12}$$

Combining the mean *matching cost* (11) and the mean *pickup cost* (12), for OD pair $(i, j) \in \mathcal{E}$ and TNC

$m \in \mathcal{M}$, we obtain the mean waiting cost of ride-hailing customers as below:

$$w_{ij}^m = \gamma_1^m \underbrace{\left[\underbrace{\sum_{k \in \mathcal{D}} t_{ki} \left(\frac{z_{kij}^m}{\sum_{k' \in \mathcal{D}} z_{k'ij}^m} \right)}_{\text{pickup cost}} + \underbrace{\frac{y_{ij}^m}{\sum_{k \in \mathcal{D}} z_{kij}^m \left(\sum_{k \in \mathcal{D}} z_{kij}^m - y_{ij}^m \right)}}_{\text{matching cost}} \right]}_{\text{waiting cost}}, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{13}$$

Similarly, under different assumptions of the distributions with respect to customers' arrival times and ride-hailing vehicles' service times, different types of steady-state mean waiting costs of ride-hailing customers can be formulated as follows:

- M/D/1 queue:

$$w_{ij}^m = \gamma_1^m \underbrace{\left[\underbrace{\sum_{k \in \mathcal{D}} t_{ki} \left(\frac{z_{kij}^m}{\sum_{k' \in \mathcal{D}} z_{k'ij}^m} \right)}_{\text{pickup cost}} + \underbrace{\frac{y_{ij}^m}{2 \sum_{k \in \mathcal{D}} z_{kij}^m \left(\sum_{k \in \mathcal{D}} z_{kij}^m - y_{ij}^m \right)}}_{\text{matching cost}} \right]}_{\text{waiting cost}}, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{14}$$

- M/G/1 queue (using Pollaczek-Khintchine formula):

$$w_{ij}^m = \gamma_1^m \underbrace{\left[\underbrace{\sum_{k \in \mathcal{D}} t_{ki} \left(\frac{z_{kij}^m}{\sum_{k' \in \mathcal{D}} z_{k'ij}^m} \right)}_{\text{pickup cost}} + \underbrace{\left[\frac{y_{ij}^m}{\sum_{k \in \mathcal{D}} z_{kij}^m \left(\sum_{k \in \mathcal{D}} z_{kij}^m - y_{ij}^m \right)} \right] \left(1 + \frac{\sigma_s^2}{\tau_s^2} \right)}_{\text{matching cost}} \right]}_{\text{waiting cost}}, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{15}$$

where σ_s^2 represents the variance of service time and τ_s is the mean of the service time, together the ratio $\frac{\sigma_s^2}{\tau_s^2}$ represents the efficiency of service time in statistics. According to the Cramér–Rao bound $\frac{\sigma_s^2}{\tau_s^2} \leq 1$.

- G/G/1 queue (using Kingman's formula):

$$w_{ij}^m = \gamma_1^m \underbrace{\left[\underbrace{\sum_{k\in\mathcal{D}} t_{ki} \left(\frac{z_{kij}^m}{\sum_{k'\in\mathcal{D}} z_{k'ij}^m} \right)}_{\text{pickup cost}} + \underbrace{\left[\frac{y_{ij}^m}{\sum_{k\in\mathcal{D}} z_{kij}^m \left(\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m \right)} \right] \left(\frac{\sigma_y^2}{\tau_y^2} + \frac{\sigma_s^2}{\tau_s^2} \right)}_{\text{matching cost}} \right]}_{\text{waiting cost}}, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{16}$$

where $\frac{\sigma_y^2}{\tau_y^2}$ and $\frac{\sigma_s^2}{\tau_s^2}$ represents the efficiency of arrival time and service time, respectively. According to the Cramér–Rao bound $\frac{\sigma_y^2}{\tau_y^2} \leq 1$ and $\frac{\sigma_s^2}{\tau_s^2} \leq 1$.

The closed-form formulas for the waiting cost of ride-hailing customers, Eqs. (13), (14), (15), (16), could cause some computational issues since the denominators, $\sum_{k'\in\mathcal{D}} z_{k'ij}^m$ and $\sum_{k\in\mathcal{D}} z_{kij}^m \left(\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m \right)$, may be zeros. We propose two approaches to overcome this challenge.

The first approach is to modify Eq. (1) slightly as follows:

$$\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m \geq \epsilon, \quad \forall\, (i,j) \in \mathcal{E}, \forall\, m \in \mathcal{M}$$

where ϵ is a small positive number. With this remedy, we can ensure that denominators in the closed-form formulas of the waiting costs, Eqs. (13), (14), (15), (16), will be positive.

The second approach is to reformulate the ride-hailing customers' waiting cost functions, Eqs. (13), (14), (15), (16), as complementarity problems. We implement the reformulation using the M/M/1 queue (13) as an example. For OD pair $(i,j) \in \mathcal{E}$, denote $w_{ij}^{\mathrm{m},m}$ as the mean *matching cost* of customers to be serviced if they choose TNC $m \in \mathcal{M}$, and denote $w_{ij}^{\max}$ as the upper bound of the waiting cost strictly larger than $w_{ij}^{\mathrm{m},m}$. How $w_{ij}^{\max}$ is quantified will be discussed in Section 8. Introduce η_{ij}^m as the multipliers, then the mean *matching cost* in Eq. (13) can be reformulated as complementarity conditions below:

$$\begin{aligned} 0 \leq w_{ij}^{\mathrm{m},m} \quad &\perp \quad \left[\sum_{k\in\mathcal{D}} z_{kij}^m \left(\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m \right) \right] w_{ij}^{\mathrm{m},m} - \gamma_1^m y_{ij}^m + \eta_{ij}^m \geq 0, \qquad \forall (i,j) \in \mathcal{E}, \forall\, m \in \mathcal{M} \\ 0 \leq \eta_{ij}^m \quad &\perp \quad w_{ij}^{\max} - w_{ij}^{\mathrm{m},m} \geq 0, \qquad \forall (i,j) \in \mathcal{E}, \forall\, m \in \mathcal{M} \end{aligned} \tag{17}$$

There are two possible situations for Eq. (17):

- If $0<w_{ij}^{\mathrm{m},m}<w_{ij}^{\max}$, then $\eta_{ij}^m = 0$ and

$$w_{ij}^{\mathrm{m},m} = \gamma_1^m \frac{y_{ij}^m}{\sum_{k\in\mathcal{D}} z_{kij}^m \left(\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m\right)}$$

- If $w_{ij}^{\mathrm{m},m} = 0$, then $\eta_{ij}^m = y_{ij}^m = 0$.

Let θ_{kij}^m represent the proportion of drivers at destination $k \in \mathcal{D}$ that are matched to customers from origin $i \in \mathcal{O}$ to destination $j \in \mathcal{D}$ for TNC $m \in \mathcal{M}$. Introduce θ_{ij}^m as the multipliers, then the proportion, θ_{kij}^m, can be reformulated as complementarity conditions as follows:

$$\begin{aligned}
&0 \le \theta_{kij}^m \quad \perp \quad \left[\sum_{k'\in\mathcal{D}} z_{k'ij}^m\right]\theta_{kij}^m - \gamma_1^m z_{kij}^m + \psi_{kij}^m \ge 0, && \forall k \in \mathcal{D}, \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \\
&0 \le \psi_{kij}^m \quad \perp \quad 1 - \theta_{kij}^m \ge 0, && \forall k \in \mathcal{D}, \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M}
\end{aligned} \tag{18}$$

Furthermore, the mean *pickup cost* in Eq. (13) can be computed as below:

$$w_{ij}^{\mathrm{p},m} = \gamma_1^m \sum_{k\in\mathcal{D}} \left(t_{ki}\theta_{kij}^m\right), \quad \forall i \in \mathcal{O}, \forall m \in \mathcal{M} \tag{19}$$

Combine the mean *pickup cost* in Eq. (19) and the mean *matching cost* in Eq.(17), together with the upper bound of the waiting cost $w_{ij}^{\max}$, the mean waiting cost of ride-hailing customers for OD pair $(i,j) \in \mathcal{E}$ and TNC $m \in \mathcal{M}$ is as follows:

$$w_{ij}^m = \min\left(w_{ij}^{\max}, \underbrace{\underbrace{w_{ij}^{\mathrm{p},m}}_{\text{pickup cost}} + \underbrace{w_{ij}^{\mathrm{m},m}}_{\text{matching cost}}}_{\text{waiting cost}}\right), \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{20}$$

Eq. (20) above is equivalent to the complementarity formulation below:

$$0 \le w_{ij}^{\max} - w_{ij}^m \perp w_{ij}^{\mathrm{p},m} + w_{ij}^{\mathrm{m},m} - w_{ij}^m \ge 0, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M} \tag{21}$$

8 The Overall Equilibrium Model and Analysis

The Equilibrium Model with given OD trip times:

Assumption 1: The OD trip times $\{ t_{ij} : (i,j) \in \mathcal{E} \}$ are known.

Under Assumption 1, there is no need to model traffic congestion. The model consists of three modules: TNC choice, traveler choice, and customer waiting. Mathematically, the model is a *generalized noncooperative game* with a system clearance mechanism that dictates the calculation of the ride-hailing customers' waiting costs, which is exogenous to the model's primary decision makers, which are the TNC companies and trip takers.

TNC Choice Module

Decomposed in each ride-hailing mode $m \in \mathcal{M}$, this is a linear program in the service variable $z^m \triangleq \left\{ z^m_{kij} : (k,i,j) \in \mathcal{D} \times \mathcal{O} \times \mathcal{D} \right\}$, for given $\{ y^m_{ij} : (i,j) \in \mathcal{E} \}$:

$$\underset{z^m_{kij} \geq 0}{\textbf{maximize}} \quad \sum_{(i,j)\in\mathcal{E}} \sum_{k\in\mathcal{D}} \left[\underbrace{(\underline{\alpha}^m_i + \alpha^m_1 t_{ij} + \alpha^m_2 d_{ij} - \beta^m_1 t_{ij} - \beta^m_2 d_{ij})}_{\text{revenue for serving customers}} - \underbrace{(\beta^m_1 t_{ki} + \beta^m_2 d_{ki})}_{\text{pickup costs}} \right] z^m_{kij}$$

$$\textbf{subject to} \quad \sum_{k\in\mathcal{D}} z^m_{kij} \geq y^m_{ij}, \quad \forall\, (i,j) \in \mathcal{E}$$

$$\sum_{(i,j)\in\mathcal{E}} z^m_{kij} = \sum_{i\in\mathcal{O}} y^m_{ik}, \quad \forall\, k \in \mathcal{D}$$

$$\textbf{and} \quad \sum_{(i,j)\in\mathcal{E}} \sum_{k\in\mathcal{D}} t_{ki}\, z^m_{kij} + \sum_{(i,j)\in\mathcal{E}} t_{ij}\, y^m_{ij} \leq N^m.$$

Note: Without the last constraint, the first two constraints are standard transportation flow constraints. Thus, the optimization problem is similar to the transportation kind. The last constraint stipulates that the number of available TNC vehicles must be sufficient to meet the various OD trip demands.

Traveler choice module

Aggregating all the trip modes, both solo driving and ride-hailing, this module is a linear program in the traveler choice variables $\left\{ y^m \triangleq y_{ij}^m : (i,j) \in \mathcal{E} \right\}_{m \in \mathcal{M}_+}$, for given $\left\{ w_{ij}^m : (i,j) \in \mathcal{E} \right\}_{m \in \mathcal{M}_+}$, where $\mathcal{M}_+ \triangleq \mathcal{M} \cup \{0\} \cup \{M+1\}$:

$$\begin{aligned} \underset{y_{ij}^m \geq 0}{\textbf{minimize}} \quad & \underbrace{\sum_{m \in \mathcal{M}} \sum_{(i,j) \in \mathcal{E}} \left(\underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} + \gamma_2^m t_{ij} + \underbrace{w_{ij}^m}_{\text{waiting cost}} \right) y_{ij}^m}_{\text{ride-hailing costs}} \\ & + \underbrace{\sum_{(i,j) \in \mathcal{E}} \left(\alpha_2^0 d_{ij} + \gamma_2^0 t_{ij} \right) y_{ij}^0}_{\text{solo-driving costs}} + \underbrace{\sum_{(i,j) \in \mathcal{E}} V_{ij}^{M+1} y_{ij}^{M+1}}_{\text{public transit costs}} \\ \textbf{subject to} \quad & y_{ij}^0 + y_{ij}^{M+1} + \sum_{m \in \mathcal{M}} y_{ij}^m = Q_{ij}, \quad \forall (i,j) \in \mathcal{E}. \end{aligned} \tag{22}$$

Note: offering flexibility, the waiting costs w_{ij}^m are modeled exogenously in a separate module that is an integral part of the overall equilibrium model. These waiting costs are the only model variables exogenous to (22) that have an impact on this module. The following lemma provides the background for the queueing-based formula to be presented later; roughly speaking, it states that if the waiting cost of a travel mode exceeds a threshold, then a traveler would choose to be a solo driver rather than hail a ride; the result quantifies the threshold.

Lemma 1. If for some pair $(i,j) \in \mathcal{E}$ and some $m \in \mathcal{M}$,

$$w_{ij}^m > \alpha_2^0 d_{ij} + \gamma_2^0 t_{ij} - \left(\underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} + \gamma_2^m t_{ij} \right)$$

then the traveler choice problem (22) has an optimal solution with $y_{ij}^m = 0$.

Proof. This is obvious because we can shift y_{ij}^m to y_{ij}^0, decreasing the former to zero and increasing the latter by the shifted amount. This adjustment will strictly decrease objective value without affecting the constraints. □

Assumption 2: For every tuple $\left\{ y^m \triangleq \{y_{ij}^m\}_{(i,j) \in \mathcal{E}} \right\}_{m \in \mathcal{M}_+}$ feasible to the traveler choice module, the TNC choice constraints are feasible.

The waiting cost map $\mathcal{Q}^m_{ij}$: For every $m \in \mathcal{M}$ and every OD pair $(i, j) \in \mathcal{E}$, the customer waiting can be modeled most generally by a set-valued map: $\mathcal{Q}^m_{ij}\left(\{z^m_{kij}\}_{k\in\mathcal{D}},\, y^m_{ij} \right)$, whose dependence on its arguments is through three separate quantities: the travel times to pickup locations $\sum_{k\in\mathcal{D}} t_{ki} \left[\frac{z^m_{kij}}{\sum_{k'\in\mathcal{D}} z^m_{k'ij}} \right]$, the OD trip demands y^m_{ij}, and the served trips $\sum_{k\in\mathcal{D}} z^m_{kij}$; thus in general, $w^m_{ij} \in \mathcal{Q}^m_{ij}\left(\{z^m_{kij}\}_{k\in\mathcal{D}},\, y^m_{ij} \right)$. In what follows, we describe one particular form of the map $\mathcal{Q}^m_{ij}$ that is derived by considering two main components of waiting:

(A) Travel times to pick-up locations: For every $m \in \mathcal{M}$, the quantity $\sum_{k\in\mathcal{D}} t_{ki} \left[\frac{z^m_{kij}}{\sum_{k'\in\mathcal{D}} z^m_{k'ij}} \right]$ gives the times for the TNC drivers to go from the last dropoff (at a destination) to a new pickup (at a new origin); the fractions $\left\{ \frac{z^m_{kij}}{\sum_{k'\in\mathcal{D}} z^m_{k'ij}} \right\}_{k\in\mathcal{D}}$ whose sum over $k \in \mathcal{D}$ is equal to unity, can be interpreted as the probability of a pickup of OD trip (i, j) by a TNC car coming from the (prior) destination k. Notice that if y^m_{ij} is positive, then the denominator of this ratio is positive, ensuring the well-definedness of this ratio; however, if $y^m_{ij} = 0$, then this ratio is ambiguous; this ambiguity, however, is immaterial because the product, $w^m_{ij}\, y^m_{ij}$, which is the only place where w^m_{ij} appears, is equal to zero no matter what value w^m_{ij} takes; this observation will be reflected in the definition of the waiting-cost map $\mathcal{Q}^m_{ij}$.

(B) Queueing-based waiting times: For every $m \in \mathcal{M}$ and each OD pair $(i, j) \in \mathcal{E}$, a queue is formed with the arrival rate of y^m_{ij} and service rate $s^m_{ij} = \sum_{k\in\mathcal{D}} z^m_{kij}$; these rates will influence the waiting time in the queueing system of the ride-hailing customers. Various queueing models will yield a steady-state queue length determined by a positive multiple of the fraction:

$$\rho_{\text{q}} \left[\frac{y^m_{ij}}{\sum_{k\in\mathcal{D}} z^m_{kij} \left(\sum_{k\in\mathcal{D}} z^m_{kij} - y^m_{ij} \right)} \right] = \rho_{\text{q}} \left[\frac{1}{\sum_{k\in\mathcal{D}} z^m_{kij} - y^m_{ij}} - \frac{1}{\sum_{k\in\mathcal{D}} z^m_{kij}} \right], \quad \text{provided that } s^m_{ij} \geq y^m_{ij},$$

where ρ_{q} is a positive constant depending on the queue model being used (M/M/1, M/D/1, M/G/1, G/G/1, etc.). Similar to the previous pickup travel times, the above formula could be problematic when $y^m_{ij} = 0$; in this case, the same observation as before applies and the ambiguity of the formula is immaterial. If $y^m_{ij} > 0$ and yet $y^m_{ij} = s^m_{ij}$, the above fraction is equal to ∞, which leads to an infinite queue length. Here, Lemma 1 comes to the rescue and motivates a truncation of the formula.

With the above explanation, letting

$$w_{ij}^{\max} \triangleq \underbrace{\alpha_2^0 d_{ij} + \gamma_2^0 t_{ij} - \min_{m \in \mathcal{M}} \left(\underline{\alpha}_i^m + \alpha_1^m t_{ij} + \alpha_2^m d_{ij} + \gamma_2^m t_{ij} \right)}_{\text{may assume positive without loss of generality}},$$

we map now formally define the special waiting-cost map whose domain is the set of nonnegative pairs $\left(\{z_{kij}^m\}_{k\in\mathcal{D}}, y_{ij}^m \right)$ satisfying $\sum_{k\in\mathcal{D}} z_{kij}^m \geq y_{ij}^m$ (cf. the constraints in (4)):

- if $y_{ij}^m > 0$, $\mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m \right)$ is the singleton

$$\left\{ \underbrace{\min \left(w_{ij}^{\max},\ \gamma_1^m \left[\sum_{k\in\mathcal{D}} t_{ki} \left[\frac{z_{kij}^m}{\sum_{k'\in\mathcal{D}} z_{k'ij}} \right] + \rho_{\rm q} \underbrace{\left[\frac{y_{ij}^m}{\sum_{k\in\mathcal{D}} z_{kij}^m \left(\sum_{k\in\mathcal{D}} z_{kij}^m - y_{ij}^m \right)} \right]}_{\text{possibly equal to } \infty} \right] \right)}_{\text{single element, denoted } q_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m \right)} \right\};$$

- if $y_{ij}^m = 0$, $\mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m \right)$ is the interval $\left[0,\, w_{ij}^{\max} \right]$.

Below, we derive some important mathematical properties of the map $\mathcal{Q}_{ij}^m(\bullet, \bullet)$. First, its domain is a closed set; moreover $\mathcal{Q}_{ij}^m(\bullet, \bullet)$ is nonempty-valued, convex-valued, and compact-valued. We next show that it is a closed, and thus upper semicontinuous set-valued map. For this purpose, let

$$\left\{ \{z_{kij}^{\nu,m}\}_{k\in\mathcal{D}},\, y_{ij}^{\nu,m} \right\}_{\nu=1}^{\infty}$$

be a sequence in the domain of $\mathcal{Q}_{ij}^m$ converging to $\left\{ \{z_{kij}^{\infty,m}\}_{k\in\mathcal{D}},\, y_{ij}^{\infty,m} \right\}$, which be in the domain of $\mathcal{Q}_{ij}^m$; and let $\{w_{ij}^{\nu,m}\}_{\nu=1}^{\infty}$ be a sequence converging to $w_{ij}^{\infty,m}$ such that $w_{ij}^{\nu,m} \in \mathcal{Q}_{ij}^m \left(\{z_{kij}^{\nu,m}\}_{k\in\mathcal{D}},\, y_{ij}^{\nu,m} \right)$ for all ν. We need to show that: $w_{ij}^{\infty,m} \in \mathcal{Q}_{ij}^m \left(\{z_{kij}^{\infty,m}\}_{k\in\mathcal{D}},\, y_{ij}^{\infty,m} \right)$. There are several cases to consider:

- If $y_{ij}^{\infty,m} > 0$ and $\sum_{k\in\mathcal{D}} z_{kij}^{\infty,m} > y_{ij}^{\infty,m}$: in this case, we have $y_{ij}^{\nu,m} > 0$ and $\sum_{k\in\mathcal{D}} z_{kij}^{\nu,m} > y_{ij}^{\nu,m}$ for all ν sufficiently large; the claim is obvious by continuity of the pointwise minimum function.

- If $y_{ij}^{\infty,m} > 0$ and $\sum_{k\in\mathcal{D}} z_{kij}^{\infty,m} = y_{ij}^{\infty,m}$: in this case, we have $y_{ij}^{\nu,m} > 0$ for all ν sufficiently large and

$\sum_{k\in\mathcal{D}} z_{kij}^{\nu,m} \geq y_{ij}^{\nu,m}$ for all ν. If the latter inequality holds an equality for infinitely many ν's, then $w_{ij}^{\nu,m} = w_{ij}^{\max}$ for all such ν's; hence $w_{ij}^{\infty,m} = w_{ij}^{\max}$ and the claim holds readily. If $\sum_{k\in\mathcal{D}} z_{kij}^{\nu,m} > y_{ij}^{\nu,m}$ for all but finitely many ν's, then we have $w_{ij}^{\infty,m} = \lim_{\nu\to\infty} w_{ij}^{\nu,m} = w_{ij}^{\max}$ because

$$\lim_{\nu\to\infty} \frac{1}{\sum_{k\in\mathcal{D}} z_{kij}^{\nu,m} - y_{ij}^{\nu,m}} = \infty.$$

Thus in all these subcases, $w_{ij}^{\infty,m} \in \mathcal{Q}_{ij}^m\left(\{z_{kij}^{\infty,m}\}_{k\in\mathcal{D}},\, y_{ij}^{\infty,m}\right)$.

- If $y_{ij}^{\infty,m} = 0$, then since $w_{ij}^{\nu,m} \in [\,0,\, w_{ij}^{\max}\,] = \mathcal{Q}_{ij}^{\infty,m}$, it follows that $w_{ij}^{\infty,m} = \lim_{\nu\to\infty} w_{ij}^{\nu,m}$ also belongs to this interval.

We have now completed the proof of the desired properties of the waiting-costs map $\mathcal{Q}_{ij}^m$. It turns out that instead of separating the definition of this map into two cases depending on whether y_{ij}^m is zero or positive, we can very succinctly express this map as:

$$\mathcal{Q}_{ij}^m\left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m\right) = \underset{w_{ij}^m\in[\,0,w_{ij}^{\max}\,]}{\textbf{argmin}}\ \frac{y_{ij}^m}{2}\,(w_{ij}^m)^2 - \left[\,y_{ij}^m\, q_{ij}^m\left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m\right)\right] w_{ij}^m.$$

One advantage of this minimization definition of the waiting costs is that by a well-known regularization, we may obtain a unique waiting cost if desired. Specifically, let $\varepsilon_i \geq 0$ for $i = 1, 2$ be two arbitrary scalars, and let

$$\begin{aligned}
&\mathcal{Q}_{ij}^m\left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m;\, \varepsilon\right) \\
&\triangleq \underset{w_{ij}^m\in[\,0,w_{ij}^{\max}\,]}{\textbf{argmin}}\ \frac{y_{ij}^m+\varepsilon_1}{2}\,(w_{ij}^m)^2 - \left[\,(y_{ij}^m+\varepsilon_2)\, q_{ij}^m\left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m\right)\right] w_{ij}^m \\
&= \min\left\{ w_{ij}^{\max},\ \frac{y_{ij}^m+\varepsilon_2}{y_{ij}^m+\varepsilon_1}\, q_{ij}^m\left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m\right)\right\}
\end{aligned} \tag{23}$$

which is uniquely well-defined if the pair $(\varepsilon_1, \varepsilon_2)$ is nonzero. We summarize the above discussion of the waiting costs in the following module.

A (special) customer waiting module

Let $\varepsilon_i \geq 0$ for $i = 1, 2$ be arbitrary. For every $m \in \mathcal{M}$, every OD pair $(i,j) \in \mathcal{E}$, and every nonnegative pair $\left(\{z_{kij}^m\}_{k\in\mathcal{D}}, y_{ij}^m\right)$ satisfying $\sum_{k\in\mathcal{D}} z_{kij}^m \geq y_{ij}^m$, let w_{ij}^m be given by (23).

In the following result, the waiting costs map $\mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m \right)$ is left unspecified but required to possess the desired properties as the special map $\mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m;\, \varepsilon \right)$.

Theorem 2. In addition to Assumptions 1 and 2, assume the set-valued maps $\mathcal{Q}_{ij}^m$ are upper semicontinuous, closed-valued, and convex-valued on their respective domains; assume also that

$$\emptyset \neq \mathcal{Q}_{ij}^m \left(\{z_{kij}^m\}_{k\in\mathcal{D}},\, y_{ij}^m \right) \subseteq \left[\, 0, w_{ij}^{\max} \,\right]$$

for all pairs $\left(\{z_{kij}^m\}_{k\in\mathcal{D}}, y_{ij}^m \right)$ in the domain of $\mathcal{Q}_{ij}^m$. Then there exist tuples

$$\left\{ \left\{ \bar{y}_{ij}^m : (i,j) \in \mathcal{E} \right\}, \left\{ \bar{z}_{kij}^m : (k,i,j) \in \mathcal{D} \times \mathcal{O} \times \mathcal{D} \right\} \right\}_{m\in\mathcal{M}}, \left\{ \bar{y}_{ij}^0 : (i,j) \in \mathcal{E} \right\}, \left\{ \bar{y}_{ij}^{M+1} : (i,j) \in \mathcal{E} \right\}$$

and $\{\bar{w}_{ij}^m : (i,j) \in \mathcal{E}\}_{m\in\mathcal{M}}$ such that $\bar{w}_{ij}^m \in \mathcal{Q}_{ij}^m \left(\{\bar{z}_{kij}^m\}_{k\in\mathcal{D}},\, \bar{y}_{ij}^m \right)$ for all pairs $(i,j) \in \mathcal{E}$ and TNC modes $m \in \mathcal{M}$ and the y-tuple and z-tuple are optimal solutions of the TNC choice module, the traveler choice module, and the TNC driver-customer matching module, respectively. □

9 Numerical Results

In this section, we validate the proposed model using the Sioux-Falls network. In order to numerically solve the model, we put together the TNC choice module in Section 5, traveler choice module in Section 6, and customer waiting module in Section 7, and reformulate the overall equilibrium as an equivalent mixed complementarity problem. Take the M/M/1 queue version of the customer waiting module, Eq. (13), as an example, the overall equivalent mixed complementarity formulation can be found in Appendix 1. The equivalent mixed complementarity problem is then solved using Knitro (Byrd et al. 2006) on the NEOS server.

The settings of the Sioux-Falls network, including the geometry and the parameters, are based on Stabler (2020), which is widely used in the literature. We select three nodes (1, 2, 4) as origins and two other nodes (13, 19) as destinations; thus, there are six OD pairs in total. In the experiments, the parameters related to TNC fee charging and traveler disutility are modified based on Ban et al. (2019) and summarized in Table 2. The problems in this section can be solved within seconds.

Table 2. Parameters of TNC fee charging and traveler disutility.

Parameters	Notation	Value
Fleet size of TNC	N^m $(m = 1, 2)$	40000, 40000
Fixed fare charge ($)	$\underline{\alpha}^m$ $(m = 1, 2)$	3, 2
Time-based fare rate ($/hr)	α_1^m $(m = 1, 2)$	15, 10
Distance-based fare rate ($/mile)	α_2^m $(m = 0, 1, 2)$	2, 2.25, 2
Conversion factor from time to cost ($/hr)	β_1^m $(m = 1, 2)$	2, 2
Conversion factor from distance to cost ($/mile)	β_2^m $(m = 1, 2)$	0.5, 2
Value of time for ride-hailing travelers while waiting ($/hr)	γ_1^m $(m = 1, 2)$	1, 0.5
Value of time for ride-hailing travelers while traveling ($/hr)	γ_2^m $(m = 0, 1, 2)$	50, 3, 15

We compare the proposed method with the two models below. The main difference between the models is in how to formulate the waiting cost of ride-hailing customers.

- Ban et al. (2019): Instead of using the queueing-based waiting cost functions, e.g., Eqs. (13), (14), (15), (16), Ban et al. (2019) estimates the mean *matching cost* component of ride-hailing customers' waiting cost using the multiplier (dual variable) with respect to Eq. (1).

Table 3. Comparison between different models.

	VMT	Deadheading Miles	Avg. Waiting Cost	% of Ride-hailing
Ban et al. (2019)	54000	25200	65	100%
Linear Waiting Cost	46390	17590	573	65%
Our Model	32963	4163	817	15%

Table 4. Results of the Sioux-Fall network for the proposed model.

OD pair	Mode Choice	Customer Disutility	Mode-specific Cost	Waiting Cost
1 → 13	Solo Driving: 450	649	649	/
	TNC I: 0	694	61	633
	TNC II: 0	679	189	490
	Transit: 50	649	649	/
1 → 19	Solo Driving: 131	1298	1298	/
	TNC I: 0	1360	119	1241
	TNC II: 154	1298	376	922
	Transit: 15	1298	1298	/
2 → 13	Solo Driving: 137	1003	1003	/
	TNC I: 0	1047	92	955
	TNC II: 154	1003	291	712
	Transit: 9	1003	1003	/
2 → 19	Solo Driving: 80	944	944	/
	TNC I: 0	1007	87	920
	TNC II: 0	986	274	712
	Transit: 20	944	944	/
4 → 13	Solo Driving: 547	649	649	/
	TNC I: 0	694	61	633
	TNC II: 0	679	189	490
	Transit: 53	649	649	/
4 → 19	Solo Driving: 177	1003	1003	/
	TNC I: 0	1066	93	973
	TNC II: 0	1045	291	754
	Transit: 23	1003	1003	/

- A model with linear waiting cost function: Notice the queueing-based waiting cost functions, e.g., Eqs. (13), (14), (15), (16), are nonlinear, this model uses a linear function as below to estimate the mean *matching cost* component of ride-hailing customers' waiting cost,

$$c_1 + c_2 y_{ij}^m - c_3 \sum_{k \in \mathcal{D}} z_{kij}^m, \quad \forall (i,j) \in \mathcal{E}, \forall m \in \mathcal{M}$$

where c_1, c_2, and c_3 are positive constants. In the numerical experiments, we set $c_1 = 100\gamma_1^m$, $c_1 = 2\gamma_1^m$, and $c_1 = \gamma_1^m$. The linear function above can be explained as follows: When the ride-hailing customers' arrival rate y_{ij}^m is larger, the customers' waiting cost increases; When the TNCs' service rate $\sum_{k \in \mathcal{D}} z_{kij}^m$ becomes larger, the customers' waiting cost decreases.

The comparison between different models is summarized in Table 3. The models are compared regarding vehicle miles traveled (VMT), deadheading miles, average waiting cost, and percentage of ride-hailing customers among all travelers. The average waiting cost and the percentage of ride-hailing customers are the mean values over all OD pairs in the network. The results show that compared with our model, Ban et al. (2019) and the model with linear waiting cost function underestimate the average waiting cost, which leads to more travelers choosing the ride-hailing service. For example, the linear waiting cost function underestimates the waiting cost by 29.9% compared with our method. Consequently, our model produces the smallest deadheading miles and VMT among the three models. To be specific, Ban et al. (2019) and the model with a linear waiting cost function overestimate VMT by 63.8% and 40.7%, respectively, compared with our model.

The results of each OD pair for the proposed model, a model with linear waiting cost, and Ban et al. (2019) are listed in Tables 4, 5, and 6, respectively, including mode choice, customer disutility, mode-specific cost, and waiting cost. We can see that the observation from Table 3 at an aggregate level is also applicable to the OD pairs. Take OD pair 1 → 19 as an example, Ban et al. (2019) and the model with linear waiting cost function underestimate the waiting cost of ride-hailing customers compared with our model. As a result, Ban et al. (2019) and the model with linear waiting cost function overestimate the mode share of ride-hailing travelers both by 23%.

Table 5. Results of the Sioux-Fall network for a model with linear waiting cost.

OD pair	Mode Choice	Customer Disutility	Mode-specific Cost	Waiting Cost
1 → 13	Solo Driving: 266	649	649	/
	TNC I: 85	649	61	588
	TNC II: 119	649	189	460
	Transit: 30	649	649	/
1 → 19	Solo Driving: 0	1298	1298	/
	TNC I: 148	912	118	794
	TNC II: 152	912	376	536
	Transit: 0	1298	1298	/
2 → 13	Solo Driving: 0	1003	1003	/
	TNC I: 138	838	92	746
	TNC II: 162	838	291	547
	Transit: 0	1003	1003	/
2 → 19	Solo Driving: 0	944	944	/
	TNC I: 54	599	87	512
	TNC II: 46	599	274	325
	Transit: 0	944	944	/
4 → 13	Solo Driving: 351	649	649	/
	TNC I: 85	649	61	588
	TNC II: 119	649	189	460
	Transit: 45	649	649	/
4 → 19	Solo Driving: 0	1003	1003	/
	TNC I: 96	731	92	639
	TNC II: 104	731	291	440
	Transit: 0	1003	1003	/

Table 6. Results of the Sioux-Fall network for Ban et al. (2019).

OD pair	Mode Choice	Customer Disutility	Mode-specific Cost	Waiting Cost
1 → 13	Solo Driving: 0	649	649	/
	TNC I: 0	229	61	168
	TNC II: 500	226	189	37
	Transit: 0	649	649	/
1 → 19	Solo Driving: 0	1298	1298	/
	TNC I: 300	252	119	133
	TNC II: 0	440	376	64
	Transit: 0	1298	1298	/
2 → 13	Solo Driving: 0	1003	1003	/
	TNC I: 300	210	92	118
	TNC II: 0	341	291	50
	Transit: 0	1003	1003	/
2 → 19	Solo Driving: 0	944	944	/
	TNC I: 0	320	87	233
	TNC II: 100	318	274	44
	Transit: 0	944	944	/
4 → 13	Solo Driving: 0	649	649	/
	TNC I: 0	228	61	167
	TNC II: 600	226	189	37
	Transit: 0	649	649	/
4 → 19	Solo Driving: 0	1003	1003	/
	TNC I: 0	335	92	243
	TNC II: 200	334	291	43
	Transit: 0	1003	1003	/

10 Conclusions and Future Research

In this study, we propose a general traffic equilibrium model that captures the complex interactions between TNCs' decisions on how to dispatch ride-hailing vehicles, travelers' choices on which mode to use (ride-hailing, solo driving, and public transit), as well as explicitly model the queueing dynamics of ride-hailing customer waiting. The overall equilibrium model is formulated as a generalized Nash equilibrium coupled with a queueing system, and the existence of an equilibrium is analyzed. To solve the proposed model, we derive an equivalent mixed complementarity formulation. Our proposed model is validated using the Sioux-Falls network and compared with Ban et al. (2019) and a linear waiting cost function. Numerical results show that without modeling the waiting cost explicitly, Ban et al. (2019) underestimates the waiting cost and, as a result, overestimates the mode share of ride-hailing travelers and VMT in the system. For example, Ban et al. (2019) overestimates VMT by 63.8% compared with our model. Similarly, if we use a linear waiting cost function, it will underestimate the waiting cost by 29.9% compared with our method. Consequently, a linear waiting cost function would output 50% more mode share of ride-hailing travelers and overestimate the VMT by 40.7%. With the general modeling framework, transportation planners can better understand ride-hailing customers' waiting costs and deadheading miles induced by ride-hailing vehicles. This could help policymakers develop appropriate incentives that capture the features of ride-hailing services to reduce congestion in the transportation system.

Future research directions of this project are listed as follows: (A) to develop solution algorithms for large-scale problems; (B) to consider the congestion effect in the traffic network; (C) to include a driver-customer matching module in the model; (D) to capture more complicated queueing dynamics between ride-hailing drivers and customers.

Appendix 1. The Overall Equivalent Mixed Complementarity Formulation.

$$0 \le z^m_{kij} \perp -R^m_{kij} - \underbrace{(\lambda^m_{ij} + \mu^m_k)}_{\substack{\text{compensation}\\ \text{and surge price}}} + t_{ki}\zeta^m \ge 0, \qquad \forall\, k \in \mathcal{D}, \forall\, (i,j) \in \mathcal{E}, m \in \mathcal{M}$$

$$0 \le \underbrace{\lambda^m_{ij}}_{\text{compensation}} \perp \sum_{k\in\mathcal{D}} z^m_{kij} - y^m_{ij} \ge \epsilon, \qquad \forall\, (i,j) \in \mathcal{E}, \forall\, m \in \mathcal{M}$$

$$\underbrace{\mu^m_k}_{\substack{\text{compensation}\\ \text{or surge price}}} \text{ free}, \quad \sum_{(i,j)\in\mathcal{E}} z^m_{kij} = \sum_{i\in\mathcal{O}} y^m_{ik}, \qquad \forall\, k \in \mathcal{D}, \forall\, m \in \mathcal{M}$$

$$0 \le \zeta^m \perp N^m - \sum_{(i,j)\in\mathcal{E}} \sum_{k\in\mathcal{D}} t_{ki}\, z^m_{kij} - \sum_{(i,j)\in\mathcal{E}} t_{ij}\, y^m_{ij} \ge 0, \qquad \forall\, m \in \mathcal{M}$$

$$0 \le y^m_{ij} \perp \underline{\alpha}^m_i + \alpha^m_1 t_{ij} + \alpha^m_2 d_{ij} + \gamma^m_2 t_{ij} + \underbrace{w^m_{ij}}_{\text{waiting cost}} - \underbrace{u_{ij}}_{\text{least disutility}} \ge 0, \quad \forall\, (i,j) \in \mathcal{E}, m \in \mathcal{M}$$

$$0 \le y^0_{ij} \perp \alpha^0_2 d_{ij} + \gamma^0_2 t_{ij} - \underbrace{u_{ij}}_{\text{least disutility}} \ge 0, \qquad \forall\, (i,j) \in \mathcal{E}$$

$$0 \le y^{M+1}_{ij} \perp V^{M+1}_{ij} - \underbrace{u_{ij}}_{\text{least disutility}} \ge 0, \qquad \forall\, (i,j) \in \mathcal{E}$$

$$\underbrace{u_{ij}}_{\text{least disutility}} \text{ free}, \quad y^0_{ij} + y^{M+1}_{ij} + \sum_{m\in\mathcal{M}} y^m_{ij} = Q_{ij}, \qquad \forall\, (i,j) \in \mathcal{E}$$

$$0 \le w^{\mathrm{m},m}_{ij} \perp \left[\sum_{k\in\mathcal{D}} z^m_{kij} \left(\sum_{k\in\mathcal{D}} z^m_{kij} - y^m_{ij}\right)\right] w^{\mathrm{m},m}_{ij} - \gamma^m_1 y^m_{ij} + \eta^m_{ij} \ge 0, \qquad \forall (i,j) \in \mathcal{E},\ \forall\, m \in \mathcal{M}$$

$$0 \le \eta^m_{ij} \perp w^{\max}_{ij} - w^{\mathrm{m},m}_{ij} \ge 0, \qquad \forall (i,j) \in \mathcal{E},\ \forall\, m \in \mathcal{M}$$

$$0 \le \theta^m_{kij} \perp \left[\sum_{k'\in\mathcal{D}} z^m_{k'ij}\right] \theta^m_{kij} - \gamma^m_1 z^m_{kij} + \psi^m_{kij} \ge 0, \qquad \forall k \in \mathcal{D},\ \forall (i,j) \in \mathcal{E},\ \forall\, m \in \mathcal{M}$$

$$0 \le \psi^m_{kij} \perp 1 - \theta^m_{kij} \ge 0, \qquad \forall k \in \mathcal{D},\ \forall (i,j) \in \mathcal{E},\ \forall\, m \in \mathcal{M}$$

$$0 \le w^{\max}_{ij} - w^m_{ij} \perp \underbrace{\gamma^m_1 \sum_{k\in\mathcal{D}} (t_{ki}\theta^m_{kij})}_{\text{pickup cost}} + \underbrace{w^{\mathrm{m},m}_{ij}}_{\text{matching cost}} - w^m_{ij} \ge 0, \qquad \forall (i,j) \in \mathcal{E},\ \forall m \in \mathcal{M}$$

References

Allyn, B. (2021). Lyft and Uber prices are high, wait times are long and drivers are scarce. *National Public Radio*. Retrieved from https://www.npr.org/2021/08/07/1025534409/lyft-and-uber-prices-are-high-wait-times-are-long-and-drivers-are-scarce.

Bai, J., So, K. C., Tang, C. S., Chen, X., & Wang, H. (2019). Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*, 21(3), 556-570.

Ban, X. J., Dessouky, M., Pang, J. S., & Fan, R. (2019). A general equilibrium model for transportation systems with e-hailing services and flow congestion. *Transportation Research Part B: Methodological*, 129, 273-304.

Banerjee, S., Kanoria, Y., & Qian, P. (2022). Large Deviations Optimal Scheduling of Closed Queueing Networks. Available at https://arxiv.org/abs/1803.04959v4.

Braverman, A., Dai, J. G., Liu, X., & Ying, L. (2019). Empty-car routing in ridesharing systems. *Operations Research*, 67(5), 1437-1452.

Byrd, R. H., Nocedal, J., & Waltz, R. A. (2006). Knitro: An integrated package for nonlinear optimization. In Large-scale nonlinear optimization (pp. 35-59). *Springer*, Boston, MA.

Castiglione, J., Cooper, D., Sana, B., Tischler, D., Chang, T., Erhardt, G. D., ... & Mucci, A. (2018). TNCs & Congestion. *San Francisco County Transportation Authority*.

Castillo, J. C., Knoepfle, D. T., & Weyl, E. G. (2017). Surge pricing solves the wild goose chase. *In Proceedings of the 2017 ACM Conference on Economics and Computation* (pp. 241-242).

Castillo, J. C., Knoepfle, D. T., & Weyl, E. G. (2023). Matching and pricing in ride hailing: Wild goose chases and how to solve them. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2890666.

Castro, F., Ma, H., Nazerzadeh, H., & Yan, C. (2021). Randomized FIFO mechanisms. Available at https://arxiv.org/abs/2111.10706.

Chen, X., & Di, X. (2023). A unified network equilibrium for e-hailing platform operation and customer mode choice. Available at https://arxiv.org/abs/2203.04865.

Daganzo, C. F., & Ouyang, Y. (2019). A general model of demand-responsive transportation services: From taxi to ridesharing to dial-a-ride. *Transportation Research Part B: Methodological*, 126, 213-224.

Di, X., & Ban, X. J. (2019). A unified equilibrium framework of new shared mobility systems. *Transportation Research Part B: Methodological*, 129, 50-78.

Erhardt, G. D., Roy, S., Cooper, D., Sana, B., Chen, M., & Castiglione, J. (2019). Do transportation network companies decrease or increase congestion?. *Science Advances*, 5(5), eaau2670.

Facchinei, F., & Pang, J. S. (2003). Finite-dimensional variational inequalities and complementarity problems. *Springer*, New York, NY.

Feng, G., Kong, G., & Wang, Z. (2021). We are on the way: Analysis of on-demand ride-hailing systems. *Manufacturing & Service Operations Management*, 23(5), 1237-1256.

Feng, S., Ke, J., Xiao, F., & Yang, H. (2022). Approximating a ride-sourcing system with block matching. *Transportation Research Part C: Emerging Technologies*, 145, 103920.

He, F., & Shen, Z. J. M. (2015). Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C: Emerging Technologies*, 58, 93-106.

Ke, J., Yang, H., Li, X., Wang, H., & Ye, J. (2020a). Pricing and equilibrium in on-demand ride-pooling markets. *Transportation Research Part B: Methodological*, 139, 411-431.

Ke, J., Yang, H., & Zheng, Z. (2020b). On ride-pooling and traffic congestion. *Transportation Research Part B: Methodological*, 142, 213-231.

Iqbal, M., (2024). Uber revenue and usage statistics. Retrieved from: `https://www.businessofapps.com/data/uber-statistics`.

Li, S., Tavafoghi, H., Poolla, K., & Varaiya, P. (2019). Regulating TNCs: Should Uber and Lyft set their own rules?. *Transportation Research Part B: Methodological*, 129, 193-225.

Li, Y., Liu, Y., & Xie, J. (2020). A path-based equilibrium model for ridesharing matching. *Transportation Research Part B: Methodological*, 138, 373-405.

Ma, H., Fang, F., & Parkes, D. C. (2022). Spatio-temporal pricing for ridesharing platforms. *Operations Research*, 70(2), 1025-1041.

Qian, X., & Ukkusuri, S. V. (2017). Taxi market equilibrium with third-party hailing service. *Transportation Research Part B: Methodological*, 100, 43-63.

Ouyang, Y., & Yang, H. (2023). Measurement and mitigation of the "wild goose chase" phenomenon in taxi services. *Transportation Research Part B: Methodological*, 167, 217-234.

Özkan, E., & Ward, A. R. (2020). Dynamic matching for real-time ride sharing. *Stochastic Systems*, 10(1), 29-70.

Patriksson, M. (2015). The traffic assignment problem: models and methods. *Courier Dover Publications.*

Rosenberg, E. (2021). Need an Uber or Lyft in Boston? Expect a long wait: 'It takes forever.' *National Broadcasting Company Boston.* Retrieved from `https://www.nbcboston.com/news/local/need-an-uber-or-lyft-in-boston-expect-a-long-wait-it-takes-forever/2347540/`.

Schaller, B. (2017). Empty seats, full streets: Fixing Manhattan's traffic problem. *Technical report, Schaller*

Consulting, Brooklyn, NY.

Schaller, B. (2018). The new automobility: Lyft, Uber and the future of American cities. *Technical report, Schaller Consulting*, Brooklyn, NY.

Sheffi, Y. (1985). Urban transportation networks. *Prentice-Hall*, Englewood Cliffs, NJ.

Spiegel, A. (2024). Why Uber wait times may be getting longer around D.C. *Axios Washington D.C.* Retrieved from https://www.axios.com/local/washington-dc/2024/05/08/dc-uber-wait-times-more-expensive.

Stabler, B. (2020). Transportation Networks for Research. Available at https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls.

Sun, L., Teunter, R. H., Hua, G., & Wu, T. (2020). Taxi-hailing platforms: Inform or Assign drivers?. *Transportation Research Part B: Methodological*, 142, 197-212.

Taylor, T. A. (2018). On-demand service platforms. *Manufacturing & Service Operations Management*, 20(4), 704-720.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1(3), 325-362.

Wang, G., Zhang, H., & Zhang, J. (2024). On-demand ride-matching in a spatial model with abandonment and cancellation. *Operations Research*, 72(3), 1278-1297.

Wang, H., & Yang, H. (2019). Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological*, 129, 122-155.

Wong, K. I., Wong, S. C., Yang, H., & Wu, J. H. (2008). Modeling urban taxi services with multiple user classes and vehicle modes. *Transportation Research Part B: Methodological*, 42(10), 985-1007.

Xu, H., Pang, J. S., Ordonez, F., & Dessouky, M. (2015). Complementarity models for traffic equilibrium with ridesharing. *Transportation Research Part B: Methodological*, 81, 161-182.

Xu, Z., Chen, Z., Yin, Y., & Ye, J. (2021). Equilibrium analysis of urban traffic networks with ride-sourcing services. *Transportation Science*, 55(6), 1260-1279.

Xu, Z., Yin, Y., & Ye, J. (2020). On the supply curve of ride-hailing systems. *Transportation Research Part B: Methodological*, 132, 29-43.

Yan, C., Zhu, H., Korolko, N., & Woodard, D. (2019). Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics*, 67(8), 705-724.

Yang, H., & Wong, S. C. (1998). A network model of urban taxi services. *Transportation Research Part B: Methodological*, 32(4), 235-246.

Zha, L., Yin, Y., & Xu, Z. (2018). Geometric matching and spatial pricing in ride-sourcing markets.

Transportation Research Part C: Emerging Technologies, 92, 58-75.

Zhang, K., & Nie, Y. M. (2021). To pool or not to pool: Equilibrium, pricing and regulation. *Transportation Research Part B: Methodological*, 151, 59-90.

Zhong, Y., Lan, Y., Chen, Z., & Yang, J. (2023). On-demand ride-hailing platforms with heterogeneous quality-sensitive customers: Dedicated system or pooling system?. *Transportation Research Part B: Methodological*, 173, 247-266.

Data Management Plan

Products of Research

The main research products will be peer-reviewed journal articles, book chapters and/or conference proceedings targeted towards the transportation science research community, plus supplemental materials such as tables, numerical data used for graphs, etc. No personal data will be used in the project, so there is no threat of identity theft.

Data Format and Content

All research products will be available online in digital form. Manuscripts will appear in a common document-viewing format, such as PDF, and supplemental materials such as tables and numerical data will be in a tabular format, such as Microsoft Excel spreadsheet, tab-delimited text, etc.

Data Access and Sharing

All participants in the project will publish the results of their work. Papers will be published in peer-reviewed scientific journals, books published in English, conference proceedings, or as peer-reviewed data reports. Beyond the data posted on USC websites, primary data and other supporting materials created or gathered in the course of work will be shared with other researchers upon reasonable request, at no more than incremental cost and within a reasonable time of the request or, if later, the filing of a patent application covering the results of such research.

All the data used in the research are included in Tables in the final report or are publicly available. For the numerical experiments, the parameters of TNC fee charging and traveler disutility can be found in Table 2, and the data of the Sioux Falls can be found in the following link:

`https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls`

Stabler, B. (2020). Transportation Networks for Research.

Reuse and Redistribution

USC's policy is to encourage, wherever appropriate, research data to be shared with the general public through internet access. This public access will be regulated by the university in order to protect privacy and confidentiality concerns, as well to respect any proprietary or intellectual property rights. Administrators will consult with the university's legal office to address any concerns on a case-by-case basis, if necessary.

Terms of use will include requirements of attribution along with disclaimers of liability in connection with any use or distribution of the research data, which may be conditioned under some circumstances.