A General Equilibrium Model for Transportation Systems with e-Hailing Services and Flow Congestion

Xuegang (Jeff) Ban* Maged Dessouky Jong-Shi Pang† Rong Fan‡

Original September 2018; revised April 20, 2019; second revision June 20, 2019

Abstract

Passengers are increasingly using e-hailing as a means to request transportation services. Adoption of these types of services has the potential to impact the travel behavior of individuals as well as increase congestion and vehicle miles driven since extra deadhead miles must be added to the trip (e.g., the extra miles from the driver location to the pick-up location of the customer). The objective of this paper is to develop a basic mathematical model to help transportation planners understand the relationship between the wide-scale use of e-hailing transportation services and deadhead miles and resulting impact on congestion. Specifically, this paper develops a general economic equilibrium model at the macroscopic level to describe the equilibrium state of a transportation system composed of solo drivers and the e-hailing service providers (e-HSPs). The equilibrium model consists of three interacting sub-models: e-HSP choice, customer choice, and network congestion; the model is completed with a "market clearance" condition describing the waiting costs in the customer's optimization problem in terms of the e-HSPs' decisions, thereby connecting the supply and demand sides of the equilibrium. We show the existence of an equilibrium of a certain kind under some mild assumptions. In numerical experiments, we illustrate the sensitivity on the usage of these modes of transport to various parameters representing cost, value of time, safety, and comfort level, as well as the resulting relationship between usage of these services and vehicle miles.

1 Introduction

Passengers are increasingly using e-hailing apps such as those provided by taxi companies and Transportation Network Companies (TNCs), e.g., Uber, Lyft, and Didi as a means to request transportation services. These e-hailing service providers (e-HSPs) are transforming the travel behavior of individuals and urban mobility patterns. However, there is a lack of models and tools that can guide transportation planners in understanding how these emerging industries will impact network congestion. Although trips

*Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195-2700, U.S.A. Email: banx@uw.edu.

†Dessouky and Pang are affiliated with the Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90089-0193, U.S.A. Emails: maged@usc.edu, jongship@usc.edu.

‡Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195-2700, U.S.A. Email: Rongfan6@uw.edu.

on these services have the potential to reduce congestion due to less time driving in search of parking if needed, they also have the potential to increase congestion and vehicle miles driven since extra deadhead miles must be added to the trip (e.g., the extra miles from the driver location to the pick-up location of the customer). It is expected that at low adoption (usage) levels of these services, the deadhead miles will be high since drivers will most likely have to drive longer distances to pick up customers; but as more people start adopting these services the deadhead miles may start to decrease at some point in the usage level since the likelihood that there is a driver closer to the origin location of a customer also increases. In fact, recent data have shown that the impact of such emerging e-hailing services to urban transportation systems can be significant (e.g., e-hailing has increased the city-wide mileage for about 14% - 19% in New York City [20]), which imposes increasing pressure to municipalities on managing and regulating those emerging mobility services. The models developed in this work can be used to assess when travelers prefer one type of e-hailing service over another as well as measuring the congestion effect of the deadhead miles caused by e-HSP vehicles.

Specifically, this paper aims to develop a general economic equilibrium model at the macroscopic (i.e., planning) level to describe the equilibrium state of a transportation system composed of solo drivers and e-HSP vehicles. Each of these companies acts as an economic agent of the system selfishly optimizing their companies' objectives. Each e-HSP provider has its own unique fare structure as well as the passengers' perception of the safety and convenience of the provider. The demand side of the system consists of the trip makers traveling between origin and destination (OD) pairs; some of these travelers choose to be solo drivers while others choose to be an e-HSP customer for their trips. We model each e-HSP as a profit-maximization agent, responding (or not) to e-hailing calls from customers who are not solo drivers. The e-HSP customers and the solo drivers selfishly minimize their disutility of travel while fulfilling the travel demands. The e-HSPs, their customers, and the solo drivers are linked by several monetary factors that include (a) trip prices, both fixed and variable, that are charged by the e-HSPs and paid by their customers, (b) time values of the travelers due to waiting for service and during travel, and (c) miscellaneous considerations such as safety and convenience to decide between solo driving and being an e-HSP customer of a particular provider type, and by various trip characteristics such as provider's behavior of route choices and traffic congestion as a result of the totality of trips made by the network users. All these factors are intrinsic to this *traffic general equilibrium model* with e-hailing transportation service providers (TGEM-ets) that distinguish it from the classical economic equilibrium model of Arrow–Debreu where the producers and consumers in an economy are linked by a market clearing condition that dictates the prices of the economic goods. In the past, mixed equilibrium has been proposed to study the equilibrating behavior of multiple players in a transportation system [14, 13, 35]. Such an equilibrium may be considered as a special case of a general equilibrium since it imposes certain restrictive assumptions on the model constraints such as separability [14, 35], linearity [36], or joint convexity [37]. In contrast, the proposed TGEM-ets contains constraints in the e-HSP providers' problems that couple these variables with the customer's variables; see equation (8), making the resulting model a Nash equilibrium problem of the generalized type; see [6, 19]. Moreover, the nonlinearity and some non-standard structure make the model much more challenging to analyze as shown in Section 7.

In the TGEM-ets, there is a cost mechanism exogenous to the e-HSPs and their customers that determine the e-HSP customers' costs due to the wait times of the arrival of e-HSP vehicles. Using this cost mechanism as the "market clearing" condition exogenous to the providers' and customers' decision making problems, the resulting TGEM-ets is formulated as a non-cooperative game with these traffic participants as its players and with the following four major components:

- e-HSPs' choices (making decisions such as where to pick up the next customer) to maximize the profits;
- travelers' choices (making driving decisions to be either a solo driver or an e-HSP customer) to minimize individual disutility;
- network equilibrium to capture traffic congestion as a result of everyone's travel behavior that is dictated by Wardrop's route choice principle;.
- market clearance conditions to define customers' waiting cost and constraints ensuring that the e-HSP OD demands are served.

To capture all these major components, the proposed equilibrium model consists of three interacting sub-models (also called principal modules). The first module, as discussed in detail in Section 3, describes the behavior of the e-HSPs who aim to maximize their profit. The second describes the consumer choice behavior assuming that each traveler minimizes his/her individual disutility, as detailed in Section 4. The third module, presented in Section 5, describes the traditional network congestion effects with the added feature of accounting for the deadhead miles of the e-HSP vehicles. In addition, the market clearance condition connects the e-HSP provider and customer models by providing a mechanism to calculate customers' waiting costs. The latter market condition is discussed in Sections 3, 4.1, and Appendix I. The overall TGEM-ets model is summarized in Section 6 wherein a graphical depiction (in Figure 1) shows the connections of these major components. Besides the proposed model formulation, one important contribution of this work is a rigorous proof of existence of an equilibrium solution to the formulated TGEM-ets, thus providing a sound analysis to a new and rather complex e-hailing transportation system that is complicated by the various novel types of economic agents in the system. Furthermore, experimental analysis shows the sensitivity of the solution to changes in the revenue and cost functions of each e-HSP mode and the disutility function of the consumers.

There have been previous studies [31, 25, 32, 26, 30, 33] that have developed network models describing urban street-hailing taxi services in a stable equilibrium state, which have also provided very useful insight to the development of the TGEM-ets model proposed in this paper. A major distinction between these earlier models and the equilibrium model of this paper is that the proposed model here considers multi-modes (including solo driving, and multiple e-hailing companies) resulting in a general equilibrium model, while the previous studies only considered taxis. Another major difference is in the way vehicles are matched with the customers. In the earlier work it is assumed that customers hail a taxi when it is driving in its zone while the model in this paper explicitly accounts for the e-HSP vehicles in response to customer requests made by a platform (phone or mobile device). This latter feature allows the model to then explicitly measure the impact of the deadhead miles on congestion, represented as vehicle miles traveled (VMT) in this paper. More recently, He and Shen [15] explicitly studied the taxi market equilibrium with both street-hailing and e-hailing services. In their setting, both customers and taxi drivers can choose to use either street-hailing or e-hailing. Similar to [30], they showed that the equilibrium condition can be expressed as a system of nonlinear equations, for which the Brouwer's fixed-point theorem can be applied to prove the existence of equilibrium. Our proposed TGEM-ets differs from [15] in three important ways. First, we consider solo driving, and multiple e-hailing transportation service providers (including e-hailing taxi), leading to a multiple player game or a general economic equilibrium problem. In particular, the complex (nonlinear) coupling constraints among different players make it extremely challenging, if not impossible, to apply standard fixed-point theorems to the TGEM-ets introduced in this paper to establish its solution existence. Rather we develop a novel penalty-based method to prove the solution existence of TGEM-ets. Secondly, in [15], the waiting times between vehicles and customers were modeled using the same method for street-hailing taxis [30]. Since we only focus on e-hailing services in this paper, we model the waiting times differently as discussed in detail in Assumption (e) in Section 2. Thirdly, while travel times are assumed to be given in [15], we consider congestion effects and allow the travel times to be endogenously determined by the traffic equilibrium module. This leads to the nonconvex inequality vehicle balance constraint that significantly complicates the analysis (such as solution existence) of the model. We should mention that there has been some work on developing equilibrium models for traditional ridesharing services (e.g., [28, 27]) where they are not modeled as "for-profit" companies like most e-hailing transportation providers.

The remainder of the paper is organized in eight sections. Section 2 lays down the setting of the model, starting from the major assumptions of the model, and introducing the ingredients of the traffic network, and the model constants, variables, and cost functions. The next three sections 3, 4, and 5 present the key modules of TGEM-ets consisting of the e-HSP's choice module, the customer choice module, and the network congestion module, respectively. Combining these modules, Section 6 formally presents the TGEM-ets and defines an equilibrium solution. A flow chart depicting the connections of these modules is provided at the end of Section 6 after the mathematical formulation of the model is presented. In Section 7, following a preview of the existence proof given in Section 6, we establish the existence of an equilibrim solution of the TGEM-ets under a set of reasonable assumptions that include in particular a condition on the number of e-HSP vehicles to meet the given travel demands. Then, in Section 8, we report the numerical results with the use of an off-the-shelf solver for solving the TGEM-ets formulated as a mixed complementarity problem and discuss the insights derived from these results. The last section provides concluding remarks and future research directions.

2 Model Setting

Transportation network analysis with e-hailing is new and complex. In building a model to address such a tremendously complex system, we have adhered to two basic yet opposite principles: (a) faithfulness to realism, and (b) tractability in analysis and computation. The proposed model aims to capture the basic behavior of e-HSP choices, customer choices, and their interactions (i.e., the congestion effect) on traffic networks. The model is based on the unique characteristics of e-hailing services. First, e-hailing is often built upon a certain platform (mostly mobile apps/devices) to dispatch vehicles to serve customers. In such a system, e-HSP vehicles/drivers largely behave according to the guidance (i.e, dispatch) of the platform (i.e., the e-HSP). Second, due to the platform dispatch, e-HSP vehicles do not need to drive around to look for customers. Rather they can stay at the drop-off locations or some preferred locations to wait for the next service call (dispatch). For simplicity, we assume in this paper an e-HSP driver always stay at the drop-off locations and wait for the next service call. Based on these two features, we state explicitly the main assumptions underlying the mathematical formulation of the TGEM-ets.

Major assumptions

(a) The model is for high-level planning purposes and assumes/describes certain stable equilibrium behavior and travel patterns.

(b) There is a given number of vehicles for each e-HSP provider to serve customer requests; this number is balanced with the e-HSP trips; see equation (7). This assumption is an approximation for practical e-HSP services since some of the drivers may not actively work for a given period of time. However it may still be reasonable to assume that the number of e-HSP drivers is relatively stable for a given day or time of day. In future research, we may relax this assumption and consider the number of drivers as a variable. For this, we need to introduce additional models to capture the behavior of how e-HSP drivers choose different e-HSP companies and how they choose their work hours.

(c) An e-HSP has a platform (e.g., based on mobile apps) to match e-HSP vehicles and customers and to dispatch/route e-HSP vehicles, largely for the benefit of the e-HSP. This is very different from traditional street-hailing taxi services where taxi drivers are more “selfish” and have more freedom to decide how to drive around and where to pickup customers. In particular, since e-HSP vehicles respond to the dispatch of the platform, we model the e-HSP vehicle’s (aggregated) choice behavior from the e-HSP’s perspective. In other words, we assume that the e-HSP vehicles from the same provider will behave for the benefit of the company. We call this behavior the “cooperative” choice behavior of the e-HSP vehicles. This is in contrast to the “selfish” behavior of traditional taxis assumed in the literature. Due to the e-hailing dispatch platform, this cooperative behavior assumption may be reasonable from a high-level planning perspective, at least compared with traditional taxis. Note that the “selfish” decision-making behavior of drivers may be captured in the proposed model by letting each driver be their own e-HSP; this refinement may be investigated in future research.

(d) Every traveler has access to a car. S/he then decides which service to use (or drive alone) and then request the service (if not driving alone) via the mobile platform. We assume that a traveler utilizes a single travel mode for an entire trip, and no transfers between modes are modeled in this paper.

(e) We model the waiting times (costs) of e-HSP vehicles and customers separately. The e-HSP waiting times are internally determined, as shown in equation (7). The customers’ waiting costs are modeled as the sum of the matching time cost with e-HSP vehicles and the traveling time cost from the e-HSP vehicle location to the pick-up location. Here we do not follow exactly the established relationships of service vehicle and customer waiting times for traditional street-hailing taxis (e.g., the ones in [30]). This is mainly due to the unique features of e-hailing services; those previously established waiting functions may or may not be applicable in this context. The determination of such functions remains an active research area [38]. To the best of the authors' knowledge, no commonly agreed functions have yet been developed to date for e-hailing services. Thus we choose to model these two sets of waiting times/costs separately. We provide more discussions on this in Appendix I.

(f) Since one important purpose of the proposed equilibrium model is to evaluate the congestion effect of the e-hailing services, we include a traffic flow module to capture the interactions of trip flows and congestion for the determination of travel times. If congestion is ignored—thus the travel times are exogenous to the model—the third module on network congestion effects in Section 5 is not necessary. However, the resulting predetermined travel-time model will have the same mathematical structure as that with the congestion module, both as a "generalized Nash equilibrium problem". The analysis and solution of the model without explicit consideration of congestion, although can be simplified, remains not straightforward. Most importantly, it would be quite un-realistic for a system equilibrium model to exclude congestion effects in the presence of a large number of trip makers.

(g) Distance and free-flow travel times in the e-HSP profit and customer disutility models are assumed to be derived from the respective minimum-free-flow-travel-time paths; see equations (5) and (12).

We are now ready to formally introduce the details of the TGEM-ets. We first describe the network structure and known parameters that define the model; this is followed by the description of the decision variables to be determined from the model. The major components of the model are then described and formulated as optimization problems (for the e-HSPs), mixed complementarity problems (MCPs, for customer choices and the route choice under traffic congestion), and conservation of e-HSP vehicles. In this paper, we include solo driving as a separate mode, and for notational convenience, we model it as a type of e-HSP (i.e., e-HSP$_0$) so that the customer choice module in Section 4 can be neatly formulated. Note that solo driving is very different from the other normal e-HSPs since it does not serve regular customers. There is only one state of the solo drivers: they just drive and do not wait or pick up customers. For this reason, solo drivers do not appear in the e-HSP choice module. Rather solo drivers only serve themselves and their choice behavior is modeled in the customer choice model in Section 4 as to minimize their individual disutility. Furthermore, since the model has assumed that the e-HSP drivers are cooperative within the same company, i.e., they have already made the decision to work for one of the e-HSPs, the fixed cost of the e-HSP drivers is assumed to be zero. Thus, the only decision of each e-HSP vehicle is which customers to serve. If a traveler decides to be a solo driver, the fixed cost portion of that trip is also assumed to be zero since the model assumes that the traveler already owns or leases a vehicle. Thus, the fixed cost of all driving modes (solo drivers and e-HSP drivers) is assumed to be zero.

Sets associated with the network. The following notations are used:

$\mathcal{N}$	set of nodes in the network
$\mathcal{A}$	set of links in the network, subset of $\mathcal{N} \times \mathcal{N}$
$\mathcal{K}$	set of OD pairs, subset of $\mathcal{N} \times \mathcal{N}$
$\mathcal{O}$	set of origins, subset of $\mathcal{N}$; $\mathcal{O} = \{ \mathrm{O}_k : k \in \mathcal{K} \}$
$\mathcal{D}$	set of destinations, subset of $\mathcal{N}$; $\mathcal{D} = \{ \mathrm{D}_k : k \in \mathcal{K} \}$ besides being the destinations of the OD pairs where customers are dropped off, these are also the locations where the e-HSP drivers initiate their next trip to pick up other customers
O_k, D_k	origin and destination (sink) respectively of OD pair $k \in \mathcal{K}$
$\mathcal{M}$	labels of the e-HSPs; $\mathcal{M} \triangleq \{1, \cdots, M\}$
$\mathcal{M}_+$	union of the solo driver/customer label (0) and the e-HSP labels; thus $\mathcal{M}_+ = \{ 0 \} \cup \mathcal{M}$
e-HSP$_m$	e-HSP of type $m \in \mathcal{M}$; an e-HSP$_0$ customer/driver is a solo driver
$\mathcal{P}_{ij}$	set of paths with start node $i \in \mathcal{N}$ and end node $j \in \mathcal{N}$; see below for the essential paths
$\mathcal{P}$	set of above paths; $\mathcal{P} \triangleq \bigcup_{(i,j) \in \mathcal{N} \times \mathcal{N}} \mathcal{P}_{ij}$.

Among all paths in the network, we are only interested in those whose start node i and end node j are either the origin or destination of some OD pairs. Depending on the roles of i and j, we may classify the path p joining them as:

- an OD-path if $i = \mathrm{O}_k$ and $j = \mathrm{D}_k$ for some $k \in \mathcal{K}$, thus p is a path serving OD pair k;
- an e-HSP-path if $i = \mathrm{D}_k$ for some $k \in \mathcal{K}$ and $j = \mathrm{O}_\ell$ for some $\ell \in \mathcal{K}$ where i and j may be the same or different locations; thus p is a path used by the driver of a vacant e-HSP vehicle, who after dropping off customers at node i, travels to node j to respond to a service call of customers waiting at j. The e-HSP-paths are one distinctive feature of the TGEM-ets.

By assumption, there are no vacant e-HSP$_m$ vehicles waiting at a node outside the set $\mathcal{D}$; these non-destination nodes in the network, i.e., nodes in the set $\mathcal{N} \setminus (\mathcal{O} \cup \mathcal{D})$ serve only as the transit nodes in the trip paths of all vehicles. Nevertheless, there may be empty e-HSP vehicles available at a destination node in the network to pick up customers at that node which happens to be the origin of another OD pair. In this paper, we assume that a vacant e-HSP vehicle may pick up a customer at any origin (despite the distance). In practice, it may be the case that an OD pair can only be served by nearby locations (or zones). The model described below can be suitably extended to handle this case by defining a maximum distance of the e-hailing service beyond which an e-HSP vehicle will not serve a customer.

In light of the assumption about the e-HSP pick-up locations and the OD trips, it would be useful to introduce some binary indicators between the OD pairs and their origins and destinations. Specifically, for a pair $(i, k) \in \mathcal{N} \times \mathcal{K}$ we define

$$\varepsilon^{\mathrm{o}}_{ik} \triangleq \left\{ \begin{array}{ll} 1 & \text{if } i = \mathrm{O}_k \\ 0 & \text{otherwise} \end{array} \right\} \quad \text{and} \quad \varepsilon^{\mathrm{d}}_{ik} \triangleq \left\{ \begin{array}{ll} 1 & \text{if } i = \mathrm{D}_k \\ 0 & \text{otherwise} \end{array} \right\}.$$

Based on these node-OD pair incidence indicators, we define some composite indicators. For a pair of nodes $(i,j) \in \mathcal{N} \times \mathcal{N}$ and an OD pair $k \in \mathcal{D}$, let

$$\delta^{\rm OD}_{ijk} \triangleq \varepsilon^{\rm o}_{ik}\,\varepsilon^{\rm d}_{jk} = \left\{ \begin{array}{ll} 1 & \text{if } i = \mathrm{O}_k \text{ and } j = \mathrm{D}_k \\ 0 & \text{otherwise} \end{array} \right\};$$

similarly, for a pair of nodes $(i',j') \in \mathcal{N} \times \mathcal{N}$ and two OD-pairs k and ℓ in $\mathcal{K}$, let

$$\delta^{\rm e-HSP}_{i'j'k\ell} \triangleq \varepsilon^{\rm d}_{i'k}\,\varepsilon^{\rm o}_{j'\ell} = \left\{ \begin{array}{ll} 1 & \text{if } i' = \mathrm{D}_k \text{ and } j' = \mathrm{O}_\ell \\ 0 & \text{otherwise} \end{array} \right\}.$$

The two composite indicators $\delta^{\rm OD}_{ijk}$ and $\delta^{\rm e-HSP}_{i'j'k\ell}$ serve as incidence indicators between given nodes i and j and the origin-destination nodes of the OD pair k and have different meanings. They will be used in the flow equations that balance path flows with OD demands and e-HSP trips en route to pick up customers; see Section 5. Notice that $\delta^{\rm e-HSP}_{i'j'k\ell} = 1$ means that the OD demand ℓ will be served by e-HSP drivers coming from i'.

Model parameters. The following are constants in the model:

Q_k	given travel demand rate (trips per unit time) of OD pair $k \in \mathcal{K}$
f^0_{ij}	(positive) free-flow travel time of a (free-flow-time based) shortest path from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$
d_{ij}	the distance of a (free-flow time based) shortest path from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$
$F^m_{\mathrm{O}_k}$	fixed fare charged by e-HSP$_m$ who picks up customers at O_k (the origin of OD pair k)
α^m_1, α^m_2	time and distance based fare rates, respectively, for e-HSP$_m$
β^m_1, β^m_2	respective conversion factors from travel time and distance to costs for e-HSP$_m$ drivers
β^m_3	the conversion factor from waiting time to costs for e-HSP$_m$ drivers
γ^m_1	value of time for e-HSP$_m$ customers while traveling in a vehicle
γ^m_2	value of time for e-HSP$_m$ customers while waiting for a vehicle to travel from its current location to the pick up location
γ^m_3	the multiplicative factor in a normalized equilibrium of the TGEM-ets
N^m	fleet size (i.e., the number of vehicles) of e-HSP$_m$.

We have $F^0_{\mathrm{O}_k} = 0$ and $\alpha^0_1 = \alpha^0_2 = 0$ as solo drivers do not need to be paid these fares. The parameters α^m_i and β^m_i are employed to calculate the e-HSP$_m$'s profit. These actual values depend on the specific characteristics of e-HSP modes. There are some reasonable relations among them across the e-HSPs; see Section 8 for more details.

Primary model variables. These are the key variables to be determined from the model; they induce certain secondary model variables (to be described next) that are needed to formulate the model conditions. Specifically, we have

Q^m_k	total demand (trips per unit time) of OD pair k served by e-HSP$_m$
z^m_{jk}	the total number of e-HSP$_m$ vehicles for the study period (also called "vehicle movements" [31]) from destination $j \in \mathcal{D}$ to serve customers of the OD pair k in response to the latter's e-hailing
t_{ij}	(congestion dependent) travel time of the shortest path from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$
h_p	vehicular flow on path $p \in \mathcal{P}$.

Two particular cases of the notation z_{jk}^m are worthy of note. First is the possibility that the node $j = \mathrm{D}_k$; this signifies that the e-HSP drivers may return to the origin O_k of the OD pair k to pick up new customers after dropping off previous customers at the destination of k. The other possibility is that $j = \mathrm{D}_\ell = \mathrm{O}_k$ for some OD pairs $\ell \neq k$; this signifies that the e-HSP drivers at the destination of OD-pair ℓ may pick up new customers at the same location who are the trip makers of OD pair k. The model accommodates both possibilities. Note that if the latter case happens for all e-HSP vehicles (i.e., they all pick up customers at their previous drop-off locations), the deadhead miles of the network will be zero. Otherwise, the deadhead miles will be positive. In particular, the total deadhead miles (denoted as DH) of the network can be expressed as:

$$\mathrm{DH} \triangleq \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{D},\, k \in \mathcal{K}} z_{jk}^m \, d_{jO_k} \,. \tag{1}$$

We have the following definitional relations between the e-HSP demand variables Q_k^m and the total known demand Q_k:

$$Q_k = \sum_{m \in \mathcal{M}_+} Q_k^m = \underbrace{\sum_{m \in \mathcal{M}} Q_k^m}_{\substack{\text{OD demands served} \\ \text{by e-HSP drivers}}} + \underbrace{Q_k^0}_{\substack{\text{trips made by} \\ \text{solo drivers}}} \,. \tag{2}$$

Derived model variables. Besides the link flow variables, there are two types of derived variables. For $m > 0$, the variables R_{jk}^m, U_k^m, u_k, and w_k^m are employed in the modules describing the e-HSP choice module, the customer choice model, and the "market clearance" conditions, respectively; the remaining variables ϕ_j^m, λ_k^m, and μ_k are dual variables of the model constraints with the latter two having the interpretation of marginal prices. Specifically, we have

x_a	flow on link $a \in \mathcal{A}$
R_{jk}^m	profit (in \$) of e-HSP$_m$ vehicles that are at destination j and plan to serve customers of OD pair k
U_k^m	disutility (\$ per trip) of e-HSP$_m$ customers of OD pair k
u_k	minimum disutility (\$ per trip) of e-HSP customers of OD pair k across all modes
w_k^m	waiting cost (\$ per trip) of customers of OD pair k to be served by e-HSP$_m$ that is due to the e-HSP vehicle driving to pick up the trip demand
$\widehat{w}_k^m$	waiting time of e-HSP$_m$ vehicles serving customers of OD pair k
ϕ_j^m, $\widehat{\phi}_j^m$	shadow price of the e-HSP vehicles balancing constraint at location $j \in \mathcal{D}$ perceived by the e-HSP$_m$ drivers and their customers, respectively
λ_k^m	marginal price of OD demand k perceived by e-HSP$_m$; under the postulate of a normalized equilibrium, this is proportional to $\widehat{\lambda}_k^m$
$\widehat{\lambda}_k^m$	marginal price perceived by customers of OD k using e-HSP$_m$; it can be interpreted as a model induced cost of matching between customers and e-HSP vehicles

We classify the link flows as derived variables because we will employ a path-flow formulation for the network congestion component of the overall model. The connection between the path flow variables h_p

and the arc flow variables x_a is standard; namely,

$$x_a = \sum_{p \in \mathcal{P} \,:\, p \text{ uses } a} h_p, \quad \text{for all links } a \in \mathcal{A}. \tag{3}$$

Thus, the link flows are derived from the path flows that we consider as primary variables.

Model functions. These are as follows:

$C_p(h)$	cost of path $p \in \mathcal{P}$, a function of the path flow vector h
$c_a(x)$	travel time of link $a \in \mathcal{A}$, a function of link flow vector $x \triangleq (x_a)_{a \in \mathcal{A}}$; an example of such a link cost is the BPR type function.

In the additive model we have $C_p(h) = \sum_{a \,:\, p \text{ uses } a} c_a(x)$. Introducing the link-path matrix Ψ with entries

$$\psi_{ap} \triangleq \left\{ \begin{array}{ll} 1 & \text{if path } p \text{ uses link } a \\ 0 & \text{otherwise} \end{array} \right\} \quad \text{for all } (a,p) \in \mathcal{A} \times \mathcal{P}$$

and letting $C(h) \triangleq (C_p(h))_{p \in \mathcal{P}}$ and $c(x) \triangleq (c_a(x))_{a \in \mathcal{A}}$ be the respective vector functions of path costs and link costs, we have the following compact relation between these two functions:

$$C(h) = \Psi^T c(\Psi h).$$

More generally, non-additive path costs may also be used [11]. Analogous to the link-path matrix Ψ, define the node-path matrix Ω with entries

$$\omega_{(ij),p} \triangleq \left\{ \begin{array}{ll} 1 & \text{if } p \in \mathcal{P}_{ij} \\ 0 & \text{otherwise} \end{array} \right\} \quad \text{for all } (i,j,p) \in \mathcal{N} \times \mathcal{N} \times \mathcal{P}.$$

This matrix Ω comes in handy when connecting the travel times t_{ij} with the path costs $C_p(h)$.

The free-flow travel times f^0_{ij} is by definition the minimum travel costs $C_p(0)$ over all paths connecting nodes i to j; i.e.,

$$f^0_{ij} = \min_{p \in \mathcal{P}_{ij}} C_p(0) \leq \min_{p \in \mathcal{P}_{ij}} C_p(h), \quad \text{for any } h \geq 0, \tag{4}$$

where the inequality holds in the additive model provided that each link cost function satisfies the natural condition: $c_a(x) \geq c_a(0)$ for all $x \geq 0$. We also assume $c_a(0) > 0$ and $C_p(0) > 0$, i.e., the free flow link or path travel time is strictly positive, which is generally true. In addition, since our development does not depend on the additivity of the path costs as the sum of the link costs, we will work with the abstract functions $C_p(h)$ and postulate that the above lower bound condition (4) holds for all pairs $(i,j) \in \mathcal{N} \times \mathcal{N}$.

We also need a couple of auxiliary variables to avoid the un-defined ratio 0/0; see Section 4.1. Mathematically, these variables allow us to define this ambiguous ratio as any number in the interval $[0,1]$.

Auxiliary variables. These are

θ^m_{jk}, ζ^m_{jk}	artificial variables employed to handle the ambiguity of the undefined ratio 0/0 in the calculation of the customers' waiting costs; see Section 4.1.

3 e-HSP Choice Module

The e-HSP choice module focuses on the optimization of an e-HSP transportation provider. As previously mentioned, we assume that drivers employed by an e-HSP are cooperative towards the company's objective. We further assume a given number of vehicles of an e-HSP, denoted as N^m for e-HSP type m. Since we focus on an one-hour study period, the total e-HSP$_m$ vehicle service time is $N^m * 1 = N^m$ vehicle-hours. This is used in the following equation (7) to balance e-HSP vehicle's occupied times, vacant times, and waiting times (denoted as $\widehat{w}_k^m$). The main objective of e-HSP$_m$ is to maximize its profit given N^m. We model different e-HSPs with similar revenue cost structures but with different parameters. In essence, an e-HSP's profit is the difference between revenue and costs. Since we integrate e-HSP and its drivers as one entity in this paper, the revenue and costs here also reflect some of the e-HSP drivers' perspective. For example, the costs include time-based costs, distance-based costs (such as gas, vehicle depreciation, etc.), and waiting costs of drivers. The revenue includes a fixed charge and time-based and distance-based charges. Other costs of the e-HSP that do not depend on the dispatch decisions are not included here. Furthermore, in practice, an e-HSP platform charges a commission fee from drivers to use the platform (around 20%). Nevertheless, this does not need to be modeled here as e-HSP and its drivers are modeled as one entity.

In summary, the per-customer (or per-pickup) profit of an e-HSP trip, for $m \in \mathcal{M}$, at location j who plans to serve OD pair k can be modeled as:

$$R_{jk}^m \triangleq \widehat{R}_{jk}^m - \beta_3^m \, \widehat{w}_k^m, \tag{5}$$

where

$$\widehat{R}_{jk}^m = F_{\mathrm{O}_k}^m - \underbrace{\beta_1^m \left(t_{j\mathrm{O}_k} + t_{\mathrm{O}_k\mathrm{D}_k} \right)}_{\text{travel time based cost}} - \underbrace{\beta_2^m \left(d_{j\mathrm{O}_k} + d_{\mathrm{O}_k\mathrm{D}_k} \right)}_{\text{travel distance based cost}} + \underbrace{\alpha_1^m \left(t_{\mathrm{O}_k\mathrm{D}_k} - f_{\mathrm{O}_k\mathrm{D}_k}^0 \right)}_{\text{time based revenue}} + \underbrace{\alpha_2^m \, d_{\mathrm{O}_k\mathrm{D}_k}}_{\text{distance based revenue}}. \tag{6}$$

We model each e-HSP type $m \in \mathcal{M}$ as a revenue taker; that is, anticipating the net profit R_{jk}^m, for all pairs (j,k) and also the demands Q_k^m as exogenous to the decision making, e-HSP$_m$ type (for $m > 0$) calculates the vehicular flows z_{jk}^m by solving a simple profit maximization problem subject to feasibility constraints. In terms of these decision variables and the pick-up travel times t_{jO_k} and OD-travel times $t_{O_kD_k}$ the e-HSP waiting times $\widehat{w}_k^m$ satisfy the constraints [31]:

$$\sum_{k\in\mathcal{K}} \left[\sum_{j\in\mathcal{D}} z_{jk}^m \right] \widehat{w}_k^m = N^m - \sum_{k\in\mathcal{K}} \sum_{j\in\mathcal{D}} z_{jk}^m t_{jO_k} - \sum_{k\in\mathcal{K}} Q_k^m t_{O_kD_k}, \quad \widehat{w}_k^m \geq 0. \tag{7}$$

Here the left hand side of the equation is the total waiting times of all e-HSP$_m$ vehicles, the second term on the right hand side is the total vacant times (i.e., traveling to pick up customers), and the third term on the right side is the total occupied times (i.e., serving customers). In the e-HSP choice module presented below, the variables $\widehat{w}_k^m$ are substituted out by the above vehicle-hours balancing constraint, resulting in an optimization problem for an e-HSP type in terms of the primary e-HSP vehicular variables

z_{jk}^m parameterized by the travel demands and travel times. Here we model the e-HSP vehicle waiting time $\widehat{w}_k^m$ endogenously in the model, similarly to how the taxi vehicle waiting times were modeled in [31].

Notice that there is no sign convention on the profit rates R_{jk}^m which are linear functions of the unknown travel times in the e-HSP. In particular, it is possible for some of these rates to be negative. However, it seems reasonable to expect that for the travel times determined from the network congestion model to be introduced in Section 5, at least one such rate should be nonnegative for each e-HSP mode m. Interestingly, such nonnegativity of the profits is not needed to guarantee the existence of an equilibrium solution to the overall TGEM-ets.

e-HSP's choice module of profit maximization: for $m \in \mathcal{M}$,

$$\begin{array}{lll}
\underset{z_{jk}^m \geq 0}{\text{maximize}} & \underbrace{\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} \widehat{R}_{jk}^m \, z_{jk}^m}_{\text{average trip profit}} - \beta_3^m \left[N^m - \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^m t_{jO_k} - \sum_{k \in \mathcal{K}} Q_k^m \, t_{O_k D_k} \right] & \\
\text{subject to} & \underbrace{\sum_{k \in \mathcal{K}} z_{jk}^m = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m}_{\text{available e-HSP vehicles for service}} & \text{for all } j \in \mathcal{D} \\
& \underbrace{\sum_{j \in \mathcal{D}} z_{jk}^m \geq Q_k^m}_{\text{OD demands served by e-HSP}_m} & \text{for all } k \in \mathcal{K} \\
\text{and} & \underbrace{\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^m \, t_{jO_k}}_{\substack{\text{trip hours of vehicles} \\ \text{en route to service calls}}} + \underbrace{\sum_{k \in \mathcal{K}} Q_k^m \, t_{O_k D_k}}_{\substack{\text{trip hours of vehicles} \\ \text{serving travel demands}}} \leq N^m. &
\end{array} \tag{8}$$

The objective is to maximize the expected profit of e-HSP$_m$ responding to service calls of all OD-pairs. The decision variables are z_{jk}^m, i.e., the number of e-HSP$_m$ vehicles at destination j serving customers of OD pair k for the study period (one-hour in this paper). Being an e-HSP$_m$ vehicle supply equation, the constraint

$$\sum_{k \in \mathcal{K}} z_{jk}^m = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m, \quad \forall \, (j, m) \in \mathcal{D} \times \mathcal{M} \tag{9}$$

means that, in a steady-state of equilibrium, the number of available e-HSP$_m$ vehicles at location j (the left-hand sum) should be equal to the total number of customers whose destination is that location (the right-hand sum); see the similar equation in [31] for street-hailing taxi services. These available e-HSP vehicles may decide to stay at j possibly to service another OD pair at that location or go to another destination to serve another OD pair. Called e-HSP$_m$ demand constraint, the inequality

$$\sum_{j \in \mathcal{D}} z_{jk}^m \geq Q_k^m, \quad (k, m) \in \mathcal{K} \times \mathcal{M} \tag{10}$$

ensures that there are enough e-HSP$_m$ vehicles to meet the OD demands that request their service. These vehicles can choose not to serve customers of a particular OD pair, in which case both the right-hand side and the left-hand side of the constraint will be zero. The third and last constraint upper bounds the amount of variable vehicle-hours by the given amount N^m. Note that, as discussed earlier, the above e-HSP choice module only applies to normal e-HSPs (i.e., $m \in M$) and not for solo driving (i.e., $m = 0$).

A distinguished (and key) feature of the individual optimization problems in a non-cooperative game (e.g., the TGEM-ets model proposed in this paper) is that each player is optimizing his/her own decision variables while treating the decision variables of the other players in the overall model as exogenous (or given); in other words, a player can only anticipate but not control the other players's decision in such a game. More detailed discussions on this aspect can be found in [6]. For the e-HSP choice model (8) presented here, z_{jk}^m is the decision variable, i.e., whether and where to pick up the next customer. The OD demands Q_k^m, and travel times t_{jO_k} and $t_{O_k D_k}$ are all exogenous to the above e-HSP module, which are determined in the customer choice module and the network congestion module, respectively, to be described below. The profits $\widehat{R}_{jk}^m$ are also exogenous and are determined by (8). As a result, let ϕ_j^m denote the un-signed dual variable of the e-HSP vehicle supply equation (9). Let λ_k^m denote the signed dual variable of the e-HSP demand constraint (10). Let μ^m denote the signed dual variable of the vehicle-hours constraint. Let $\perp$ denote perpendicularity which in this context describes the complementary slackness between the slack of a constraint and its dual variable. The optimality conditions of (8) are given by the following complementarity conditions:

$$\text{for } m \in \mathcal{M} \left\{ \begin{array}{llll} 0 \leq z_{jk}^m & \perp & -\widehat{R}_{jk}^m - \beta_3^m \, t_{jO_k} - \phi_j^m - \lambda_k^m + t_{jO_k} \, \mu^m \geq 0, & \forall \, (j, k) \in \mathcal{D} \times \mathcal{K} \\ \phi_j^m \text{ free}, & & \displaystyle\sum_{k \in \mathcal{K}} z_{jk}^m = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m, \quad \forall \, j \in \mathcal{D} & \\ 0 \leq \lambda_k^m & \perp & \displaystyle\sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m \geq 0, \quad \forall \, k \in \mathcal{K} & \\ 0 \leq \mu^m & \perp & N^m - \left[\displaystyle\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^m \, t_{jO_k} + \sum_{k \in \mathcal{K}} Q_k^m \, t_{O_k D_k} \right] \geq 0 & \end{array} \right\}. \tag{11}$$

To establish the existence of a solution of the overall TGEM-ets, a postulate about the dual variable of the third complementarity above, λ_k^m, will be made that connects it with the customer's matching cost with the e-HSP vehicles. More details on this are provided in Subsection 6.1.

4 Customer Choice Module

A traveler for an OD pair k can choose either to use an e-HSP service or to drive alone (i.e., solo driving). If s/he chooses to use e-HSP service, s/he (i.e., the customer) will need to wait for the e-HSP vehicle and pay the fare; if s/he decides to drive along, the cost of driving his/her own car will need to be paid.

These along with the travel times are part of the travel disutility. For an e-HSP customer, additional disutilities may include the inconvenience of riding an e-HSP such as security concerns, comfort level, loss of productivity, among others. We model an e-HSP$_m$ customer's disutility for $m \in \mathcal{M}$ as:

$$V_k^m = \underbrace{F_{\mathrm{O}_k}^m + \alpha_1^m \left(t_{\mathrm{O}_k\mathrm{D}_k} - f_{\mathrm{O}_k\mathrm{D}_k}^0 \right) + \alpha_2^m \, d_{\mathrm{O}_k\mathrm{D}_k}}_{\text{total fare}} + \underbrace{\gamma_1^m \, t_{\mathrm{O}_k\mathrm{D}_k}}_{\substack{\text{travel time} \\ \text{based disutility}}} + \underbrace{w_k^m}_{\substack{\text{disutility due to} \\ \text{waiting for e-HSPs'} \\ \text{travel to pick up}}} . \tag{12}$$

For a solo driver, the disutility can be expressed as:

$$V_k^0 = \underbrace{\gamma_1^0 \, t_{\mathrm{O}_k\mathrm{D}_k}}_{\substack{\text{travel time} \\ \text{based disutility}}} + \underbrace{\beta_2^0 \, d_{\mathrm{O}_k\mathrm{D}_k}}_{\substack{\text{distance} \\ \text{based disutility}}} . \tag{13}$$

Here the disutility is expressed in monetary values. The first three terms on the right-hand side of equation (12) are the total fare the customer will need to pay. They consist of the fixed portion of the fare, the extra fee for traveling due to congestion, and the normal distance based fee. The last two terms represent the travel time costs and waiting time costs anticipated by the customer. That is, they represent the value of time associated while traveling or waiting. The value of time in particular due to traveling is mode specific to represent cases where customers may feel less safe or inconvenienced in traveling in a particular mode, thus, increasing the cost for traveling for using that particular mode. The customer cost w_k^m for $m \in \mathcal{M}$ due to waiting will be discussed in Subsection 4.1. Notice that in the definition of V_k^m, the matching cost between customer and vehicles is not included; this cost will be recovered once the constraints of the customer choice mode are introduced; see the discussion later in this section. If solo driving is selected, the customer disutility includes both distance-based and travel time based costs, as shown in (13).

Note that both the travel times and distances should be based on the actual travelled paths in the e-HSP vehicle profit and customer disutility functions. In this paper, to simplify the model, we assume that the distances (d_{jO_k} and $d_{O_kD_k}$) and free-flow travel times ($f_{\mathrm{O}_k\mathrm{D}_k}^0$) are fixed as those of the minimum-free-flow-travel-time paths, while travel time variables ($t_{\mathrm{O}_k\mathrm{D}_k}$ and $t_{j\mathrm{O}_k}$) are based on the actually traveled paths. The model could easily be extended to relax the assumption that the distance and free-flow travel time variables in the e-HSP vehicle profit and customer disutility are based only on the actual shortest paths at equilibrium. For this, a path index needs to be added to the z variables, which however will significantly increase the problem size of the model for even relatively small networks. For example, the dimension of the z variables would then become the product of the number of destinations, OD pairs, modes, and paths, i.e., $|\mathcal{D}|$ times $|\mathcal{K}|$ times $|\mathcal{M}|$ times $|\mathcal{P}|$. In the well-known Sioux Falls network, we have $|\mathcal{D}| = 24$, $|\mathcal{K}| = 524$, and $|\mathcal{M}| = 3$. Even if we only consider 20 paths for each OD pair, the dimension of z would be 167,680. Since the Sioux Falls network is a small-size network in practice, the resulting dimension of this extended model is very large. Overall, our assumption represents a certain tradeoff between the model's ability to capturing realism and its computational efficiency using the applicable state-of-the-art solver (such as the PATH solver that we will employ in the numerical experiments).

An important feature of the disutility as defined in equations (12) and (13) is that it is exogenous to the customer choice decisions. This is because all the variables in (12) and (13) are determined in the e-HSP choice and congestion modules over which a customer has no control; thus the customer can only anticipate such a disutility and cannot decide on its value. Defined by (16), the customer waiting cost is part of the disutility in (12), which is also exogenous to customer choices.

In this paper, we model the choice behavior of each traveler (i.e., each customer since solo driving is treated as a special type of e-HSP) as to minimize his/her own disutility. As discussed above, the disutility of each customer is exogenous in the customer's choice decision. Also, while not aware of the TNC's allocations of vehicles to meet the trip demands, the customers anticipate these allocations as upper bounds of their demands. The customers are not concerned about the e-HSP vehicle balance constraints (9) and thus do not include them in the optimization. In summary, we model a customer's choice behavior to determine the travel demands Q_k^m for $(m,k) \in \mathcal{M}_+ \times \mathcal{K}$ by an aggregate disutility minimization problem with the disutilities V_k^m and the TNC's vehicle allocations z_{jk}^m as exogenous variables subject to the e-HSP demand constraints (10) and also a balancing equation of each trip demand.

Customer choice module:

$$
\begin{aligned}
\underset{Q_k^m \geq 0}{\text{minimize}} \quad & \sum_{m \in \mathcal{M}_+} \sum_{k \in \mathcal{K}} V_k^m \, Q_k^m \\
\text{subject to} \quad & \sum_{j \in \mathcal{D}} z_{jk}^m \geq Q_k^m, \quad \text{for all } (k, m) \in \mathcal{K} \times \mathcal{M} \\
& \sum_{m \in \mathcal{M}_+} Q_k^m = Q_k, \quad \text{for all } k \in \mathcal{K}.
\end{aligned}
\tag{14}
$$

It is not difficult to show that the above model (14) aggregated over OD pairs k and travel modes m is equivalent to minimizing each customer's disutility separated by OD pairs, as can be seen in the equivalent optimality formulation (15) below. Letting $\widehat{\lambda}_k^m$ be the multiplier of the e-HSP$_m$ demand constraint for OD pair k and u_k for $k \in \mathcal{K}$ be the multipliers of the equality constraints, the optimality conditions of the above linear program in Q_k^m can be cast as the following mixed complementarity conditions:

$$
\begin{aligned}
& 0 \leq Q_k^m \perp \underbrace{V_k^m + \widehat{\lambda}_k^m}_{\text{denoted } U_k^m} - u_k \geq 0 \quad && \forall (k, m) \in \mathcal{K} \times \mathcal{M}_+ \\
& 0 \leq \widehat{\lambda}_k^m \perp \sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m \geq 0, && \forall (k, m) \in \mathcal{K} \times \mathcal{M} \\
& u_k \text{ free}, \quad \sum_{m \in \mathcal{M}_+} Q_k^m = Q_k && \forall k \in \mathcal{K}.
\end{aligned}
\tag{15}
$$

The first and third constraints have a similar interpretation to the classical Wardrop's user equilibrium principle of route choice. In this context, this principle states that customer chooses e-HSP$_m$ for travel between OD pair k (i.e., when Q_k^m is positive for $m \in \mathcal{M}$) only if the associated combined disutility U_k^m is minimum across all e-HSP choices. This is analogous to the shortest-path route choice in a Wardrop user equilibrium model; see e.g. [7, Section 1.4.5]. Here $U_k^m = V_k^m + \widehat{\lambda}_k^m$ is the *total* disutility of e-HSP$_m$ customers. Being the dual variable of the middle complementarity constraint in (15), $\widehat{\lambda}_k^m$ is also known as the shadow price of the constraint in basic linear programming theory. This added disutility may be considered as a model induced cost of matching between customers and e-HSP vehicles. To understand this interpretation, we note that the difference $\sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m$ represents the surplus of e-HSP vehicles that plan to serve an OD pair k. When such surplus is zero, we have $\widehat{\lambda}_k^m \geq 0$ based on the complementarity. In this case, the customers' matching time with e-HSP vehicles can be positive or zero. This is consistent with the nonnegativity of $\widehat{\lambda}_k^m$. If the surplus is positive, we must have $\widehat{\lambda}_k^m = 0$ by complementarity. In this case, a customer's matching time is zero since there are more vehicles to serve the customers, which again is consistent with the value of $\widehat{\lambda}_k^m$. Therefore, we can relate the dual variable $\widehat{\lambda}_k^m$ as an economic surrogate of the customer's matching with e-HSP vehicles. To the best of the authors' knowledge, this use of the shadow price, which is an economic concept, as related to matching is new and one of the contributions of the modeling framework in this paper. The customer matching time modeled this way is internally determined by the model; it captures the essence of customer waiting by relating $\widehat{\lambda}_k^m$ with the vehicle surplus. In practice, the matching time is larger if there are more customers than the vehicles, and is smaller if the opposite is true. Notice that for planning purposes and under a stable state, the available e-HSP vehicles should be always larger than (or equal to) the number of customers they can serve, thus the e-HSP demand constraint, leading to the second complementarity above. In a recent study [38], customer's matching time was related to the ratio of the number of the service vehicles and the number of the customers, instead of using the vehicle surplus (the difference of the two) as we do here.

We remark that the same e-HSP demand constraints (10) appear in both (8) and (15). The variables therein play different roles in the respective optimization problems, with z_{jk}^m being the primary variables and Q_k^m being the anticipated variables in the former and with their roles reversed in the latter. It is important to point out that mathematically, the multipliers (or equivalently, dual variables) of these constraints may be different as they represent the marginal prices of the corresponding constraints per the respective minimizing agents. Subsequently, we will introduce a *normalized* relationship of these multipliers λ_k^m and $\widehat{\lambda}_k^m$ in proving an existence result of the overall equilibrium model. Another noteworthy remark is that while the e-HSP modules are separable according to the e-HSP types $m = 1, \cdots, M$, the customers' choice module decouples according to OD pairs $k \in \mathcal{K}$.

4.1 Customer waiting cost due to pick-up

The e-HSP customers' total waiting cost has two components: the first one w_k^m, appearing in the disutility V_k^m, is related to the travel time of the e-HSP vehicles to come from their previous waiting locations to the pick-up locations. The second one is the multiplier $\widehat{\lambda}_k^m$ of the corresponding trip constraint (10) in the customer choice module (14) which is interpreted as the matching cost, as explained above. We focus on the former travel time in this subsection, and model it as the average of the travel times of e-HSP vehicles from all possible locations to the origin of an OD pair. We then use the multiplicative factor γ_2^m to covert travel times to cost (disutility). This leads to the following expression for customers' waiting cost:

$$w_k^m \;=\; \gamma_2^m \,\frac{\displaystyle\sum_{j\in\mathcal{D}} z_{jk}^m\, t_{j\mathrm{O}_k}}{\displaystyle\sum_{j\in\mathcal{D}} z_{jk}^m}, \quad \text{for all } m = 1, \cdots, M. \tag{16}$$

In order to properly handle this fraction, which is not well defined when the denominator is equal to zero, we let θ_{jk}^m equal to this fraction when the denominator is not zero and equal to an arbitrary scalar in the interval $[\,0,1\,]$ when the denominator is zero. We can then define:

$$w_k^m \;=\; \gamma_2^m \sum_{j\in\mathcal{D}} \theta_{jk}^m\, t_{j\mathrm{O}_k}.$$

The advantage of the scalar θ_{jk}^m as defined is due to the following minimization description:

$$\theta_{jk}^m \,\in\, \operatorname*{argmin}_{\theta\,\in\,[\,0,1\,]} \left\{ -z_{jk}^m\,\theta + \frac{1}{2}\left[\,\sum_{j\,'\in\mathcal{D}} z_{j\,'k}^m\right]\theta^2 \right\}; \tag{17}$$

in turn, (17) is a simple univariable bounded-constrained convex quadratic program (on θ) that is characterized by the complementarity conditions:

$$\begin{aligned} 0 \,\le\, \theta_{jk}^m \quad &\perp \quad \left[\,\sum_{j\,'\in\mathcal{D}} z_{j\,'k}^m\right]\theta_{jk}^m - z_{jk}^m + \zeta_{jk}^m \,\ge\, 0 \\ 0 \le \zeta_{jk}^m \quad &\perp \quad 1 - \theta_{jk}^m \,\ge\, 0. \end{aligned} \tag{18}$$

Thus bypassing the derived variable w_k^m, we can express the disutility V_k^m as: for $m > 0$,

$$V_k^m \;=\; F_{\mathrm{O}_k}^m + \alpha_1^m\,(\,t_{\mathrm{O}_k\mathrm{D}_k} - f_{\mathrm{O}_k\mathrm{D}_k}^0\,) + \alpha_2^m\, d_{\mathrm{O}_k\mathrm{D}_k} + \gamma_1^m\, t_{\mathrm{O}_k\mathrm{D}_k} + \gamma_2^m \sum_{j\in\mathcal{D}} \theta_{jk}^m\, t_{j\mathrm{O}_k}, \tag{19}$$

where θ_{jk}^m along with ζ_{jk}^m satisfies (18).

As formulated above, the two sets of waiting times (costs), $\widehat{w}_k^m$ for the e-HSP vehicles, and $w_k^m + \lambda_k^m$ for the e-HSP customers, are not explicitly linked. They both can be recovered from the model solutions: the former from the vehicle-hours balance constraints (7) for each e-HSP type and the latter from the expression (16). In the case where one type of e-HSP vehicles represents taxis, this post-solution determination of the waiting times/costs is different from a certain Cobb-Douglas type relation discussed extensively in [30, 31, 32, 33]. We will address more about our modeling approach of waiting times in the context of this type of explicit connection in Appendix I.

5 Network Congestion Effects

In a transportation network with e-hailing services, two types of vehicular traffic contributes to the overall congestion: (i) the customers (including the solo drivers and e-HSP customers) for any OD pair k; that can be served by any mode (solo, e-HSPs); and (ii) the vacant e-HSP vehicles going from a node $j \in \mathcal{D}$ to the origin of the next customer. We assume that once customers make their mode choices (from the customer choice module) and the e-HSP drivers make their decisions about which OD pair to serve, drivers' (for any mode) route choice decision in the network will be solely to minimize the travel time. It is known in the literature of traffic equilibrium that the network congestion model can be formulated as a complementarity problem. The only distinctive feature here is that the trips of the (vacant) e-HSP vehicles picking up customers add to the flows, and thus congestion of the network. Such additional trips/demands appear in the first equation below that balances the flows and demands. The overall congestion module of the TGEM-ets is as follows:

$$\begin{array}{lll} \underbrace{\sum_{p \in \mathcal{P}_{ij}} h_p}_{\text{path flows from } i \text{ to } j} & = & \underbrace{\sum_{k \in \mathcal{K}} \delta_{ijk}^{\text{OD}} Q_k}_{\text{fixed demand}} + \underbrace{\sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\text{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m}_{\text{variable demand}}, \text{for all } (i,j) \in \mathcal{N} \times \mathcal{N}, \\ 0 \leq h_p & \perp & C_p(h) - t_{ij} \geq 0, \quad \text{for all } p \in \mathcal{P}_{ij}. \end{array} \tag{20}$$

The path flow equation for nodes $(i,j) \in \mathcal{N} \times \mathcal{N}$ implies that if neither i nor j is an origin or destination node of an OD pair, then all paths connecting i to j are neither OD-paths or e-HSP pick-up paths; these paths must have zero flows because the right-side of the equation corresponding to the pair of nodes i and j is equal to zero. For completeness throughout the paper, for a pair (i,j) such that neither i nor j is in $\mathcal{O} \cup \mathcal{D}$, we take $t_{ij} = 0$ and $h_p = 0$ for all paths $p \in \mathcal{P}_{ij}$. The complementarity condition in the above equation expresses the well-known Wardrop principle of driver behavior; i.e., vehicles will take the routes with minimum travel times. The flow-demand balancing equation takes into account two types of trips: the OD trips and the (vacant) e-HSP pick-up trips, both contributing to network congestion. Other than this extension, the above network congestion model is exactly that of a traffic equilibrium. As such, the following result can be proved which ensures the nonnegativity of the minimum travel times under a mild assumption of the path costs; cf. [7, Proposition 1.5.6].

Proposition 1. Suppose that the path costs $C_p(h)$ satisfy (4). Then (20) holds if and only if the following complementarity conditions hold,

$$\begin{array}{l} 0 \leq t_{ij} \perp \displaystyle\sum_{p \in \mathcal{P}_{ij}} h_p - \left[\sum_{k \in \mathcal{K}} \delta_{ijk}^{\text{OD}} Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\text{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \right] \geq 0, \text{for all } (i,j) \in \mathcal{N} \times \mathcal{N}, \\ 0 \leq h_p \perp C_p(h) - t_{ij} \geq 0, \quad \text{for all } p \in \mathcal{P}_{ij}. \end{array} \tag{21}$$

Proof. Since the free-flow times f_{ij}^0 are positive, (4) implies that each $C_p(h)$ is a positive function for all $p \in \mathcal{P}$. Hence, the following condition holds trivially for all $(i,j) \in \mathcal{N} \times \mathcal{N}$:

$$\left[\sum_{p \in \mathcal{P}_{ij}} h_p C_p(h) = 0, \quad h \geq 0 \right] \Rightarrow [\, h_p = 0 \; \forall \, p \in \mathcal{P}_{ij} \,]. \tag{22}$$

It suffices to show that (21) implies the first equation in (20). Assume not; then for some $(i,j) \in \mathcal{N} \times \mathcal{N}$, we have

$$\sum_{p \in \mathcal{P}_{ij}} h_p > \sum_{k \in \mathcal{K}} \delta^{\text{OD}}_{ijk}\, Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta^{\text{e-HSP}}_{ijk\ell} \sum_{m \in \mathcal{M}} z^m_{i\ell} \;\geq\; 0.$$

Thus $t_{ij} = 0$ by complementarity. Hence

$$\sum_{p \in \mathcal{P}_{ij}} h_p\, C_p(h) \;=\; 0.$$

By assumption, it follows that $h_p = 0$ for all $p \in \mathcal{P}_{ij}$. But this is a contradiction. □

For the analysis of the overall model, to be summarized in the next section, it would be convenient to use the complementarity formulation (21) to describe the network congestion, restricting the travel time variables t_{ij} to be nonnegative in particular.

Also note that in Assumption (c) in Section 2, we assume that e-HSP drivers are cooperative in the sense that they follow the dispatching orders from the platform regarding which customers to serve and which routes to take. In the current e-hailing services, to the best of our knowledge, a platform routes its vehicles based on the shortest-path type of routing, which mostly resembles the user equilibrium (UE) type of routing. This is why the network congestion module here is described as similarly to an (extended) UE model, as shown above. In the future, when a platform indeed decides to route its vehicles to minimize the total travel time of all the vehicles (i.e., following the system optimal principle for all the vehicles of the platform), the current congestion module needs to be modified to capture this by extending the module to a more general type (or a mixed equilibrium type with each e-HSP minimizing the total travel time of all its own vehicles). We expect that a similar mathematical analysis can be carried out for such an extension. We also note that our analysis in the next section does not require the path cost functions $C_p(h)$ to be additive; we only need these functions satisfy a very mild assumption; see Lemma 2. Furthermore, as it is known in the literature (see e.g. [7]), the above network congestion module is not necessarily the optimality condition of a minimization problem when the path cost structure a non-additive.

5.1 Bounds on primary variables

Since $Q_k > 0$, it follows that for each OD pair $k \in \mathcal{K}$ with origin node i and destination j (i.e., $i = O_k, j = D_k$), there must exist a path $\bar{p} \in \mathcal{P}_{ij}$ with $h_{\bar{p}} > 0$. Hence $t_{ij} = C_{\bar{p}}(h)$ by complementarity. Moreover, $t_{ij} \leq C_p(h)$ for any path $p \in \mathcal{P}_{ij}$. Hence,

$$t_{\text{O}_k\text{D}_k} \;=\; \min_{p \in \mathcal{P}_{ij}} C_p(h) \;\geq\; f^0_{\text{O}_k\text{D}_k}.$$

This inequality yields that the disutility U^m_k as given by (19) is always nonnegative provided that the travel times $t_{\text{O}_k\text{D}_k}$ are determined according to (20)

Next, we derive some upper bounds for the primary variables of the model. Since $\sum_{m \in \mathcal{M}_+} Q^m_k = Q_k$, it follows that $Q^m_k \leq Q_k$ for all $m \in \mathcal{M}_+$. Since $\sum_{k \in \mathcal{K}} z^m_{jk} = \sum_{k' \in \mathcal{K}} \varepsilon^{\text{d}}_{jk'}\, Q^m_{k'}$, it follows that $\sum_{k \in \mathcal{K}} z^m_{jk} \leq \sum_{k' \in \mathcal{K}} Q^m_{k'}$

for all $(j,m) \in \mathcal{D} \times \mathcal{M}$. Let

$$\overline{Q} > \sum_{k \in \mathcal{K}} Q_k \quad \text{and} \quad Q_{\max} > \max_{k \in \mathcal{K}} Q_k. \tag{23}$$

It follows from the path flow equation in (20) that for all $p' \in \mathcal{P}_{ij}$,

$$h_{p'} \leq \sum_{p \in \mathcal{P}_{ij}} h_p \leq \sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \leq c \sum_{k \in \mathcal{K}} Q_k$$

for some constant $c > 0$. Thus, for a constant $\bar{h} > c \sum_{k \in \mathcal{K}} Q_k$, we have

$$h_p \leq c \sum_{k \in \mathcal{K}} Q_k < \bar{h}, \quad \forall\, p \in \mathcal{P}. \tag{24}$$

Moreover, we have

$$t_{i'j'} \leq \max_{(i,j) \in \mathcal{N} \times \mathcal{N}} \max_{p \in \mathcal{P}_{ij}} \max_{0 \leq h \leq \bar{h}\mathbf{1}} C_p(h), \quad \text{for all } (i',j') \in \mathcal{N} \times \mathcal{N},$$

Let $\bar{t}$ be greater than the right-hand constant in the last bound. Similar to (24), we have the following bound on the travel times; namely,

$$t_{i'j'} \leq \max_{(i,j) \in \mathcal{N} \times \mathcal{N}} \max_{p \in \mathcal{P}_{ij}} \max_{0 \leq h \leq \bar{h}\mathbf{1}} C_p(h) < \bar{t}, \quad \text{for all } (i',j') \in \mathcal{N} \times \mathcal{N}. \tag{25}$$

Based on the two positive bounds $\bar{h}$ and $\bar{t}$ as derived above, we can show that a pair of path flows and travel times:

$$\mathbf{h} \triangleq \{ h_p : p \in \mathcal{P} \}, \quad \text{and} \quad \mathbf{t} \triangleq \{ t_{ij} : (i,j) \in \mathcal{N} \times \mathcal{N} \}$$

satisfies (20) if and only if it is a solution of a variational inequality defined on the bounded rectangle $\mathcal{H} \triangleq \left\{ (\mathbf{h}, \mathbf{t}) \geq 0 : \mathbf{h} \leq \bar{h}\,\mathbf{1}, \text{ and } \mathbf{t} \leq \bar{t}\,\mathbf{1} \right\}$. This assertion is similar to that employed to prove the existence of a solution to the classical traffic equilibrium problem; cf. [7, Proposition 2.2.14]. Rather than directly proving the assertion, we embed it in the proof of the existence of an equilibrium solution to the overall TGEM-ets. Similar to the tuples $\mathbf{h}$ and $\mathbf{t}$, we let

$$\mathbf{z} \triangleq \left\{ z_{jk}^m : (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \right\} \quad \text{and} \quad \mathbf{Q} \triangleq \{ Q_k^m : (k,m) \in \mathcal{K} \times \mathcal{M}_+ \},$$

and call them an *e-HSP vehicle profile* and a *OD demand profile*, respectively. Any such pair of vehicle and demand profiles *induces* a tuple of travel times $\mathbf{t}$ via the network congestion conditions (20), or equivalently their complementarity formulation (21).

6 The Overall Equilibrium Model

For ease of reference, we first summarize the three principal modules in the overall TGEM-ets:

- with profits given by (5), the e-HSP optimization problems (i.e., profit maximization) (8) for $m = 1, \cdots, M$, or their optimality conditions (11); details are provided in Section 3;

• with waiting costs given by (16) and disutilities given by (19), the customer choice optimization module (14), or its equivalent complementarity formulation (15); details are provided in Section 4;

• network congestion modeled by the equilibrium conditions (21); details are provided in Section 5.

As shown in Section 3, the e-HSP$_m$ waiting times $\widehat{w}_k^m$ can be recovered from expression (7) after a TGEM-ets equilibrium solution is obtained. More about this recovery and its relation to the customer waiting costs will be provided in Appendix I.

TGEM-ets Equilibrium. By definition, this is a tuple $\{\, \mathbf{z}, \mathbf{Q}, \mathbf{h}, \mathbf{t} \,\}$ for which there exist

$$\boldsymbol{\theta} \triangleq \left\{\, \theta_{jk}^m \,:\, (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \,\right\},$$

such that for each $m \in \mathcal{M}$, the tuple $\left\{\, z_{jk}^m \,:\, (j,k) \in \mathcal{D} \times \mathcal{K} \,\right\}$ is an optimal solution of (8) with each $\widehat{R}_{jk}^m$ given by (6); the tuple $\{\, Q_k^m \,:\, (k,m) \in \mathcal{K} \times \mathcal{M}_+ \,\}$ is an optimal solution of (14) with V_k^m given by (19); $(\mathbf{h}, \mathbf{t})$ satisfies (20), or equivalently (21); and θ_{jk}^m satisfies (17) for all $(j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M}$. $\square$

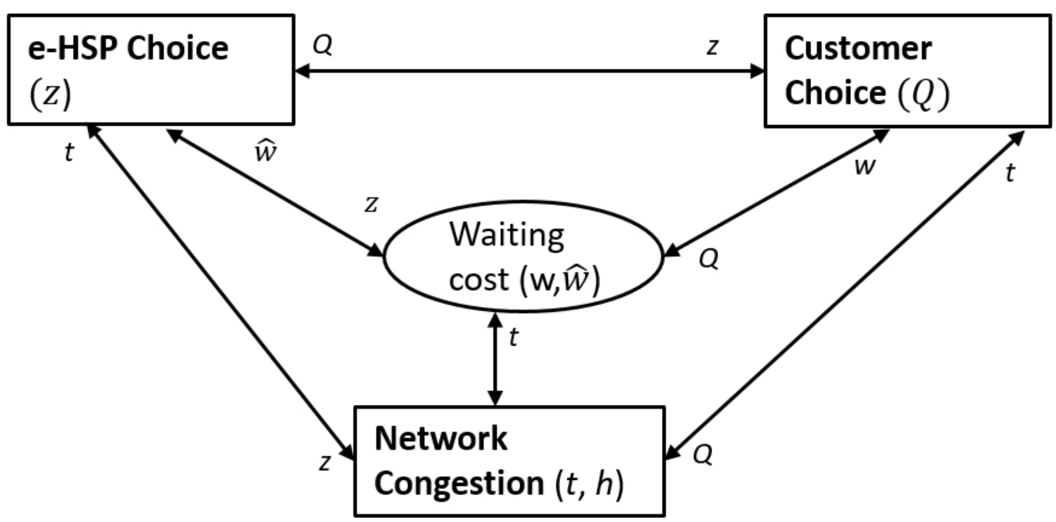

Figure 1: Interactions of major model components of TGEM-ets

To better illustrate the interactions among the major model components, Figure 1 provides a graphical depiction of the connections among these modules (e-HSP choice, customer choice, and network congestion) and the waiting cost component. The variables in the parenthesis in each module/component indicate the main decision variables (i.e., endogenously determined) of that particular component. For example, the congestion module will determine the minimum route travel times t and the route flow h. The links show the connections among the four components. Each link has two arrows. The variables at an arrow of a link indicate the exogenous variables to the module next to the arrow, which are derived from the module in the other end of the link. For example, the revenue R_{jk}^m receives $\mathbf{Q}$ from the customer's choice module; the disutility V_k^m receives $\mathbf{z}$ from the e-HSP choice module; both R_{jk}^m and V_k^m receive $\mathbf{t}$ from the congestion module.

6.1 A preview of the proof

The demonstration of the existence of an equilibrium solution to the proposed TGEM-ets model is a non-trivial task; in fact, this is a major contribution of our work along with the formulation of the overall model for a highly complex e-hailing transportation system. We accomplish this task via a sequence of steps that we will outline momentarily. Before doing so, we highlight the technical challenges in the proof from several perspectives, starting from the simplified situation of ignoring the network congestion effects. Removing this network module in the model implies that we take the travel times t_{jO_k} and $t_{O_kD_k}$ as given constants independent of the traffic flows. In addition to being a significant simplification of reality, this maneuver, along with other simplifications, ease the proof of existence of an equilibrium somewhat but other important bottlenecks persist.

Without the network congestion module, the formulated equilibrium model may be considered a non-cooperative multi-agent optimization problem with the principal decision makers being the e-HSP acting as the supply side of the system and deciding which service calls to respond to, and the customers deciding between the e-hailing modes or solo driving. Each decision making agent can only anticipate the decisions of the other agents and take them as exogenous to their optimization problems. Due to the fact that the constraints of each e-HSP's optimization problem (8) contains the customers' variables Q_k^m which are decision variables in (14), and to the presence of the same OD-demand constraints (10) in both the e-HSPs' (8) and customer's (14) problems, the resulting non-cooperative game is of the generalized type [5, 6] with share constraints. Unlike a standard game where each agent's optimization problem has only "private" constraints, i.e., these constraints contain only the agent's own variables, a generalized Nash game is much more challenging to treat analytically (weaker theory) and computationally (less guarantee of consistent solvability). In particular, the (sufficient) conditions for existence of an equilibrium solution currently existing in the literature are typically quite restrictive and not easy to satisfy in applied formulations such as the one we have on hand.

With the inclusion of the network congestion module, the overall equilibrium model may continue to be considered as a generalized non-cooperative game with one additional player whose objective is to equilibrate the traffic flows, i.e., as a game with side constraints [29], or as a "Multiple Optimization Problem with Equilibrium Constraints" (MOPEC) as it is termed in [9]. For games of this kind, the general approach to establish the existence of an equilibrium solution is to cast the problem as a variational inequality or complementarity problem (VI/CP) [7] and apply the theory for these problems. For the model that we have on hand, we need to be careful about the following features that make this general theory and a standard fixed-point approach not readily applicable:

- the generalized nature of the game as mentioned above and the challenges associated with the multiplicative coupling of variables in the constraints;
- adding to the nonconvex coupling of variables is the lack of explicit bounds on some of the model

variables; both deficiencies make the basic requirements of a self-map from a compact convex set into itself stipulated in a fixed-point theorem difficult to satisfy;
• the lack of required monotonicity or other favorable properties of the defining function in the resulting VI/CP formulation of the model (see below) that would enable a direct application of existence results available in the literature.

For all these reasons, we provide a detailed proof of the existence of an equilibrium solution of the e-hailing model via the novel idea of penalization embedded in the theory of variational inequalities. Readers who are not interested in such proof can skip the next section and proceed directly to the Section 8 on numerical results. In the proof, we focus on a special type of equilibrium known as a *normalized equilibrium* that was proposed by Rosen [19] in his classic work of generalized Nash games with coupled shared constraints. Specifically, a normalized equilibrium of the TGEM-ets postulates that the multipliers corresponding to a common coupled constraint are proportional and the proportionality constant is dependent on the players only. In this context, this requirement means that there exist positive constants, which we denote γ_3^m such that $\widetilde{\lambda}_k^m = \gamma_3^m \lambda_k^m$ for all pairs $(k,m) \in \mathcal{K} \times \mathcal{M}$. Our main existence result, Theorem 1, asserts the existence of a normalized TGEM-ets equilibrium for any positive scalars $\{\gamma_3^m\}_{m \in \mathcal{M}}$ under the specified assumptions. It is important to point out that the existence result in [19] cannot be applied to the TGEM-ets because of the reasons given above. For more discussion about multipliers of shared constraints in generalized Nash games, see the recent article [21] and the references therein.

7 Existence of a Normalized Equilibirum

There are several main steps in the proof. First, we obtain a mixed complementarity problem (MCP) formulation of the model in Subsection 7.1. Next, we highlight the complicating constraints in the formulation whose relaxation significantly simplifies the analysis. The way we obtain such a relaxation is via a penalty of the violation of such constraints; see Subsection 7.2. Relying on a feasibility assumption and a couple key lemmas (Subsection 7.4), the last step in the proof consists of showing that a solution of the overall model can be recovered by letting the penalty tends to infinity; see Subsection 7.3. Combined together, these steps constitute an innovative approach to show the existence of a solution to a rather complex equilibrium model that is not amenable to direct analysis by standard methods.

In terms of specific results, there are 4 lemmas and the main existence Theorem 1. Lemma 1 pertains to Step 1 that results in the penalized variational formulation of the model; the lemma asserts the existence of a solution to this penalized variational inequality. This is followed by Lemma 2 that asserts a particular property of a solution of the variational model; this lemma is very similar to a classic result in the complementarity formulation of the standard traffic equilibrium problem (see e.g. [7, Proposition 1.4.6]). Lemma 3 introduces a feasibility assumption that is needed for the final existence problem and under which two preliminary limiting properties of the penalized variational problem as the penalty parameter goes to infinity are derived. Lemma 4 is a technical result from linear complementarity theory that enables the completion of the proof of the main existence Theorem 1 via Lemma 3. We next present the details of the proof.

7.1 First step: The complementarity formulation

As outlined above, the proof of existence of a TGEM-ets equilibrium begins with the formulation of such an equilibrium as a solution of the following MCP:

$$
\begin{array}{lll}
0 \leq z_{jk}^m \quad \perp \quad -\widehat{R}_{jk}^m - \beta_3^m \, t_{jO_k} - \phi_j^m - \lambda_k^m + t_{jO_k} \, \mu^m \, \geq \, 0 & & \forall \, (\, j, k, m \,) \, \in \, \mathcal{D} \times \mathcal{K} \times \mathcal{M} \\
\phi_j^m \text{ free} \qquad \displaystyle\sum_{k \in \mathcal{K}} z_{jk}^m \, = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m & & \forall \, (\, j, m \,) \, \in \, \mathcal{D} \times \mathcal{M} \\
0 \leq \lambda_k^m \quad \perp \quad \displaystyle\sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m \, \geq \, 0 & & \forall \, (\, k, m \,) \, \in \, \mathcal{K} \times \mathcal{M} \\
0 \leq \mu^m \quad \perp \quad N^m - \left[\displaystyle\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^m \, t_{jO_k} + \sum_{k \in \mathcal{K}} Q_k^m \, t_{O_k D_k} \right] \geq \, 0 & & \forall \, m \, \in \, \mathcal{M} \\
0 \leq Q_k^0 \quad \perp \quad V_k^0 - u_k \, \geq \, 0 & & \\
0 \leq Q_k^m \quad \perp \quad V_k^m - u_k + \underbrace{\gamma_3^m \, \lambda_k^m}_{\text{this is the normalized postulate}} \, \geq \, 0 & & \forall \, (\, k, m \,) \, \in \, \mathcal{K} \times \mathcal{M} \\
u_k \text{ free,} \qquad \displaystyle\sum_{m \in \mathcal{M}_+} Q_k^m \, = \, Q_k & & \forall \, k \, \in \, \mathcal{K} \\
0 \leq t_{ij} \quad \perp \quad \displaystyle\sum_{p \in \mathcal{P}_{ij}} h_p - \left[\sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \right] \geq \, 0 & & \forall \, (\, i, j \,) \, \in \, \mathcal{N} \times \mathcal{N} \\
0 \leq h_p \quad \perp \quad C_p(h) - t_{ij} \, \geq \, 0 & & \forall \, p \, \in \, \mathcal{P}_{ij}, \; \forall \, (\, i, j \,) \, \in \, \mathcal{N} \times \mathcal{N} \\
0 \leq \theta_{jk}^m \quad \perp \quad \left[\displaystyle\sum_{j' \in \mathcal{D}} z_{j'k}^m \right] \theta_{jk}^m - z_{jk}^m + \zeta_{jk}^m \, \geq \, 0 & & \forall \, (\, j, k, m \,) \, \in \, \mathcal{D} \times \mathcal{K} \times \mathcal{N} \\
0 \leq \zeta_{jk}^m \quad \perp \quad 1 - \theta_{jk}^m \, \geq \, 0 & & \forall \, (\, j, k, m \,) \, \in \, \mathcal{D} \times \mathcal{K} \times \mathcal{M}.
\end{array}
$$

Note that the above formulation is directly from the e-HSP choice (11), customer choice (15), waiting costs (18), (19), and network congestion module (21). There are two main complicating factors in the above complementarity formulation: one is the bilinear (thus nonlinear) term $t_{jO_k} \mu^m$ in the z-complementarity condition; the second is the lack of symmetry of the multipliers ϕ_j^m and μ^m which are present in the z_{jk}^m-complementarity condition but not in the Q_k^m-complementarity condition. The latter feature is the result of the generalized nature of the game; i.e., the coupling constraints

$$
\sum_{k \in \mathcal{K}} z_{jk}^m \, = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m \quad \text{and} \quad N^m - \left[\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^m \, t_{jO_k} + \sum_{k \in \mathcal{K}} Q_k^m \, t_{O_k D_k} \right] \geq \, 0
$$

appear only in the e-HSP choice module but not in the customer choice module. To deal with these complications caused by the coupling constraints, we penalize their violations using a positive parameter

ρ. Note that while the variables z_{jk}^m and Q_k^m are also coupled in the e-HSP demand constraints (10), they can be kept in the penalized VI formulation below by means of a scaling maneuver.

7.2 Second step: the VI with penalization of the complicating constraints

For a scalar α, let $\alpha_+ \triangleq \max(0, \alpha)$ be its nonnegative part. Define, for each $\rho > 0$, the vector function (note the multiplicative factor γ_3^m in front of R_{jk}^m which was not in the original complementarity formulation; this is another manipulation that is needed in the proof):

$$
\mathbf{F}^\rho(\mathbf{w}) \triangleq
\left(\begin{array}{lcl}
-\gamma_3^m \left(\widehat{R}_{jk}^m + \beta_3^m \, t_{jO_k} \right) + & : & (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \\
\underbrace{\rho \left[\left(\displaystyle\sum_{\ell \in \mathcal{K}} z_{j\ell}^m - \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m \right) + t_{jO_k}^m \left[\displaystyle\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^m \, t_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q_{k'}^m \, t_{O_{k'}D_{k'}} - N^m \right]_+ \right]}_{\text{penalty term}} & & \\
\hline
V_k^m & : & (k,m) \in \mathcal{K} \times \mathcal{M}_+ \\
\hline
C_p(h) - t_{ij} & : & p \in \mathcal{P}_{ij}, \ (i,j) \in \mathcal{N} \times \mathcal{N} \\
\hline
\displaystyle\sum_{p \in \mathcal{P}_{ij}} h_p - \sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k - \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \, Q_k^m & : & (i,j) \in \mathcal{N} \times \mathcal{N} \\
\hline
\left[\displaystyle\sum_{j' \in \mathcal{D}} z_{j'k}^m \right] \theta_{jk}^m - z_{jk}^m & : & (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M}
\end{array}\right)
$$

and the polyhedron $\mathcal{W}$ consisting of tuples $\mathbf{w} \triangleq (\mathbf{z}, \mathbf{Q}, \mathbf{h}, \mathbf{t}, \boldsymbol{\theta})$ with $(\mathbf{h}, \mathbf{t}) \in \mathcal{H}$ and $(\mathbf{z}, \mathbf{Q}, \boldsymbol{\theta}) \geq 0$ satisfying:

$$
\begin{array}{rclll}
\displaystyle\sum_{k \in \mathcal{K}} z_{jk}^m & \leq & \underbrace{\displaystyle\sum_{k' \in \mathcal{K}} Q_{k'}}_{\text{model constant}} & \forall \, (j,m) \in \mathcal{D} \times \mathcal{M} & (\eta_j^m) \\
\displaystyle\sum_{j \in \mathcal{D}} z_{jk}^m & \geq & Q_k^m & \forall \, (k,m) \in \mathcal{K} \times \mathcal{M} & (\tilde{\lambda}_{jk}^m) \\
\displaystyle\sum_{m \in \mathcal{M}_+} Q_k^m & = & Q_k & \forall \, k \in \mathcal{K} & (u_k) \\
\theta_{jk}^m & \leq & 1 & \forall \, (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} & (\zeta_{jk}^m).
\end{array} \tag{26}
$$

Note that in the set $\mathcal{W}$, the only coupling between the e-HSP service variables z_{jk}^m and e-HSP demand variables Q_k^m is through the e-HSP demand constraints; the e-HSP vehicle supply equations and the vehicle-hours bound constraints are transferred to the function $\mathbf{F}^\rho$ via a penalization. It is worthwhile to note that $\mathcal{W}$ contains upper bounds of the variables $(\mathbf{h}, \mathbf{t})$, whereas such bounds do not appear in the complementarity formulation of the original TGEM-ets model (see Subsection 7.1). Including the bounds in $\mathcal{W}$ renders this set compact; in turn, this compactness leads to Lemma 1 below. To restore the formulation of the original model, we need to show that these bounds are redundant; a precise result is presented in Lemma 2. The first constraint in (26) is a relaxation of the former omitted e-HSP$_m$ vehicle supply equation and is included in $\mathcal{W}$ to ensure the boundedness of the z_{jk}^m variables. Also the functions corresponding to λ, u, ζ do not appear in the definition of $\mathbf{F}^\rho$ since they will be treated as (and thus recovered from) multipliers of the constraints in set $\mathcal{W}$. This will become clear when we write the complementarity condition of VI $(\mathbf{F}^\rho; \mathcal{W})$ in the following (in Lemma 1). Our goal is to show that the omitted constraints will be satisfied in the limit when the penalty parameter ρ tends to ∞, therefore recovering a solution to the full model. To accomplish this goal, we first state the following existence result for the penalized VI defined by the pair $(\mathbf{F}^\rho; \mathcal{W})$. The result is immediate since $\mathcal{W}$ is a compact polyhedron; no proof is needed.

Lemma 1. If the path cost functions $C_p(h)$ are continuous, and $\mathcal{W}$ is nonempty, then the VI $(\mathbf{F}^\rho; \mathcal{W})$ has a solution for every $\rho > 0$. $\square$

Introducing multipliers s_{ij} and v_p for the upper bound constraints $t_{ij} \leq \bar{t}$ and $h_p \leq \bar{h}$, we can write the

MCP of the VI $(\mathbf{F}^{\rho}; \mathcal{W})$ as follows:

$$\begin{array}{lll}
0 \,\leq\, z_{jk}^m \;\perp\; -\gamma_3^m \left(\widehat{R}_{jk}^m + \beta_3^m \, t_{jO_k} \right) - \widetilde{\lambda}_k^m + \eta_j^m + & & \forall\,(\,j,k,m\,) \,\in\, \mathcal{D}\times\mathcal{K}\times\mathcal{M} \\
\qquad \rho \left[\begin{array}{l} \left(\displaystyle\sum_{\ell\in\mathcal{K}} z_{j\ell}^m - \sum_{k\,':\,j=\mathrm{D}_{k\,'}} Q_{k\,'}^m \right) + \\ t_{jO_k}^m \left[\displaystyle\sum_{k\,'\in\mathcal{K}} \sum_{j\,'\in\mathcal{D}} z_{j\,'k\,'}^m \, t_{j\,'O_{k\,'}} + \sum_{k\,'\in\mathcal{K}} Q_{k'}^m \, t_{O_{k\,'},D_{k\,'}} - N^m \right]_+ \end{array} \right] & \geq 0 & \\
0 \,\leq\, \eta_j^m \;\perp\; \displaystyle\sum_{k\,'\in\mathcal{K}} Q_{k\,'} - \sum_{k\in\mathcal{K}} z_{jk}^m \,\geq\, 0 & & \forall\,(\,j,k,m\,) \,\in\, \mathcal{D}\times\mathcal{K}\times\mathcal{M} \\
0 \,\leq\, \widetilde{\lambda}_k^m \;\perp\; \displaystyle\sum_{j\in\mathcal{D}} z_{jk}^m - Q_k^m \,\geq\, 0 & & \forall\,(\,k,m\,) \,\in\, \mathcal{K}\times\mathcal{M} \\
0 \,\leq\, Q_k^0 \;\perp\; V_k^0 - u_k \,\geq\, 0 & & \\
0 \,\leq\, Q_k^m \;\perp\; V_k^m - u_k + \widetilde{\lambda}_k^m \,\geq\, 0 & & \forall\,(\,k,m\,) \,\in\, \mathcal{K}\times\mathcal{M} \\
u_k \text{ free,} \quad \displaystyle\sum_{m\in\mathcal{M}_+} Q_k^m \,=\, Q_k & & \forall\,k \,\in\, \mathcal{K} \\
0 \,\leq\, h_p \;\perp\; C_p(h) - t_{ij} + v_p \,\geq\, 0 & & \forall\,p \,\in\, \mathcal{P}_{ij},\ \forall\,(\,i,j\,) \,\in\, \mathcal{N}\times\mathcal{N} \\
0 \,\leq\, s_{ij} \;\perp\; \bar{t} - t_{ij} \,\geq\, 0 & & \forall\,(\,i,j\,) \,\in\, \mathcal{N}\times\mathcal{N} \\
0 \,\leq\, v_p \;\perp\; \bar{h} - h_p \,\geq\, 0 & & \forall\,p \,\in\, \mathcal{P} \\
0 \,\leq\, \theta_{jk}^m \;\perp\; \left[\displaystyle\sum_{j\,'\in\mathcal{D}} z_{j\,'k}^m \right] \theta_{jk}^m - z_{jk}^m + \zeta_{jk}^m \,\geq\, 0 & & \forall\,(\,j,k,m\,) \,\in\, \mathcal{D}\times\mathcal{K}\times\mathcal{N} \\
0 \,\leq\, \zeta_{jk}^m \;\perp\; 1 - \theta_{jk}^m \,\geq\, 0 & & \forall\,(\,j,k,m\,) \,\in\, \mathcal{D}\times\mathcal{K}\times\mathcal{M} \\
0 \,\leq\, t_{ij} \;\perp\; \displaystyle\sum_{p\in\mathcal{P}_{ij}} h_p - \left[\sum_{k\in\mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell)\in\mathcal{K}\times\mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m\,\in\,\mathcal{M}} z_{i\ell}^m \right] + s_{ij} \,\geq\, 0 & & \forall\,(\,i,j\,) \,\in\, \mathcal{N}\times\mathcal{N}.
\end{array}$$

In the following, we recall that the upper bounds $\bar{t}$ and $\bar{h}$ satisfy:

$$\bar{h} \,>\, \sum_{k\in\mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell)\in\mathcal{K}\times\mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m\,\in\,\mathcal{M}} z_{i\ell}^m, \quad \forall \text{ nonnegative } (\,\mathbf{z},\mathbf{Q}\,) \text{ satisfying (26)}$$

and

$$\bar{t} \,>\, \max_{(i,j)\in\mathcal{N}\times\mathcal{N}} \, \max_{p\in\mathcal{P}_{ij}} \, \max_{0\leq h\leq\bar{h}\mathbf{1}} \, C_p(h).$$

Lemma 2. Suppose that for every pair $(i,j) \in \mathcal{N}\times\mathcal{N}$, $\min\limits_{p\in\mathcal{P}_{ij}} C_p(h) \geq 0$. Then for every solution of the complementarity formulation of the VI $(\mathbf{F}^{\rho}; \mathcal{W})$, it holds that $v_p = s_{ij} = 0$ for all $p \in \mathcal{P}$ and all $(i,j) \in \mathcal{N}\times\mathcal{N}$.

Proof. Assume for contradiction that $v_{\bar{p}} > 0$ for some $\bar{p} \in \mathcal{P}_{ij}$. We then have $h_{\bar{p}} = \bar{h} > 0$. This implies

$$0 \,=\, C_{\bar{p}}(h) - t_{ij} + v_{\bar{p}} \,>\, C_{\bar{p}}(h) - t_{ij}.$$

Thus $t_{ij} > C_{\bar{p}}(h) \geq 0$. Hence,

$$h_{\bar{p}} \leq \sum_{p \in \mathcal{P}_{ij}} h_p + s_{ij} = \left[\sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \right] < \bar{h},$$

which is a contradiction. Similarly, suppose $s_{ij} > 0$ for some $(i,j) \in \mathcal{N} \times \mathcal{N}$. We then have $t_{ij} = \bar{t} > 0$. Hence, similar to the last displayed inequality, we have for all $\bar{p} \in \mathcal{P}_{ij}$,

$$h_{\bar{p}} < \sum_{p \in \mathcal{P}_{ij}} h_p + s_{ij} = \left[\sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \right] < \bar{h},$$

implying $v_{\bar{p}} = 0$. Hence $t_{ij} \leq C_{\bar{p}}(h) < \bar{t}$, which again is a contradiction. □

With v_p and s_{ij} both zero, the complementarities related to the bounds of t_{ij} and h_p are therefore not explicitly needed; i.e., adding or removing them will not change the proposed TGEM-ets model. This result enables us to study the latter model via the (penalized) VI formulation. Letting for $m > 0$,

$$\begin{aligned} \phi_j^{\rho;m} &\triangleq -\frac{\rho}{\gamma_3^m} \left[\sum_{\ell \in \mathcal{K}} z_{j\ell}^m - \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^m + \gamma_3^m \, \eta_j^m \right], \quad \lambda_k^m \triangleq \frac{\tilde{\lambda}_k^m}{\gamma_3^m}, \\ \text{and} \quad \varphi^{\rho;m} &\triangleq \frac{\rho}{\gamma_3^m} \left[\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^m \, t_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q_{k'}^m \, t_{O_{k'}D_{k'}} - N^m \right]_+, \end{aligned} \tag{27}$$

and recalling $U_k^m \triangleq V_k^m + \gamma_3^m \, \lambda_k^m$, we can write the above complementarity conditions of the VI $(\mathbf{F}^\rho; \mathcal{W})$ by dividing the first complementarity by γ_3^m and substituting the above defined variables: $\phi_j^{\rho;m}$, λ_k^m, and $\varphi^{\rho;m}$, dropping the complementarity involving the variables η_j^m and those involving the bounds on t_{ij} and h_p, which become unnecessary by Lemma 2, this results in the following conditions where all the variables are dependent on the parameter ρ (through the dependence on $\phi_j^{\rho;m}$ and $\varphi^{\rho;m}$):

$$\begin{array}{llll}
0 \leq z_{jk}^m & \perp & -\widehat{R}_{jk}^m - \beta_3^m \, t_{jO_k} - \phi_j^{\rho;m} - \lambda_k^m + t_{jO_k} \, \varphi^{\rho;m} \geq 0 & \forall \, (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \\
0 \leq \lambda_k^m & \perp & \displaystyle\sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m \geq 0 & \forall \, (k,m) \in \mathcal{K} \times \mathcal{M} \\
0 \leq Q_k^m & \perp & U_k^m - u_k \geq 0 & \forall \, (k,m) \in \mathcal{K} \times \mathcal{M}_+ \\
u_k \text{ free,} & & \displaystyle\sum_{m \in \mathcal{M}_+} Q_k^m = Q_k & \forall \, k \in \mathcal{K} \\
0 \leq h_p & \perp & C_p(h) - t_{ij} \geq 0 & \forall \, p \in \mathcal{P}_{ij}, \ \forall \, (i,j) \in \mathcal{N} \times \mathcal{N} \\
0 \leq \theta_{jk}^m & \perp & \left[\displaystyle\sum_{j' \in \mathcal{D}} z_{j'k}^m \right] \theta_{jk}^m - z_{jk}^m + \zeta_{jk}^m \geq 0 & \\
0 \leq \zeta_{jk}^m & \perp & 1 - \theta_{jk}^m \geq 0 & \forall \, (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \\
0 \leq t_{ij} & \perp & \displaystyle\sum_{p \in \mathcal{P}_{ij}} h_p - \left[\sum_{k \in \mathcal{K}} \delta_{ijk}^{\mathrm{OD}} \, Q_k + \sum_{(k,\ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m \in \mathcal{M}} z_{i\ell}^m \right] \geq 0 & \forall \, (i,j) \in \mathcal{N} \times \mathcal{N}.
\end{array}$$

The resulting complementarity conditions are one step closer to the original complementarity formulation of the overall TGEM-ets model, except for the two relaxed constraints: the e-HSP vehicle supply equations and the vehicle-hours inequalities.

7.3 Third step: the limiting argument explained

To recover the relaxed constraints and their associated multipliers, we apply a limiting argument by taking an arbitrary sequence $\{\rho_\nu\}$ of positive scalars tending to ∞. For each ν, let $\mathbf{w}^\nu$ be a corresponding solution of the VI $(\mathbf{F}^{\rho_\nu}; \mathcal{W})$. The sequence $\{\mathbf{w}^\nu\}$ belongs to the compact set $\mathcal{W}$; it therefore contains a convergent subsequence. By working with the latter subsequence, we may assume, without loss of generality, that the entire sequence $\{\mathbf{w}^\nu\}$ converges to a limit $\mathbf{w}^\infty \triangleq (\mathbf{z}^\infty, \mathbf{Q}^\infty, \mathbf{h}^\infty, \mathbf{t}^\infty, \boldsymbol{\theta}^\infty)$ which must belong to $\mathcal{W}$. The goal of this limiting argument is to show that such a limit must be a desired equilibrium solution of the TGEM-ets. Among the conditions that need to be verified, the following three are the most crucial: (a) the limiting tuple $\mathbf{w}^\infty$ satisfies

$$\sum_{\ell \in \mathcal{K}} z_{j\ell}^{\infty;m} = \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^{\infty;m} \quad \forall \, (j, m) \in \mathcal{D} \times \mathcal{M}; \tag{28}$$

(b) for each $m > 0$, the variables $\{z_{jk}^{\infty;m}\}$ satisfy the following constraint corresponding to the limiting OD demands $Q_k^{\infty;m}$ and travel times $t_{jO_k}^\infty$ and $t_{O_kD_k}^\infty$:

$$N^m - \left[\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} z_{jk}^{\infty;m} \, t_{jO_k}^\infty + \sum_{k \in \mathcal{K}} Q_k^{\infty;m} \, t_{O_kD_k}^\infty \right] \geq 0; \tag{29}$$

and (c) the same variables $\{z_{jk}^{\infty;m}\}$ are optimal for the e-HSP$_m$ problem associated with the exogenous variables $\{Q_k^{\infty;m}, t_{jO_k}^\infty, t_{O_kD_k}^\infty\}$. The challenge in this proof is the possible unboundedness of the remaining variables:

$$\phi_j^{\nu;m}; \; \lambda_k^{\nu;m}; \; \varphi^{\nu;m}; \; \zeta_{jk}^{\nu;m}; \text{ and } u_k^\nu, \tag{30}$$

which are the short-hands of the respective variables that are dependent on ρ_ν. These variables prevent us from directly passing to the limit $\nu \to \infty$ in the complementarity conditions corresponding to ρ_ν. Among the variables in (30), the ones that require the most attention are $\varphi^{\nu;m}$ which appear in the bilinear (thus nonlinear) term: $t_{jO_k}^{\nu;m} \varphi^{\nu;m}$. The other variables in (30) appear linearly in the complementarity conditions; they can be handled by a result from linear complementarity theory to be introduced later.

7.4 Fourth step: a feasibility assumption and a lemma

So far, no assumption has been imposed on the number of vehicles N^m in relation to the (given) OD demands Q_k. Since the trip makers have the option of solo driving if there are not sufficient number of e-HSP vehicles or if their waiting times are too long, one would hope that the existence of an equilibrium solution could be established without requiring any such relation. Mathematically, this appears quite difficult, if not impossible, because of the interconnections among the various players in the system in addition to the network congestion effects. In what follows, we impose a reasonable "feasibility assumption" under which the existence proof can be completed. In particular, the condition (31) involving the positive scalar ε is a type of a Slater condition that is needed to deal with the nonlinearity of the products: $z^m_{jk}\, t_{jO_k}$ and $Q^m_k\, t_{O_kD_k}$ of the unknowns. Let $\mathcal{T}$ be the set of travel times induced by pairs of vehicle-demand profiles $(\mathbf{z}, \mathbf{Q})$ satisfying the first three conditions in (26) that define the polyhedron $\mathcal{W}$.

Lemma 3. Suppose that there exists a scalar $\varepsilon > 0$ such that for each nonnegative tuple $\mathbf{Q}$ satisfying $\sum_{m\in\mathcal{M}_+} Q^m_k = Q_k$ for all $k \in \mathcal{K}$ and for each tuple $\mathbf{t} \in \mathcal{T}$, there exists $\mathbf{z}$ such that the tuple $(\mathbf{z}, \mathbf{Q}, \mathbf{t})$ satisfies the constraints in (8) for all $m > 0$ with

$$N^m - \left[\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{D}} z^m_{jk}\, t_{jO_k} + \sum_{k\in\mathcal{K}} Q^m_k\, t_{O_kD_k}\right] \geq \varepsilon \quad \forall\, m \in \mathcal{M}. \tag{31}$$

The following two statements hold:

(a) the limiting tuple $(\mathbf{z}^\infty, \mathbf{Q}^\infty, \mathbf{t}^\infty)$ satisfies (28) and (29);

(b) for each $m \in \mathcal{M}$, the sequence $\{\varphi^{\nu;m}\}$ is bounded.

Proof. Corresponding to the pair $(\mathbf{Q}^\nu, \mathbf{t}^\nu)$, let $\bar{\mathbf{z}}^\nu$ be such that the tuple $(\bar{\mathbf{z}}^\nu, \mathbf{Q}^\nu, \mathbf{t}^\nu)$ satisfies the constraints in (8) for all $m > 0$ with

$$N^m - \left[\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{D}} \bar{z}^{\nu;m}_{jk}\, t^\nu_{jO_k} + \sum_{k\in\mathcal{K}} Q^{\nu;m}_k\, t^\nu_{O_kD_k}\right] \geq \varepsilon \quad \forall\, m \in \mathcal{M}.$$

By definition of the VI $(\mathbf{F}^{\rho_\nu}; \mathcal{W})$, we have

$$\mathbf{F}^{\rho_\nu}(\mathbf{w}^\nu)^T(\,\mathbf{w} - \mathbf{w}^\nu\,) \geq 0, \quad \forall\, \mathbf{w} \in \mathcal{W};$$

since the tuple $\overline{\mathbf{w}}^\nu = (\bar{\mathbf{z}}^\nu, \mathbf{Q}^\nu, \mathbf{h}^\nu, \mathbf{t}^\nu, \boldsymbol{\theta}^\nu)$ is clearly an element of $\mathcal{W}$ and differs from $\mathbf{w}^\nu$ only in the $\mathbf{z}$-variables, it follows that

$$\begin{aligned}
0 &\leq \mathbf{F}^{\rho_\nu}(\mathbf{w}^\nu)^T(\,\overline{\mathbf{w}}^\nu - \mathbf{w}^\nu\,) \\
&= \sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{D}}\sum_{k\in\mathcal{K}} \left[-\gamma^m_3\left(\widehat{R}^{\nu;m}_{jk} + \beta^m_3\, t^\nu_{jO_k}\right) + \rho_\nu\left(\sum_{\ell\in\mathcal{K}} z^{\nu;m}_{j\ell} - \sum_{k':\,j=\mathrm{D}_{k'}} Q^{\nu;m}_{k'}\right)\right]\left(\bar{z}^{\nu;m}_{jk} - z^{\nu;m}_{jk}\right) + \\
&\quad \rho_\nu \sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{D}}\sum_{k\in\mathcal{K}} t^\nu_{jO_k}\left[\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{D}} z^{\nu;m}_{j'k'}\, t^\nu_{j'O_{k'}} + \sum_{k'\in\mathcal{K}} Q^{\nu;m}_{k'}\, t^\nu_{O_{k'}D_{k'}} - N^m\right]_+\left(\bar{z}^{\nu;m}_{jk} - z^{\nu;m}_{jk}\right) \\
&= -\gamma^m_3 \underbrace{\sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{D}}\sum_{k\in\mathcal{K}}\left(\widehat{R}^{\nu;m}_{jk} + \beta^m_3\, t^\nu_{jO_k}\right)\left(\bar{z}^{\nu;m}_{jk} - z^{\nu;m}_{jk}\right)}_{\text{bounded}} - \rho_\nu \sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{D}}\left(\sum_{\ell\in\mathcal{K}} z^{\nu;m}_{j\ell} - \sum_{k':\,j=\mathrm{D}_{k'}} Q^{\nu;m}_{k'}\right)^2 + \\
&\quad \rho_\nu \sum_{m\in\mathcal{M}}\left[\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{D}} z^{\nu;m}_{j'k'}\, t^\nu_{j'O_{k'}} + \sum_{k'\in\mathcal{K}} Q^{\nu;m}_{k'}\, t^\nu_{O_{k'}D_{k'}} - N^m\right]_+ \sum_{j\in\mathcal{D}}\sum_{k\in\mathcal{K}} t^\nu_{jO_k}\left(\bar{z}^{\nu;m}_{jk} - z^{\nu;m}_{jk}\right)
\end{aligned}$$

using the condition that

$$\sum_{\ell \in \mathcal{K}} \bar{z}_{j\ell}^{\nu;m} \; = \; \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^{\nu;m} \; \leq \; \sum_{k' \in \mathcal{K}} Q_k.$$

Moreover, since

$$N^m - \left[\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{D}} \bar{z}_{jk}^{\nu;m} \, t_{jO_k}^{\nu} + \sum_{k \in \mathcal{K}} Q_k^{\nu;m} \, t_{O_k D_k}^{\nu} \right] \; \geq \; \varepsilon$$

by the choice of $\mathbf{z}^\nu$, we deduce

$$\sum_{j \in \mathcal{D}} \sum_{k \in \mathcal{K}} t_{jO_k}^{\nu} \left(\bar{z}_{jk}^{\nu;m} - z_{jk}^{\nu;m} \right) \; \leq \; - \left[\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^{\nu;m} \, t_{j'O_{k'}}^{\nu} + \sum_{k' \in \mathcal{K}} Q_{k'}^{\nu;m} \, t_{O_{k'} D_{k'}}^{\nu} - N^m \right] - \varepsilon.$$

Since $t_+ t \geq (t_+)^2$ for any scalar t, we deduce

$$\begin{aligned} 0 \; \leq \; & \text{ a bounded term} - \rho_\nu \sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{D}} \left(\sum_{\ell \in \mathcal{K}} z_{j\ell}^{\nu;m} - \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^{\nu;m} \right)^2 \\ & - \rho_\nu \sum_{m \in \mathcal{M}} \left[\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^{\nu;m} \, t_{j'O_{k'}}^{\nu} + \sum_{k' \in \mathcal{K}} Q_{k'}^{\nu;m} \, t_{O_{k'} D_{k'}}^{\nu} - N^m \right]_+^2 \\ & - \varepsilon \sum_{m \in \mathcal{M}} \underbrace{\rho_\nu \left[\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^{\nu;m} \, t_{j'O_{k'}}^{\nu} + \sum_{k' \in \mathcal{K}} Q_{k'}^{\nu;m} \, t_{O_{k'} D_{k'}}^{\nu} - N^m \right]_+}_{= \gamma_3^m \, \varphi^{\nu;m}} . \end{aligned}$$

This shows that

- $\lim\limits_{\nu \to \infty} \left(\underbrace{\sum_{\ell \in \mathcal{K}} z_{j\ell}^{\nu;m} - \sum_{k' : j = \mathrm{D}_{k'}} Q_{k'}^{\nu;m}}_{\text{denoted } e_j^{\nu;m}} \right) \; = \; 0$ for all $m \in \mathcal{M}$ and $j \in \mathcal{D}$;
- $\lim\limits_{\nu \to \infty} \underbrace{\left[\sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z_{j'k'}^{\nu;m} \, t_{j'O_{k'}}^{\nu} + \sum_{k' \in \mathcal{K}} Q_{k'}^{\nu;m} \, t_{O_{k'} D_{k'}}^{\nu} - N^m \right]_+}_{\text{set as } \widehat{e}^{\nu;m} = \frac{\varphi^{\nu;m}}{\rho_\nu}} \; = \; 0$; and
- the sequence $\{\varphi^{\nu;m}\}$ is bounded for all $m \in \mathcal{M}$.

These conclusions establish the two claims of the lemma. $\square$

Final step: passing to the limit by a linear complementarity lemma

We may augment the complementarity conditions of the VI $(\mathbf{F}^{\rho_\nu};\mathcal{W})$ by the conditions:

$$\phi_j^{\rho_\nu;m} \text{ free,} \qquad \sum_{\ell\in\mathcal{K}} z_{j\ell}^{\nu;m} - \sum_{k':\,j=\mathrm{D}_{k'}} Q_{k'}^{\nu;m} - e_j^{\nu;m} = 0, \quad \forall\,(j,m) \in \mathcal{D}\times\mathcal{M}$$

$$0 \le \varphi^{\nu;m} \quad \perp \quad \widehat{e}^{\nu;m} + N^m - \left[\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{D}} z_{j'k'}^{\nu;m}\, t_{j'O_{k'}}^{\nu} + \sum_{k'\in\mathcal{K}} Q_{k'}^{\nu;m}\, t_{O_{k'}D_{k'}}^{\nu}\right] \ge 0,$$

re-arrange the augmented conditions, and write them in the following form to facilitate the application of a general complementarity result to complete the existence proof:

$$\begin{array}{lll}
0 \le z_{jk}^{\nu;m} & \perp \; -\widehat{R}_{jk}^{\nu;m} - \beta_3^m\, t_{jO_k}^{\nu} - \phi_j^{\nu;m} - \lambda_k^{\nu;m} + t_{jO_k}^{\nu}\,\varphi^{\nu;m} \ge 0 & \forall\,(j,k,m) \in \mathcal{D}\times\mathcal{K}\times\mathcal{M} \\[2ex]
0 \le Q_k^{\nu;0} & \perp \; V_k^{\nu;0} - u_k^{\nu} \ge 0 & \\
0 \le Q_k^{\nu;m} & \perp \; V_k^{\nu;m} - u_k^{\nu} + \gamma_3^m\,\lambda_k^{\nu;m} \ge 0 & \forall\,(k,m) \in \mathcal{K}\times\mathcal{M} \\[2ex]
0 \le \varphi^{\nu;m} & \perp \; \widehat{e}^{\nu;m} + N^m - \left[\displaystyle\sum_{k'\in\mathcal{K}}\sum_{j'\in\mathcal{D}} z_{j'k'}^{\nu;m}\, t_{j'O_{k'}}^{\nu} + \sum_{k'\in\mathcal{K}} Q_{k'}^{\nu;m}\, t_{O_{k'}D_{k'}}^{\nu}\right] \ge 0 & \forall\, m \in \mathcal{M} \\[2ex]
0 \le t_{ij}^{\nu} & \perp \; \displaystyle\sum_{p\in\mathcal{P}_{ij}} h_p^{\nu} - \left[\sum_{k\in\mathcal{K}} \delta_{ijk}^{\mathrm{OD}}\, Q_k + \sum_{(k,\ell)\in\mathcal{K}\times\mathcal{K}} \delta_{ijk\ell}^{\mathrm{e-HSP}} \sum_{m\in\mathcal{M}} z_{i\ell}^{\nu;m}\right] \ge 0 & \forall\,(i,j) \in \mathcal{N}\times\mathcal{N} \\[2ex]
0 \le h_p^{\nu} & \perp \; C_p(h^{\nu}) - t_{ij}^{\nu} \ge 0 & \forall\, p \in \mathcal{P}_{ij},\ \forall\,(i,j) \in \mathcal{N}\times\mathcal{N} \\
0 \le \theta_{jk}^{\nu;m} & \perp \; \left[\displaystyle\sum_{j'\in\mathcal{D}} z_{j'k}^{\nu;m}\right] \theta_{jk}^{\nu;m} - z_{jk}^{\nu;m} + \zeta_{jk}^{\nu;m} \ge 0 & \forall\,(j,k,m) \in \mathcal{D}\times\mathcal{K}\times\mathcal{N} \\[2ex]
\hline
\phi_j^{\rho_\nu;m} \text{ free,} & \displaystyle\sum_{k\in\mathcal{K}} z_{jk}^{\nu;m} - \sum_{k':\,j=\mathrm{D}_{k'}} Q_{k'}^{\nu;m} = e_j^{\nu;m} & \forall\,(j,m) \in \mathcal{D}\times\mathcal{M} \\[2ex]
u_k^{\nu} \text{ free,} & \displaystyle\sum_{m\in\mathcal{M}_+} Q_k^{\nu;m} = Q_k & \forall\, k \in \mathcal{K} \\[2ex]
\hline
0 \le \lambda_k^{\nu;m} & \perp \; \displaystyle\sum_{j\in\mathcal{D}} z_{jk}^{\nu;m} - Q_k^{\nu;m} \ge 0 & \forall\,(k,m) \in \mathcal{K}\times\mathcal{M} \\
0 \le \zeta_{jk}^{\nu;m} & \perp \; 1 - \theta_{jk}^{\nu;m} \ge 0 & \forall\,(j,k,m) \in \mathcal{D}\times\mathcal{K}\times\mathcal{M},
\end{array}$$

where the primary variables $(z_{jk}^{\nu;m}, Q_k^{\nu;m}, t_{ij}^{\nu}, h_p^{\nu}, \theta_{jk}^{\nu;m})$; thus the derived variables $R_{jk}^{\nu;m}$ and $V_k^{\nu;m}$, together with the multipliers $\varphi^{\nu;m}$ in the first group of complementarity conditions, are all bounded, while the multipliers $(\phi_j^{\rho_\nu;m}, u_k^{\nu})$ in the group of equality constraints and the multipliers $(\lambda_k^{\nu;m}, \zeta_{jk}^{\nu;m})$ in the group of inequality constraints are not necessarily bounded. These multipliers appear linearly in the first group of complementarity conditions. Due to such linearity, in spite of the possible unboundedness of the multipliers, we can apply the lemma below to complete the proof of the existence of a TGEM-ets equilibrium. Proof of the lemma follows from the theory of complementary cones in linear complementarity theory [4].

Lemma 4. Consider the following mixed complementarity conditions:

$$\begin{aligned}
0 \le x \quad &\perp \quad G(x) + A^T\mu + B^T\lambda \ge 0 \\
\mu \text{ free} \quad & \quad Cx - e = 0 \\
0 \le \lambda \quad &\perp \quad Dx - f \ge 0,
\end{aligned}$$

where G is continuous. Let $\{(x^k, \mu^k, \lambda^k)\}$ be a sequence of solutions corresponding to the convergent sequence $\{(e^k, f^k)\}$ with $\lim_{k\to\infty} e^k = e^\infty$ and $\lim_{k\to\infty} f^k = f^\infty$. If $\lim_{k\to\infty} x^k = x^\infty$, then there exists $(\mu^\infty, \lambda^\infty)$ such that the triple $(x^\infty, \mu^\infty, \lambda^\infty)$ is a solution corresponding to (e^∞, f^∞). □

Summarizing the above derivations, we have proved the following existence result.

Theorem 1. Under the assumptions of Lemmas 1, 2, and 3, a normalized TGEM-ets equilibrium exists for all scalars $\gamma_3^m > 0$ for all $m \in \mathcal{M}$. □

By showing that the omitted e-HSP vehicle supply equations and vehicle-hours bound constraints are satisfied by the limiting solution as the penalty parameter ρ tends to ∞, the theorem establishes that an equilibrium solution exists. It is important to note, however, that such constraints are not shown to hold for finite ρ. The latter finite recovery of the omitted constraints would be a kind of an *exact penalty* result for complementarity problems that does not exist to date.

7.5 Discussion about uniqueness

For a highly nonlinear equilibrium model as complex as the TGEM-ets, it is difficult to expect the uniqueness of a solution. In fact, there is no such uniqueness result in the theory of generalized Nash equilibrium problems known to date. The main reason is due to the coupling of variables in the players' constraints; this is further complicated by the nonlinearities throughout the model. In general, there are two ways to deal with the multiplicity of solutions. One, while an equilibrium solution is not globally unique, the local uniqueness of such a solution can be analyzed via the theory of variational inequalities; see [7, Chapter 5]. Since the practical significance of a locally unique equilibrium solution is not clear, we refrain from lengthening the paper by carrying out such an analysis. An alternative way to deal with the multiplicity of solutions is to select one such solution with the aid of a secondary objective. Mathematically, this solution selection can be formulated as a mathematical program with equilibrium constraints [16] set up to optimize the secondary objective among all equilibrium solutions. For instance, one such objective could be the deadhead miles given by (1). This may also include methods on how modelers plan to deal with the uncertainty due to the multiplicity of solutions, such as risk-averse or risk-neutral design schemes [2, 3]. Details of this issue are best left for future investigation.

8 Numerical Experiments

Formulated as an MCP, the TGEM-ets model is solved using the PATH solver in GAMS in the numerical experiments in this section; more details about PATH and GAMS can be found in [8, 12], respectively.

Through the extensive developments and continued refinements by its principal architects: Michael Ferris (at the University of Wisconsin at Madison) and Todd Munson (at Argonne National Laboratory), this solver has become very effective for solving MCPs in practice and has been routinely used for solving equilibrium problems arising in real-world energy modeling, electricity markets, transportation, and other science/engineering fields. Although there is an accompanying theory for the convergence of the algorithms behind this solver that can be found in the monograph by Facchinei and Pang [7] on this subject, this convergence theory has never been shown to be directly applicable to the applied problems; yet, computationally, the solver has been successful in consistently obtaining practically acceptable solutions with some occasional adjustments of the solver's default settings (e.g., the starting points) to ensure its successful termination. For large-scale problems, specialized algorithms may need to be developed to enhance the solution of the proposed TGEM-ets more efficiently. Such an algorithmic development is beyond the scope of this paper.

8.1 Results of the small network

The model and solution method are tested on two networks: a small network and the Sioux-Falls network. The first network is a "4-node-9-link" small network, as shown in Figure 2. Parameters of the network are shown in Table 1, which are needed in the network congestion module. For most of the numerical experiments for this network, node 1 is the origin, and nodes 2, 3, and 4 are the destinations, with travel demands being 50, 40, and 50 respectively. Later in this subsection, we also test the scenarios with varying travel demands from nodes 2, 3, and 4 to node 1.

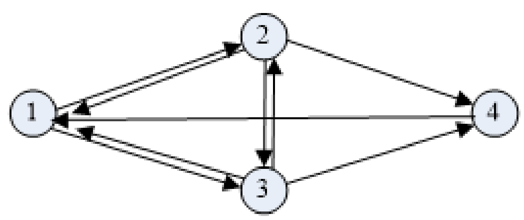

Figure 2: Illustration of the small network

Parameters related to the e-HSP vehicle fees and the customer disutility are provided in Table 2. The proposed model and analysis method in this paper can be applied to any types of e-HSP services as long as their characteristics (e.g., the parameters in the table) can be modeled. To illustrate, we assume two e-HSP modes (in addition to solo-driving): e-HSP I ($m = 1$) is more expensive, but is perceived as safer (or more comfortable), while e-HSP II ($m = 2$) is less expensive but is also perceived as less safe or less comfortable. In the current situation, e-HSP I is similar to e-hailing taxi services and e-HSP II is similar to TNC services (such as Uber, Lyft, or Didi). To capture the different characteristics of the two e-HSP modes, we set $\alpha_1^1 \geq \alpha_1^2$ and $\beta_1^1 = \beta_1^2$, and $\alpha_2^1 \geq \alpha_2^2$ and $\beta_2^1 \leq \beta_2^2$. We also set $\beta_2^2 = \beta_2^0$ to indicate that the distance-based cost for solo driving is the same as that for e-HSP II services (and larger than that for e-HSP I services). The parameters in Table 2 serve as the "base" parameters and are used in most of the numerical experiments in this section, unless stated otherwise. The e-HSP vehicle fee parameters

Table 1: Parameters of the small network

Links	From node	To node	Free flow travel time (h)	Length (mile)	Capacity
1	1	2	0.3	10	40
2	1	3	0.5	20	40
3	2	3	0.4	20	60
4	2	4	0.4	10	40
5	3	4	0.3	20	40
6	4	1	1.0	40	60
7	2	1	0.4	15	50
8	3	1	0.4	20	60
9	3	2	0.5	20	40

are based on those from the City of Seattle [23, 22], with some modifications. We conduct sensitivity analysis later in this section regarding how different choices of some of the parameters ($\alpha_1^m, \alpha_2^m, \beta_2^m, \gamma_1^m$) may change the customers' choice of modes and the total system VMT and VHT. Other parameters are not tested because either the model results are not very sensitive to these parameters or they are the same for the two types of e-HSP services.

Table 2: Parameters for e-HSP fee charging and customer disutility

Parameters	Notation	Small network	Sioux-Falls
Fixed fare ($)	$F^m(m = 1, 2)$	3, 2	3, 2
Time-based fare rate ($/hr)	$\alpha_1^m(m = 1, 2)$	20, 15	15, 10
Distance-based fare rate ($/mile)	$\alpha_2^m(m = 1, 2)$	2, 1.5	2.25, 2
Conversion factor (from travel time to cost, $/hr)	$\beta_1^m(m = 1, 2)$	2, 2	2, 2
Conversion factor (from travel distance to cost, $/mile)	$\beta_2^m(m = 0, 1, 2)$	0.95, 0.55, 0.9	2 ,0.5, 2
Conversion factor (from OSP vehicle waiting time to cost)	$\beta_3^m(m = 1, 2)$	0.2, 0.1	2 ,1
Value of time of customer (while traveling, $/hr)	$\gamma_1^m(m = 0, 1, 2)$	40, 7, 18	50, 3, 15
Value of time of customer (while waiting, $/hr)	$\gamma_2^m(m = 0, 1, 2)$	0, 3, 2	0, 1, 0.5
The number of e-HSP vehicles	$N^m(m = 1, 2)$	400, 400	40000, 40000

For this small network, the TGEM-ets model contains 175 variables (also 175 complementarity conditions) and the solution (CPU) time is about 0.5 seconds on a PC with a 3.4GHz processor and 16GB memory. The results of solving TGEM-ets verify that the equilibria are reached for both customers' mode choices (in terms of choosing solo driving or one of the two e-HSP modes) and path choices (in terms of choosing the minimum travel time path when traveling between an OD pair). Table 3 shows the results of this base case, including, for each OD pair, the mode choices and minimum disutility of the customers, the number of used paths, and the minimum path travel time. In the table, the selected modes and their disutility are highlighted in bold text. The VMT and vehicle hours traveled (VHT) are 5463.47 vehicle-miles and

3.71 vehicle-hours, respectively, and the deadhead miles are 2683.53 vehicle-miles. It turns out that, as expected, the deadhead miles of a particular scenario is exactly the difference between its VMT and the VMT of the scenario when only solo driving is used (for the same network, same demand pattern).

Table 3: Results of the small network (base case)

OD Pair	**Mode Choice**	**Customer Disutility**	**# of Used Paths**	**Min Path TT**
$1->2$	**Solo: 4.499** e-HSP I: 0 **e-HSP II: 45.501**	**Solo: 44.962** e-HSP I: 44.974 **e-HSP II: 44.962**	1	0.887
$1->3$	**Solo: 40** e-HSP I: 0 e-HSP II: 0	**Solo: 58.657** e-HSP I: 61.779 e-HSP II: 59.293	1	0.991
$1->4$	Solo: 0 **e-HSP I: 44.165** **e-HSP II: 5.835**	Solo: 70.89 **e-HSP I: 67.317** **e-HSP II: 67.317**	2	1.297

We next test how the results (especially customer mode choices and the VMT/VHT) vary with respect to the changes of the parameters. In particular, we show the total VMT, total VHT, and the VHTs of the two e-HSP modes, in these tables. Four set of parameters are selected for this purpose: time-based fare rate ($\alpha_1^m, m = 1, 2$), distance-based fare rate ($\alpha_2^m, m = 1, 2$), distance-based conversation factor ($\beta_2^m, m = 0, 1, 2$), and value of time of customers ($\gamma_1^m, m = 1, 2$). Table 4—Table 7 show the results of changing these parameters. Notice that the parameters not shown in each table are the same as those in the base case in Table 2. When changing the distance-based conversation factors, we keep the factors the same for solo driving and e-HSP II, i.e., $\beta_2^0 = \beta_2^2$. When changing customers' value of time parameters, we set the parameter for solo driving fixed as the base case ($\gamma_1^0 = 40$), and change those for the two e-HSP modes accordingly. Since each parameter represents a certain "cost" to the customers, one would expect that increasing the value of the parameter associated with a particular mode, customers' choice of that mode will decrease (or at least remain the same). The results show such a trend for most of the parameters as shown in the tables.

The four tables show that once the e-HSP usage (by either mode) increases, the total system VMT/VHT also increases and in many cases fairly dramatically. For example, Table 5 illustrates that from everyone choosing solo driving to everyone choosing e-HSP modes, the VMT increase is more than doubled (from 2779.94 to 6329.94). These results and findings are for this specific demand pattern, which is totally asymmetric (demands are only from 1 to 2, 3, and 4, but no demands going back to 1 at all). To test how the demand symmetry may affect the network congestion (represented by the total VMT), we add more OD pairs to the original setting. In particular, node 1 is also a destination with 2, 3, and 4 as the origins. Varying demands are tested. From 2 to 1 and 4 to 1, three demand scenarios are tested: 10, 30, and 50. From 3 to 1, three demand scenarios are tested as well: 10, 25, and 40. The combinations of different demands of each OD pair will produce 27 demand scenarios, from highly asymmetric (10 from

Table 4: Results of changing time-based fare rate ($\alpha_1^m, m = 1, 2$)

α_1^1	α_1^2	**Solo%**	**e-HSP I%**	**e-HSP II%**	**VMT**	**VHT (e-HSP I)**	**VHT (e-HSP II)**	**VHT**
20	8.6	28.57	0.00	71.43	5529.94	0.00	3.10	3.76
20	12.6	28.57	35.72	35.71	5529.94	1.12	1.97	3.76
20	14.5	28.57	44.62	26.81	5529.94	2.16	0.94	3.76
15	10	0.00	35.71	64.29	6329.94	1.12	2.91	4.03
20.5	10	28.57	0.00	71.43	5529.94	0.00	3.10	3.76
20	15	31.74	31.51	36.75	5463.47	1.74	1.24	3.71
32	15	64.29	0.00	35.71	4779.94	0.00	1.97	3.37
40	30	100.00	0.00	0.00	2779.94	0.00	0.00	2.48

all three nodes to 1), to perfectly symmetric (2 to 1 and 4 to 1 have 50, and 3 to 1 has 40, which is exactly the same as the demands from 1 to the other nodes). To quantify the level of symmetry of the OD demands, we introduce the demand "symmetry indicator", which is calculated as the total demands from the destinations (2,3,4) to the origin 1 divided by 140 (which is the total demand initially from origin 1 to all destinations). The relationship between e-HSP mode usage and the total VMT of the network is shown in Figure 3. In the figure, the vertical axis is the percentage of VMT increase over the VMT when all customers choose solo driving. Each plot in Figure 3 shows, for a given symmetry indicator (from 0 to 1 with 0.2 as the increment), the increase of the total VMT over the solo-VMT (i.e., the case with all customers choosing solo driving) with respect to the percentage of customers who choose e-HSP modes. We can see that as the symmetry indicator increases, the VMT increase will get reduced in general. Furthermore, there are more fluctuations in the changes of VMT increase with respect to the percentage of e-HSP use when the symmetry indicator is not zero. The plots indicate that when demand symmetry increases, it is more likely for a e-HSP vehicle to pickup the next customer from a nearby location, thus reducing the deadhead miles and the total network VMT.

8.2 Results of the Sioux-Falls network

Our next set of experiments is based on the Sioux-Falls network that has been studied extensively in the transportation network modeling literature. The geometry and parameters of the network can be found in [24], which are omitted in this paper. We initially select five nodes (1,2,4,7,9) as the origins and five other nodes (13,19,20,23,24) as the destination nodes, resulting in 25 OD pairs. The origins and destinations are selected so that they can mimic morning commuting trips from the suburbs to the center of the city. The demand for each OD pair is the original OD demand multiplied by 10, to create a similar congestion effect in the network. For each OD pair, we explicitly enumerate all its paths (some have nearly 5,000 paths; see [3]) and select the 20 shortest paths (in terms of free flow travel time) in the network congestion model. The travel time of the longest path in the 20 selected paths for an OD pair is on average about 2 times longer than the shortest path of the same OD pair. To test how demand

Table 5: Results of changing distance-based fare rate ($\alpha_2^m, m = 1, 2$)

α_2^1	α_2^2	**Solo%**	**e-HSP I%**	**e-HSP II%**	**VMT**	**VHT (e-HSP I)**	**VHT (e-HSP II)**	**VHT**
1.5	1.2	0.00	100.00	0.00	6329.94	4.03	0.00	4.03
1.6	1.2	0.00	71.43	28.57	6329.94	3.09	0.93	4.03
1.71	1.2	0.00	58.00	42.00	6329.94	2.35	1.68	4.03
1.84	1.2	0.00	35.71	64.29	6329.94	1.12	2.91	4.03
1.9	1.4	0.00	71.43	28.57	6329.94	3.10	0.93	4.03
1.96	1.4	0.00	56.48	43.52	5579.94	2.98	1.05	4.03
2.01	1.4	20.79	35.71	43.50	5747.80	1.12	2.23	3.83
1.9	1.5	28.57	71.43	0.00	5529.94	3.10	0.00	3.76
2.03	1.5	64.29	35.71	0.00	4779.94	1.97	0.00	3.37
2.23	1.5	64.29	0.00	35.71	4779.99	0.00	1.97	3.37
2	1.8	28.57	71.43	0.00	5529.98	3.10	0.00	3.76
2.2	1.8	100.00	0.00	0.00	2779.94	0.00	0.00	2.48

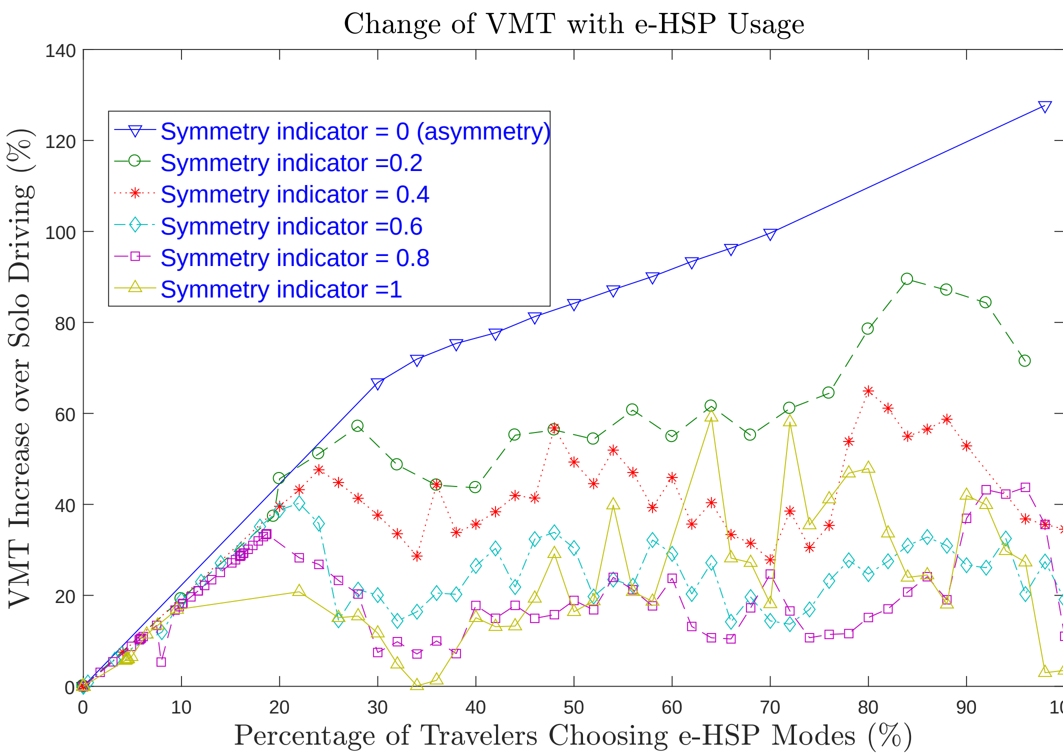

Figure 3: VMT changes with percentage of customers choosing e-HSP modes (small network)

symmetry and e-HSP usage impact the VMT changes for the Sioux-Falls network, we add 25 more OD pairs by reversing the roles of the 10 distinguished nodes: (1,2,4,7,9) and (13, 19, 20, 23, 24). In the

Table 6: Results of changing distance-based conversation factor ($\beta_2^m, m = 0, 1, 2$)

$\beta_2^0 = \beta_2^2$	β_2^1	**Solo%**	**e-HSP I%**	**e-HSP II%**	**VMT**	**VHT (e-HSP I)**	**VHT (e-HSP II)**	**VHT**
0.92	0.7	28.57	71.43	0.00	5529.94	3.10	0.00	3.76
0.92	0.82	28.57	35.85	35.58	5529.94	1.13	1.97	3.37
0.92	0.88	64.29	0.00	35.71	4779.94	0.00	1.97	3.37
0.66	0.25	100.00	0.00	0.00	2779.93	0.00	0.00	2.48
0.76	0.25	85.14	14.86	0.00	3611.94	0.80	0.00	2.83
0.94	0.25	42.35	34.25	23.39	5240.51	1.89	0.74	3.58
0.76	0.15	82.55	17.45	0.00	3757.21	0.94	0.00	2.89
0.77	0.15	66.16	33.84	0.00	4674.84	1.86	0.00	3.32

Table 7: Results of changing customer value of time ($\gamma_1^m, m = 1, 2$)

γ_1^1	γ_1^2	**Solo%**	**e-HSP I%**	**e-HSP II%**	**VMT**	**VHT (e-HSP I)**	**VHT (e-HSP II)**	**VHT**
3	18	0.00	100.00	0.00	6329.99	4.03	0.00	4.03
4	18	28.57	71.43	0.00	5529.94	3.10	0.00	3.76
7	18	64.29	0.00	35.71	4779.94	0.00	1.97	3.37
7	17.3	28.57	66.11	5.32	5529.94	2.88	0.21	3.76
7	16.3	0.00	35.71	64.29	6329.94	1.12	2.91	4.03
7	15	0.00	0.00	100.00	6329.94	0.00	4.03	4.03
8	17.3	28.57	35.71	35.71	5529.94	1.12	1.97	3.76
8	17	0.00	0.00	100.00	6329.98	0.00	4.03	4.03

additional 25 OD pairs, the former 5 nodes are destinations and the latter 5 nodes are origins. Thus the total number of OD pairs is 50. Varying demands are provided to the additional 25 OD pairs to test different demand symmetry levels. Similarly we define the demand "symmetry indicator" for the Sioux Fall network as the total demands from the destinations (13, 19, 20, 23, 24) to the origins (1,2,4,7,9) divided by 77,000 (which is the total demand initially from all origins to all destinations).

The TGEM-ets model for the Sioux Falls network results in 6736 variables. The PATH solver can solve the problem successfully, with the solution (CPU) time about 40 seconds on a PC with a 3.4GHz processor and 16GB memory. In some cases, different starting points were selected when the solver failed with the default ones. The results, especially the trends of mode choices and VMT, of this network are very similar to those for the small network. Details are omitted here to save space. Figure 4 illustrates how VMT changes with respect to the percentage of customers choosing e-HSP modes, under different demand symmetry indicators. Similar to that of the small network in Figure 3, it is nearly a linear increase with the increase of e-HSP usage when the demand is completely asymmetric (symmetry indicator is zero), possibly because it is for the extremely asymmetric demand pattern and the origins and destinations are relatively far away from each other (meaning an e-HSP vehicle has to travel a long distance to be back to the origins before it can pick up the next customer). As the demand symmetry indicator increases, the VMT increases get reduced nearly monotonically, with more fluctuations. It can also be seen clearly from the figure that as long as the e-HSP usage is not zero, the total VMT of the network is generally larger than that of solo driving. Such VMT increase is larger when e-HSP usage is larger or the demand symmetry indicator is smaller (i.e., the demand is more asymmetric). The specific shapes of such trends may be different for different networks. However, we suspect that the general trend should be true for other networks as well.

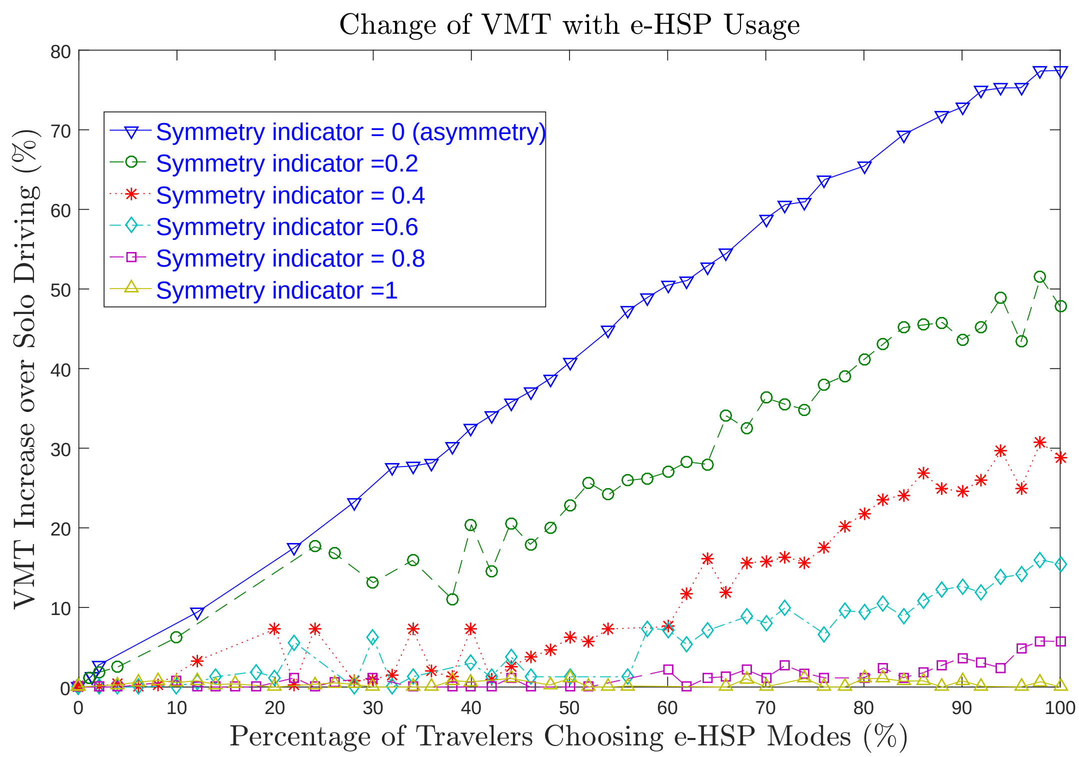

Figure 4: VMT changes with percentage of customers choosing e-HSP modes (Sioux-Falls Network)

We can also observe that it is possible that at certain specific demand symmetry indicator (e.g., 100% for perfect symmetry) and specific e-HSP usage (e.g., a little more than 60%), the VMT increase can be zero. This is because for these specific cases, it is possible that e-HSP vehicles can just pick up customers at their drop-off locations and there is no deadhead miles. Also, at very high e-HSP usage levels the deadhead miles can go down substantially depending on the network topology (e.g., high demand symmetry) since it is likely that consecutive service requests are close to one another. However, from the figure, when the demand symmetry is small and the e-HSP usage is large, it is less likely that the VMT increase is zero due to e-hailing, i.e., there are always significant deadhead miles. These observation may provide useful insight on planning/regulating e-hailing services in a traffic network. For example, for a real urban traffic network, as the morning or afternoon commuting demand is usually very asymmetric (i.e, in the morning more people travel to the city from suburban, while in the afternoon more people travel reversely), significant e-HSP usage (in terms of transporting the percentage of total network demand) may noticeably increase the total VMT of the system. This might have contributed to some recent practical observations related to e-hailing. For example, in New York City, it was reported that e-hailing services have created more congestion during peak periods than non-peak periods [20]. It was also reported in [18] that e-hailing leads to mild increase of congestion in a southern city in China. We note here that in practice the actual impact of e-hailing services on VMT/VHT and the overall congestion can be more complex. This is because there are other factors (e.g., the interactions of e-hailing with transit, ridesharing, parking, etc.) that may influence how e-hailing services may impact the total VMT/congestion, which however are not modeled in the current paper. For example, e-hailing services may reduce the amount of driving made by travelers [1]. Another important factor of solo driving that we did not model in this paper is the searching for parking at the destination. Previous research showed that in urban areas, up to 30% of the traffic could be those vehicles searching for a parking space. However, it is unclear about the percentage of such searching for parking traffic in the morning commuting period—this should be significantly lower as most likely people already have arrangements for parking if they choose to drive to work everyday. In any case, the tradeoff between the VMT increase due to e-hailing (e.g., deadhead miles, induced demands) and the potential VMT decrease (e.g., the reduction of searching for parking traffic, ridesplitting such as Uber pool /citeLiRideSplitting2019, serving first/last miles to facilitate transit, etc.) needs to be carefully studied by city transportation managers before an e-HSP related policy is implemented. On the other hand, for e-HSPs, in addition to encouraging more customers to use ridesplitting, they should also try different strategies to reduce deadhead miles caused by e-HSP vehicles. This is particularly critical when the demand pattern is very asymmetric (e.g., peak periods), which is less so when the demand pattern is more symmetric (e.g., some recreational or shopping trips). For this, the proposed model (and its planned extensions as summarized in the next section) can potentially serve as a tool by both transportation planners and e-HSPs for their e-hailing related planning.

9 Conclusions

This paper develops a general economic equilibrium model to study the congestion impact when consumers have the choice to use an e-hailing service provider (e-HSP). The model consists of three interacting modules: e-HSP choice module, customer choice module, and network congestion module, connected by a waiting cost mechanism. Through a novel penalty-based method, we show that there exists an equilibrium solution of the normalized type to the model that is not expected to be unique. The model can be used by transportation planners to understand the adoption levels of these e-HSPs as a function of different cost parameters such as fares and value of time, as well as the resulting impact on congestion as a function of the usage of these types of transportation services. The model can also be used by e-HSPs to understand the impact of their fare structure on their usage and on the overall traffic network congestion. Experimental results show the sensitivity of the relationship between VMT/VHT and the network demand symmetry pattern. That is, when there is little in the symmetry in the network OD demand, the VMT increases significantly with the e-HSP usage due to the increase in deadhead miles. However, as the symmetry increases the impact on deadhead miles significantly reduces with increased e-HSP usage.

Although there has been a significant amount of prior models studying the impact of taxi services on congestion, most of this prior work focused on taxis driving in zones waiting to be hailed by customers (i.e., street-hailing). The model developed in this paper is one of the first attempts in understanding the congestion impacted by e-HSPs which more closely resemble TNCs (such as Uber and Lyft type of providers) and e-hailing taxies. The research community is at the beginning stages of understanding the traveler behavior with the increasing adoption of these transportation services, as well as their overall impact to network congestion. We thus hope that the model proposed in this paper can motivate more research in this important and interesting area. For this, there is a significant amount of future work that can be studied. For example, some of the assumptions made in this paper could be relaxed such as (1) e-HSP drivers behavior in a cooperative manner; (2) the relationship of the waiting times of e-HSP vehicles and their customers are not modeled explicitly; (3) the distance and free-flow travel time variables in the e-HSP profit and customer disutility models are path independent (i.e., based only on the minimum free-flow travel time path which can be pre-determined); (4) the e-HSP drivers have already made a decision to work for a particular provider and for certain work hours; and (5) many e-HSP specific features are not considered such as surcharging, trip cancellation, etc. Regarding the first assumption, we can consider each e-HSP driver as selfish and develop an e-HSP driver model specifically. In this case, we may need to consider both the e-HSP choice behavior (such as dispatching) and e-HSP drivers' choice behavior (such as where to wait for the next customer, trip acceptance/cancellation). Other features such as commission fee (i.e., fare split between e-HSP and its drivers) can also be modeled. The second assumption can be relaxed by explicitly modeling the matching between e-HSP vehicles and customers, for which the methods in [30] for traditional taxis and more recently in [38] for TNCs may provide useful insight. The third assumption can be relaxed by adding a path index to the variable z. This however will increase the dimension of the problem significantly; see the discussions in Section 4. Future research needs to investigate this extended model especially on how to reduce its dimension and improve computational efficiency. Relaxing the fourth assumption will require the development of a (bilevel) optimization model to capture how the driver selects a provider and/or his/her work hours. The fifth assumption can be relaxed by specially modeling an e-HSP platform and its drivers' behaviors related to surcharging, trip cancellation, etc. This paper also focuses on identifying a user equilibrium solution and we show there exists such a solution to the model. However, there may be multiple equilibrium solutions, which may have contributed to the fluctuations of the results in the numerical section. Therefore, it would be interesting to study if there might be some conditions for the model to possess a unique equilibrium solution.

In the proposed model, we assume e-HSP vehicles only serve a single group of customers for a given OD pair, i.e., no ridesplitting is modeled. Although this is largely the case in the current situation (the use of ridesplitting services is pretty low in current e-HSPs [17]), when e-HSPs become the dominant transportation mode in a network in the future (e.g., if automated vehicles are mature and become dominant in the market), ridesplitting will be probably the norm of urban transportation. To capture this, the proposed model needs to be extended to integrate ridesplitting/ridesharing network models, e.g., the one in [28]. Furthermore, we assume that every traveler has access to a car, which may not be realistic. In future research, this can be relaxed by modeling travelers into two groups as those with and without access to cars. The group that does not have access to cars will not have the solo-driving option. We also assume that the total travel demand that uses solo and e-HSP modes for an OD pair is given and fixed. In many cities, transit services also play an important role in meeting travel demands. Future research may consider this by either including transit directly as an additional mode or making the total demand for solo and e-HSP modes elastic (probably dependent on the minimum disutility of the OD pairs). If transit is included as an additional mode, transfers among solo-driving, e-HSPs, and transit also need to be explicitly modeled.

All these extensions need to be tested and further developed using real-world transportation networks and e-hailing data. This is currently challenging due to the lack of such data and the general reluctance of e-HSPs for sharing their data. Another challenge related to testing the proposed (and extended) models on real world transportation networks is related to the path-based formulation of the model. Being quite intuitive and adhering to Wardrop's user equilibrium principle, the path-based formulation is quite common in modeling traffic equilibrium, which is adopted in our model as a starting point in formulating a general equilibrium model of the emergent transportation system with e-hailing services. The computational challenge of path-based models is well recognized. The present paper focuses on one version of the model formulation, a proof of the existence of an equilibrium, and some preliminary numerical results to support the viability of the model. In future research, we need to develop specific algorithms to solving large-scale problems. For this, link-based formulations or techniques such as column generation to deal with potentially large number of paths may be applied. Last but not least, we also need to extend the proposed model to dynamic transportation networks to better capture traffic dynamics and more importantly to incorporate operational decisions of e-HSPs (such as dispatching and surcharging), e-HSP drivers (such as decisions on where to wait to pick up the next customer), and customers. Results on this may be reported in subsequent papers.

Acknowledgements

The authors would like to thank the five anonymous referees for their insightful comments that helped improve the early versions of the paper. Mr. Jingxing Wang from the University of Washington helped produce some of the numerical results; his effort is highly appreciated. The work of Xuegang (Jeff) Ban and Rong Fan is based on research supported by the National Science Foundation under Grants CMMI-1719551. The work of Maged Dessouky is based on research sponsored by Metrans. The work of Jong-Shi Pang is based on research partially supported by the National Science Foundation under Grant CMMI-1538605. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or

Metrans.

References

[1] F. Alemi and G. Circella and D. Sperling. Adoption of Uber and Lyft in California and Impacts on the Use of Other Travel Modes. In 97th Annual Meeting of the Transportation Research Board, Washington, DC.

[2] X. Ban and S. Lu and M.C. Ferris and H. Liu. Risk-averse second best toll pricing. In W.H.K. Lam, S.C. Wong, H.K. Lo, editors. *Transportation and Traffic Theory*. Springer (2009) pp. 197-218.

[3] X. Ban, M.C. Ferris, L. Tang, and S. Lu. Risk-neutral second best toll pricing. *Transportation Research Part B* 48 (2013) 67–87.

[4] R.W. Cottle, J.S. Pang, and R.E. Stone. *The Linear Complementarity Problem*. SIAM Classics in Applied Mathematics 60, Philadelphia (2009). [Originally published by Academic Press, Boston (1992).

[5] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *4OR* 5 (2007) 173-–210.

[6] F. Facchinei and J.S. Pang. Nash equilibria: The variational approach. In Y. Eldar and D. Palomar, editors. *Convex Optimization in Signal Processing and Communications*. Cambridge University Press (Cambridge 2009) pp. 443–493.

[7] F. Facchinei and J.S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer (New York 2003).

[8] M.C. Ferris, and T.S. Munson. PATH 4.6 User Manual. *University of Wisconsin at Madison* (1998).

[9] M.C. Ferris and R.J.B. Wets. MOPEC: multiple optimization problems with equilibrium constraints. `http://www.cs.wisc.edu/~ferris/talks/chicago-mar.pdf`

[10] M. Furuhata, M. Dessouky, F. Ordóñez, M. Brunet, X. Wang, and S. Koenig. Ridesharing: the state-of-the-art and future directions. *Transportation Research Part B: Methodological* 57 (2013) 28–46.

[11] S.A. Gabriel and D. Bernstein. The traffic equilibrium problem with nonadditive path costs. *Transportation Science* 31 (1997) 337–348.

[12] GAMS Website. Internet Link (Accessed on August 30, 2016): `http://www.gams.com` (2016).

[13] P.T. Harker. Multiple equilibrium behavior on networks. *Transportation Science* 22 (1988) 39–46.

[14] A. Haurie and P. Marcotte. On the relationship between Nash–Cournot and Wardrop equilibria. *Networks* 15 (1985) 295–308.

[15] F. He and Z.J.M. Shen. Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C* 58 (2015) 93–106.

[16] Z.Q. Luo, J.S. Pang, and D. Ralph. *Mathematical Programs With Equilibrium Constraints*. Cambridge University Press (Cambridge, England 1996).

[17] W. Li, Z. Pu, Y. Li, and X. Ban. Characterization of ridesplitting based on observed data: A case study of Chendu, China. *Transportation Research Part C* 100 (2019) 330–353.

[18] Y.M. Nie. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. *Transportation Research Part C* 79 (2017) 242–256.

[19] J. Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society* 33 (1965) 520–534.

[20] Schaller Consulting. Unsustainable? The Growth of App-Based Ride Services and Traffic, Travel and the Future of New York City (Accessed on April 20, 2018) `http://schallerconsult.com/rideservices/unsustainable.htm` (2017).

[21] D.A. Schiro, J.S. Pang, and U.V. Shanbhag. On the solution of affine generalized Nash equilibrium problems with shared constraints by Lemke's method. *Mathematical Programming, Series A* 146 (2013) 1–46.

[22] Seattle Uber Prices Internet Link (Accessed on August 30, 2016) `http://uberestimate.com/prices/Seattle/` (2016).

[23] Taxi fares: How much does a ride cost? Internet Link (Accessed on August 30, 2016): `http://www.seattle.gov/your-rights-as-a-customer/file-a-complaint/taxi-for-hire-and-tnc-complaints/taxi-fares-how-much-does-a-ride-cost` (2016).

[24] Transportation Network Testing Problem – Sioux Falls. Internet Link (Accessed on August 30, 2016): `https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls` (2016).

[25] K.I. Wong, S.C. Wong, and H. Yang. Modeling urban taxi serices in congested road networks with elastic demand. *Transportation Research Part B* 35 (2001) 819–842.

[26] K.I. Wong, S.C. Wong, H. Yang, and J.H. Wu. Modeling urban taxi services with multiple user classes and vehicle modes. *Transportation Research Part B* 42 (2008) 985–1007.

[27] H. Xu, F. Ordóñez, and M.M. Dessouky. A traffic assignment model for a ridesharing transportation market. *Journal of Advanced Transportation* 49 (2015) 793–816.

[28] H. Xu, J.S. Pang, F. Ordóñez, and M.M. Dessouky. Complementarity models for traffic equilibrium with ridesharing. *Transportation Research Part B* 81 (2015) 161–182.

[29] J.S. Pang and G. Scutari. Nonconvex games with side constraints. *SIAM Journal on Optimization* 21 (2011) 1491–1522.

[30] H. Yang, C.W.Y. Leung, S.C. Wong, and M.G.H. Bell. Equilibria of bilateral taxi-customer searching and meeting on networks. *Transportation Research Part B* 44 (2010) 1067–1083.

[31] H. Yang, and S.C. Wong. A network model of urban taxi services. *Transportation Research Part B* 32 (1998) 235–246.

[32] H. Yang, S.C. Wong, and K.I. Wong. Demand-supply equilibrium of taxi services in a network under competition and regulation. *Transportation Research Part B* 36 (2002) 799–819.

[33] H. Yang, and T. Yang. Equilibruim properties of taxi markets with search frictions. *Transportation Research Part B* 45 (2011) 696–713.

[34] H. Yang, and H.J. Huang. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. *Transportation Research Part B* 38 (2004) 1–15.

[35] H. Yang, X. Zhang, and Q. Meng. Stackelberg games and multiple equilibrium behavior on networks. *Transportation Research Part B* 41 (2007) 841—861.

[36] Y. Yang, Y. Yin, and H. Lu. Designing emission charging schemes for transportation conformity. *Journal of Advanced Transportation* 48 (2007) 766–781.

[37] X. Yang, R. Ma, and X. Ban. Mixed equilibria with common constraints on transportation networks. *Networks and spatial economics, accepted* (2016).

[38] L. Zha, Y. Yin, and Z. Xu. Geometric matching and spatial pricing in ride-sourcing markets. *Transportation Research Part C, under review* (2018).

Appendix I: Relation between vehicle and customer waiting times

In this appendix, we test if a relation exists for the vehicle and customer waiting times modeled in this paper. The test pertains to a Cobb-Douglas type relation between these two sets of waiting times as per the (taxi) model in [30, 31, 32, 33]. For this, we first define $\widetilde{w}_k^m = w_k^m + \widehat{\lambda}_k^m = w_k^m + \gamma_3^m \lambda_k^m$ as the total waiting cost of customers of OD pair k who use e-HSP$_m$ services. Then we aim to ascertain the validity of the following equations [30]:

$$(\widetilde{w}_k^m)^{\eta_1} (\widehat{w}_k^m)^{\eta_2} = \frac{1}{A_k} (Q_k^m)^{\frac{1}{2}-\eta_1} \left(\sum_{j \in \mathcal{D}} z_{jk}^m \right)^{\frac{1}{2}-\eta_2}, \quad \forall\, m > 0, \tag{32}$$

for some positive constant exponents η_1 and η_2 and "location-specific parameter" A_k, which we have taken to be OD-pair dependent. Mathematically, this type of power law is based on the underlying assumption that all variables involved in the power terms are positive; since these variables are derived from a very complex model, one has to be careful to ascertain whether this basic requirement is met in order for such a law to be applicable. Notice also that $\widetilde{w}_k^m$ actually represents the customer's waiting cost, which however should have the same trend with the customer's waiting time. Therefore, $\widetilde{w}_k^m$ is used in (32) and later the fitting model (34).

In what follows, we propose to test this power law based on the solutions obtained from our equilibrium model. Specifically, the question we ask is the following: with the customer waiting times $\widetilde{w}_k^m$, the customer trip demands Q_k^m, and e-HSP service choices z_{jk}^m computed from the TGEM-ets model, and adopting a small tolerance $\varepsilon > 0$ to guard against possible zero quantities within the logarithmic terms below, do there exist nonnegative constants η_1 and η_2 (properly restricted), and e-HSP waiting times $\widehat{w}_k^m$ such that (7) holds and

$$\begin{aligned} & -\eta_1 \log (\widetilde{w}_k^m + \varepsilon) - \eta_2 \log \widehat{w}_k^m + \left(\tfrac{1}{2} - \eta_1 \right) \log (Q_k^m) + \left(\tfrac{1}{2} - \eta_2 \right) \log \left(\sum_{j \in \mathcal{D}} z_{jk}^m + \varepsilon \right) \\ & \quad \approx \text{ an OD-dependent constant } A_k, \quad \forall\, m > 0. \end{aligned} \tag{33}$$

We address this question by considering the following constrained nonlinear least-squares minimization problem for each m and lower and upper bounding the constants η_1 and η_2 by 0 and 1 respectively:

$$\begin{aligned} \underset{\eta_1,\, \eta_2,\, \widehat{w}_k^m}{\text{minimize}} \quad & \sum_{k \in \mathcal{K}} \left(\begin{array}{l} -\eta_1 \log (\widetilde{w}_k^m + \varepsilon) - \eta_2 \log \widehat{w}_k^m + \\ \left(\tfrac{1}{2} - \eta_1 \right) \log (Q_k^m) + \left(\tfrac{1}{2} - \eta_2 \right) \log \left(\displaystyle\sum_{j \in \mathcal{D}} z_{jk}^m + \varepsilon \right) - A_k \end{array} \right)^2 \\ \text{subject to} \quad & \text{(7) on the waiting times } \widehat{w}_k^m \\ & 0 \leq \eta_1 \leq 1, \quad \text{and} \quad 0 \leq \eta_2 \leq 1. \end{aligned} \tag{34}$$

This least-squares approach can easily be adopted to the situation where the parameters A_k are not fixed a priori; in this case, A_k can be taken to be an additional variable in the above least-squares minimization.

Using the small network in Figure 2 , we show that such a power-law relation roughly holds for the vehicle and customer waiting times calculated from the proposed model. In the test, we use the sensitivity results obtained by using different values of the parameters of the model to produce Table 4–Table 7. For each set of parameters, the TGEM-ets model is solved. We then use the solution from TGEM-ets as the input to solve the fitting model (34). Figure 5 shows the distribution of the objective values after solving all the fitting problems. These values are very close to zero, suggesting that a power law of the Cobb-Douglas type (32) could remain a good approximation of the relation between the e-HSP vehicles' waiting times and their customers' waiting times, as in the case postulated for a taxis-only model.

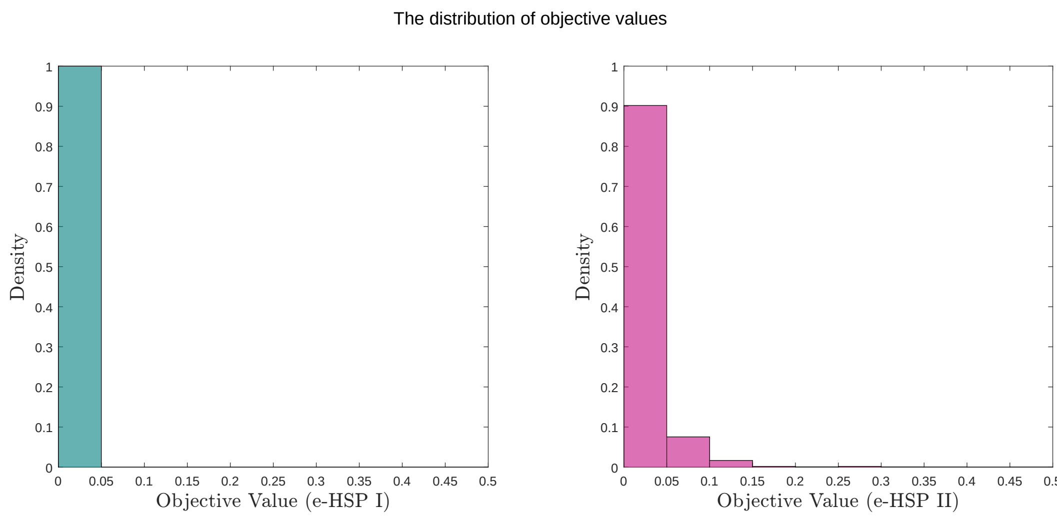

Figure 5: Distribution of the objective values (small network)

Solving the fitting model (34) for each TGEM-ets problem will result in a coefficient pair (η_1 and η_2). Figure 6 and Figure 7 show, respectively, the η_1 and η_2 parameters in the waiting time function (32) for the two e-HSP modes. Interestingly, these values are clustered around a few critical values (e.g., for e-HSP I, they are 0, 0.2, 0.4 for η_1 and 0, 0.1, 0.3 for η_2). While Figure 5–Figure 7 indicate a rather good fitting of the expression (32), there are some noticeable differences from the relationship established in the past for traditional street-hailing taxis. For example, η_1 and η_2 are not the same for different sets of parameters. At this time, our investigation of the relation of the two sets of waiting times is very preliminary; it appears that there are some similarities and differences between the waiting times of e-hailing services and those of the street-hailing taxi services. Further investigation of such a relation is important and deserves to be pursued in future research.

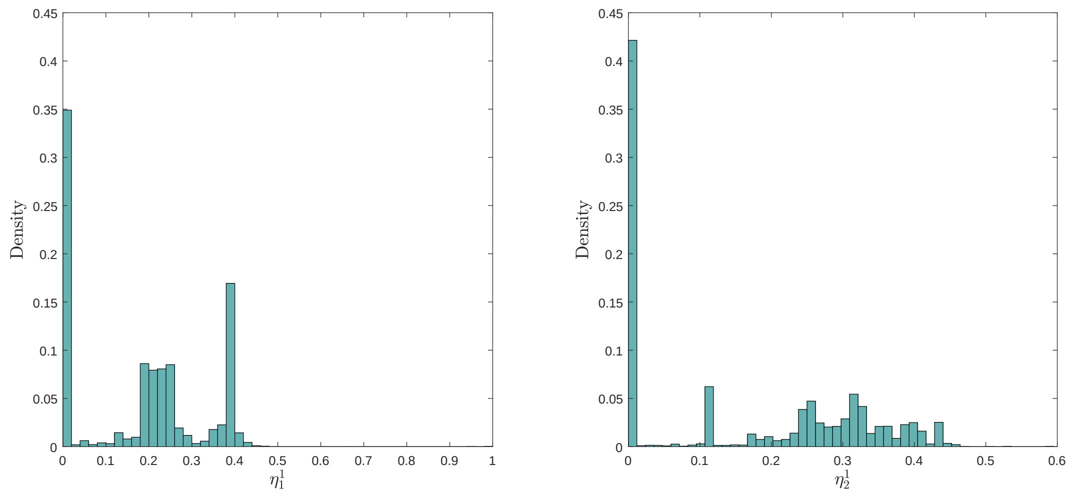

Figure 6: Waiting time fitting results (small network, e-HSP I)

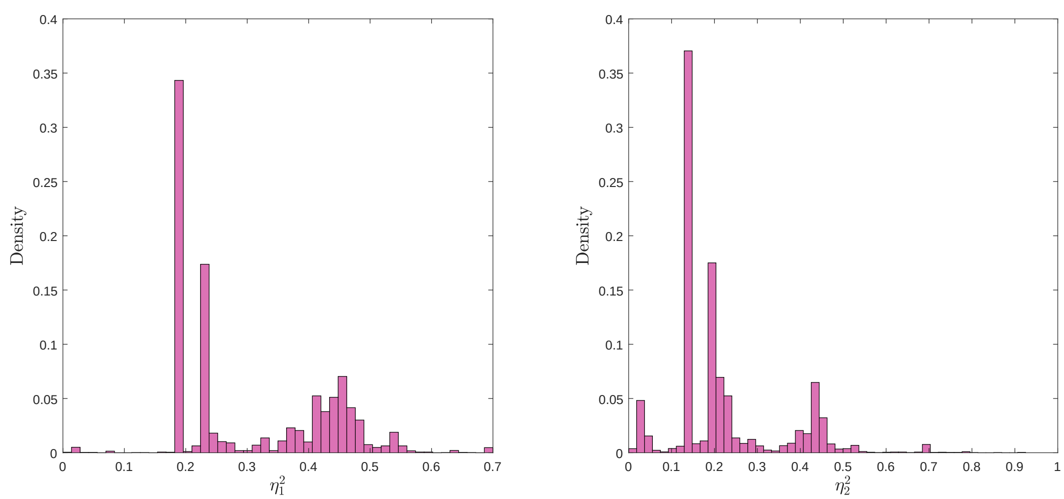

Figure 7: Waiting time fitting results (small network, e-HSP II)