Abstract
The emergence of Transportation Network Companies (TNCs) such as Uber, Lyft, and DiDi where customers request service through e-hailing has had a tremendous impact on the travel behavior of individuals. However, there is still a lack of models and tools to quantify the expected use of the TNCs and the resulting impact on congestion. Adoption of these types of services has 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). In addition, these TNCs have had a significant negative impact on the demand for taxi services. The objective of this paper is to develop a basic mathematical model to help transportation planners understand the relationship between the wide-scale adoption of these 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) that include emergent TNCs and e-hailing taxi services. The equilibrium model consists of three interacting models: TNC 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 TNC’s decisions, thereby connecting the supply and demand sides of the equilibrium. We show the existence of an equilibrium under certain 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
A set of nascent transportation industries, such as shared mobility (including transportation network companies (TNCs) such as Uber, Lyft, Didi and ridesharing/carsharing) and smart parking apps, 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. For example, claims by TNCs that they are part of the solution to reduce congestion are difficult to assess and quantify due to lack of traffic models that incorporate their capabilities. Although trips 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 transportation service providers 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 there is a driver closer to the origin location 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 [16]), which imposes increasing pressure to municipalities on managing and regulating those emerging mobility services. The objective of this paper is to develop a basic model to help transportation planners understand the relationship between the wide-scale adoption of these services and deadhead miles and resulting impact on congestion.

We note that these emerging TNC companies are often mistakenly referred to as ridesharing transportation companies. However, at least in the current stage, they operate quite differently from ridesharing companies such as Carma and BlaBlaCar where the incentive is for the driver to partially recover the cost of their trip through ridesharing as opposed to taking on customers for profit maximization [9]. The focus on this paper is in the latter type of companies, e.g. TNCs, that are becoming an increasingly important part of the current e-hailing transportation service industry.

In addition to having an impact on congestion, TNCs have had a significant negative impact on the demand of taxi services; in fact the distinction between TNCs and taxi companies are becoming less clear. Nevertheless, differences between TNCs and traditional taxi companies still exist; the latter are fully regulated, must carry commercial insurance coverage, have strict background checks and have high entrance costs for example. In this paper, we assume that the use of taxis is only through e-hailing, while standing on the street and hailing a taxi while it is driving in its zone is not permitted. Clearly in certain cities such as New York City the primary demand for taxis is directly hailing the taxi while it is driving, whereas in most other cities in the US (such as Los Angeles and Seattle), there are limited taxi trips of this type and the primary use is through e-hailing. Despite being perceived as more expensive, there are many individuals who still use taxis even in an e-hailing mode over TNC services, most likely due to a number of reasons such as security, convenience, etc. These new transportation services and e-hailing taxi companies are referred to as e-hailing service providers (e-HSPs) in this paper. 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, the latter consisting of the emergent TNC services plus e-hailing taxi services. Each of these companies acts as an economic agent of the system selfishly optimizing their companies' objectives. Comprising the supply side of the transportation system, the TNCs have very different fare structures from the taxis. 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 who responds (or chooses 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 [13, 12, 30]. 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 [13, 30], linearity [31], or joint convexity [32]. 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 (7), making the resulting model a Nash equilibrium problem of the generalized type; see [5]. 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 wait times of the arrival of the 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 main 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.
The proposed equilibrium model consists of three interacting sub-models (also called modules). The first module describes the behavior of the e-HSPs who aim to maximize their profit. The second describes the consumer choice model assuming that each traveler minimizes his/her individual disutility. The third module 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 model connects the e-HSP provider and customer models by providing a mechanism to calculate customers’ waiting costs. Besides the proposed model formulation, one major contribution of this work is a proof of existence of an equilibrium solution to the formulated TGEM-ets. 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 has been previous studies [26, 20, 27, 21, 25, 28] that have developed network models describing urban street-hailing taxi services in a stable equilibrium state. 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, taxis, and TNCs) 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 [14] 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 [25], 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 differ from [14] in three important ways. First, we consider solo driving, e-hailing taxies, and e-hailing TNCs as different modes, 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 the standard fixed-point theorem to TGEM-ets 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 [14], the waiting times between vehicles and customers were modeled using the same method for street-hailing taxis [25]. Since we only focus on e-hailing services in this paper, we model the waiting times differently as discussed in detail in Assumption (f) in Section 2. Thirdly, while travel times are assumed to be given in [14], we consider congestion effect and allow the travel times to be endogenously determined by the traffic equilibrium sub-module. This leads to the nonconvex inequality vehicle balance constraint that significantly complicates the analysis (such as solution existence) of the model. There also has been some work on developing equilibrium models for traditional ridesharing services (e.g., [23, 22]) where they are not modeled as “for-profit” companies like TNCs. Indeed, drivers in such services are regular travelers who have their own unique origins and destinations and pick up additional customers on their trip for some form of compensation. Such services are not considered in this paper.
The remainder of the paper is organized in eight sections. Section 2 lays down the setting of the model, starting from 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 the model consisting of the e-HSP’s choice module, the customer choice module, and the network congestion module, respectively. Combining all the previous sections, Section 6 formally presents the TGEM-ets and defines an equilibrium solution. In Subsection 6.1 we address the relation of the vehicle and customer waiting times; we give a preview of the existence proof in Subsection 6.2. In Section 7, we establish the existence of an equilibrium 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

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 request. 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 in a traffic network consisting of trip makers of different types each with its own objectives. The dynamic operational behavior of the e-HSP vehicles or customers, e.g., unmatched trips or cancelled trips, is not modeled. In addition, we only model basic dispatch behavior of the platform (e.g., which customers to serve and which route to take). Other (more advanced) features, such as surcharges and incentives, are not considered.

(b) There is a given number of vehicles for each e-HSP type (taxi or TNC) to serve customer requests; this number is balanced with the e-HSP trips; see equation 6.

(c) The e-HSPs provide paid ride services via e-hailing; no e-HSP vehicle drives around to seek a customer. This is an important feature of e-hailing vehicles compared with traditional taxis (that often drive around for customers). In practice, after serving a previous demand, an e-HSP vehicle may remain at the drop-off location or drive to a preferred location to wait for the next service request. That is, it is a two-step decision: where to wait (i.e., the preferred location) and which service call to respond next. In this paper, we simplify this two-step decision to a one-step decision by assuming that a driver will stay at
the drop-off location and wait for the next service call. To exactly capture drivers’ two-step decisions, we will need to model drivers’ choices of their preferred location which may be done in future research.

(d) 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 (taxi or TNC) 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. That is, the e-HSP drivers have the incentive to make trips in order to gain future demands by the provider. Under this cooperative behavior assumption, the e-HSP is identified with the respective e-HSP vehicles/drivers as one entity. We realize that this assumption may not be entirely valid as individual drivers can indeed have their own choices of accepting or rejecting ride requests, and where to wait for the next customer. A more sophisticated modeling approach should consider an e-HSP and its vehicles/drivers separately with their interactions properly described. Such a model will be more complex. As a first step in modeling the equilibrium state of the traffic system with e-hailing services, we leave such advanced features for refined modeling in future research.

(e) Customers decide which service to use (or drive alone) and then request the service (if not driving alone) via the mobile platform. We assume they utilize a single travel mode (solo, or taxi, or TNC) for an entire trip, and no transfers between modes are modeled.

(f) We model the waiting times of e-HSP vehicles and customers separately. The e-HSP waiting times are internally determined, as shown in equation (6). The customers’ waiting times are modeled as the sum of the matching time with e-HSP vehicles and the traveling time 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 [25]). 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 [33]. 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 separately. However we explore the relation between the e-HSP vehicle and customer waiting times empirically by testing a Cobb-Douglas type connection between these times similar to that previously employed for taxis (see Subsection 6.1).

(g) 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.

(h) 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 (11).

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. On one hand, if we build a model to adhere to every single aspect of practice, we will end up with such a large and complicated mathematical model that no rigorous analysis can possibly be performed to produce useful insights; any attempt to obtain a “solution” of such a model is at best a heuristic. On the other hand, an analytically trivial model would be a much over-simplification of practice and become a simple mathematical exercise of modeling and analysis. So, a compromise has to be reached. In this sense, we recognize that the above assumptions may be simplifications of realism. Nevertheless, as a first attempt to model the emergent e-hailing industry with various system participants and the complexity of vehicular flows in a congested traffic network, the model has achieved a reasonable compromise between these two principles. See the Conclusion Section 9 for discussion on future research to relax some of these assumptions.

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. For notational convenience, we include the solo drivers as an e-HSP; they are distinguished from the other e-HSPs by their cost structures, which are different among the e-HSPs themselves. There is only one state of the solo drivers, they just drive and do not wait or pick up customers. 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 the customer 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 customer already owns or leases a vehicle. Thus, the fixed cost of all driving modes is assumed to be zero. Also no direct fare is paid for solo drivers. Travelers taking an e-hailing service are called e-HSP customers; their values of time will be impacted by the security, comfort level, loss of productivity, and so on by using a particular e-HSP service. We distinguish two types of time values, due to waiting and traveling. For an e-HSP customer, the waiting is for the arrival of the e-HSP vehicle; for a solo driver, the waiting is zero. Hereafter in this paper, we focus on one-hour study period when we develop the network model with e-hailing services.
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 $\mathcal{N}$; $\mathcal{O} = \{ O_k : k \in \mathcal{K} \}$
- **$\mathcal{D}$**: set of destinations, subset of $\mathcal{N}$; $\mathcal{D} = \{ D_k : k \in \mathcal{K} \}$

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. See Assumption (c) above.

- **$O_k$, $D_k$**: origin and destination (sink) respectively of OD pair $k \in \mathcal{K}$
- **$\mathcal{M}$**: labels of the e-HSPs; $\mathcal{M} \triangleq \{ 1, \ldots , M \}$; in the numerical experiments, we take $\mathcal{M} = \{ 1, 2 \}$ where 1 is for taxi and 2 for TNC
- **$\mathcal{M}_+$**: union of the solo driver label (0) and the e-HSP labels; thus $\mathcal{M}_+ = \{ 0 \} \cup \mathcal{M}$
- **$\mathcal{M}_m$**: 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:

- a OD-path if $i = O_k$ and $j = D_k$ for some $k \in \mathcal{K}$, thus $p$ is a path serving OD pair $k$;
- an e-HSP-path if $i = D_k$ for some $k \in \mathcal{K}$ and $j = O_\ell$ 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 that extends beyond the classical traffic equilibrium problem where only the paths joining OD pairs are of interest. Here we need to be concerned also with the e-HSP-paths that join destinations to origins because these are the pick-up trips made by e-HSP vehicles, something that is absent in the classical problem. We could think of the e-HSP-paths as OD paths provided we accept the fact that the origin of an e-HSP-path is a destination node of an OD-pair and the destination of the e-HSP-path is the origin node of (the same or a different) OD-pair.

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 broader 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
\[
\epsilon_o^{ik} = \begin{cases} 
1 & \text{if } i = O_k \\
0 & \text{otherwise}
\end{cases}
\text{ and } \epsilon^d_{ik} = \begin{cases} 
1 & \text{if } i = D_k \\
0 & \text{otherwise}
\end{cases}
\]
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
\[
\delta_{ij}^{\text{OD}} \triangleq \epsilon_o^{ik} \epsilon_d^{jk} = \begin{cases} 
1 & \text{if } i = O_k \text{ and } j = D_k \\
0 & \text{otherwise}
\end{cases}
\]
similarly, for a pair of nodes \((i', j') \in \mathcal{N} \times \mathcal{N}\) and two OD-pairs \(k\) and \(\ell\) in \(\mathcal{K}\), let
\[
\delta_{i'j'k\ell}^{\text{e-HSP}} \triangleq \epsilon_d^{i'k} \epsilon_o^{j'\ell} = \begin{cases} 
1 & \text{if } i' = D_k \text{ and } j' = O_\ell \\
0 & \text{otherwise}
\end{cases}
\]
The two composite indicators \(\delta_{ij}^{\text{OD}}\) and \(\delta_{i'j'k\ell}^{\text{e-HSP}}\) 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 \(\delta_{i'j'k\ell}^{\text{e-HSP}} = 1\) means that the OD demand \(\ell\) will be served by e-HSP drivers coming from \(i'\). The variable \(z^{m}_{i'\ell}\) to be introduced later will quantify the number of e-HSP vehicles responding to such a service.

**Model parameters.** The following are constants in the model:

- \(Q_k\) given travel demand rate (trips per unit time) of OD pair \(k \in \mathcal{K}\)
- \(C^m\) average fixed daily cost paid by an e-HSP\(_m\) driver for \(m \in \mathcal{M}\);
- \(f_{ij}^0\) (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 \(j \in \mathcal{N}\)
- \(F^m_{O_k}\) fixed fare charged by e-HSP\(_m\) who picks up customers at \(O_k\) (the origin of OD pair \(k\))
- \(\alpha^m_1, \alpha^m_2\) time and distance based fare rates, respectively, for e-HSP\(_m\)
- \(\beta^m_1, \beta^m_2\) respective conversion factors from travel time and distance to costs for e-HSP\(_m\) drivers
- \(\gamma^m_1\) value of time for e-HSP\(_m\) customers while traveling in a vehicle
- \(\gamma^m_2\) value of time for e-HSP\(_m\) customers while waiting for a vehicle
- \(\gamma^m_3\) disutility of extra cost for e-HSP\(_m\) customers due to matching with e-HSP vehicles
- \(N^m\) fleet size (i.e., the number of vehicles) of e-HSP\(_m\).

We have \(F^0_{O_k} = 0\) and \(\alpha^0_1 = \alpha^0_2 = 0\) as solo drivers do not need to be paid these fares. The parameters \(\alpha^m_i\) and \(\beta^m_i\) are employed to calculate the e-HSP\(_m\)'s profit. There are some reasonable relations among them across the e-HSPs. For instance, to capture the different characteristics of conventional taxis and
TNC vehicles, we may have \( \alpha_1^1 \geq \alpha_1^2 \) and \( \beta_1^1 = \beta_1^2 \) (where \( m = 1 \) refers to taxis and \( m = 2 \) refers to TNC), and \( \alpha_2^1 \geq \alpha_2^3 \) and \( \beta_2^1 \leq \beta_2^3 \). This relationship covers the conventional thought that taxis are more expensive but safer. We also set \( \beta_2^3 = \beta_2^0 \) to indicate that the distance-based cost for solo driving is the same as that for TNC services (and larger than that for taxi services).

**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” in [26]) from destination \( j \in D \) 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 \( i \in N \) to node \( j \in N \)
- \( h_p \): vehicular flow on path \( p \in \mathcal{P} \).

Two particular cases of the notation \( z^m_{jk} \) are worthy of note. First is the possibility that the node \( j = D_k \); 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 \( j = D_\ell = O_k \) for some OD pairs \( \ell \neq k \); this signifies that the e-HSP drivers at the destination of OD-pair \( \ell \) 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:

\[
DH \triangleq \sum_{m \in M} \sum_{j \in D, k \in K} z^m_{jk} d_{jO_k}.
\]  

The total number of e-HSP\(_m\) vehicles available at node \( j \in D \) for the study period (one hour in this paper) to pick up customers is thus given by the sum \( \sum_{k':j=D_{k'}} Q^m_{k'} = \sum_{k' \in K} \varepsilon^d_{jk'} Q^m_{k'} \); see also equation (7).

We have the following definitional relations between the (variable) e-HSP demand variables \( Q^m_k \) and the total known demand \( Q_k \):

\[
Q_k = \sum_{m \in M^+} Q^m_k = \sum_{m \in M} Q^m_k + Q^0_k.
\]  

OD demands served by e-HSP drivers

Trips made by solo drivers

**Derived model variables.** Besides the link flow variables, there are two types of derived variables. For \( m > 0 \), the variables \( R^m_{jk}, U^m_k, u_k, \) and \( w^m_k \) are employed in the modules describing the e-HSP choice module, the customer choice model, and the “market clearance” conditions, respectively; the remaining variables \( \phi^m_j, \lambda^m_k, \) and \( \mu_k \) are dual variables of the model constraints with the latter two having the interpretation of marginal prices. Specifically, we have
flow on link \( a \in A \)

profit (in $) of e-HSP \( m \) vehicles that are at destination \( j \) and plan to serve customers of OD pair \( k \)

disutility ($ per trip) of e-HSP \( m \) customers of OD pair \( k \)

minimum disutility ($ per trip) of e-HSP customers of OD pair \( k \) across all modes

waiting cost ($ per trip) of customers of OD pair \( k \) to be served by e-HSP \( m \)

waiting time of e-HSP \( m \) vehicles serving customers of OD pair \( k \); \( m \) denotes the waiting cost

shadow price of the e-HSP vehicles balancing constraint at location \( j \in D \) perceived by the e-HSP \( m \) drivers and their customers, respectively

marginal price of OD demand \( k \) perceived by customers of OD pair \( k \) to be served by e-HSP \( m \); it is interpreted as the customer matching time with e-HSP vehicles; this variable is added to the customer disutility through \( w^m_k \) as an extra waiting cost

dual variable of the last constraint in e-HSP choice module (7).

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,

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

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 \( h \)
- \( c_a(f) \) travel time of link \( a \in A \), a function of link flow vector \( f \triangleq (f_a)_{a \in A} \);
  
  an example of such a link cost is the BPR type function.

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

\[
\psi_{ap} \triangleq \begin{cases} 
1 & \text{if path } p \text{ uses link } a \\
0 & \text{otherwise}
\end{cases} \quad \text{for all } (a,p) \in A \times \mathcal{P}
\]

and letting \( C(h) \triangleq (C_p(h))_{p \in \mathcal{P}} \) and \( c(f) \triangleq (c_a(f))_{a \in A} \) 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 [10]. Analogous to the link-path matrix \( \Psi \), define the node-path matrix \( \Omega \) with entries

\[
\omega_{(ij),p} \triangleq \begin{cases} 
1 & \text{if } p \in \mathcal{P}_{ij} \\
0 & \text{otherwise}
\end{cases} \quad \text{for all } (i,j,p) \in \mathcal{N} \times \mathcal{N} \times \mathcal{P}.
\]

This matrix \( \Omega \) comes in handy when connecting the travel times \( t_{ij} \) with the path costs \( C_p(h) \).
The free-flow travel times \( f_{ij}^0 \) is by definition the minimum travel costs \( C_p(0) \) over all paths connecting nodes \( i \) to \( j \); i.e.,
\[
f_{ij}^0 = \min_{p \in P_{ij}} C_p(0) \leq \min_{p \in P_{ij}} C_p(h), \quad \text{for any } h \geq 0,
\]
(4)
where the inequality holds in the additive model provided that each link cost function satisfies the natural condition: \( c_a(f) \geq c_a(0) \) for all \( f \geq 0 \). 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
\[
\theta_{jk}^m, \quad \zeta_{jk}^m \quad \text{artificial variables employed to handle the ambiguity of the undefined ratio 0/0 in the calculation of the waiting costs; see Section 4.1}
\]

### 3 e-HSP Choice Module

The e-HSP choice module focuses on the optimization of an e-HSP (TNC or taxi company employing only e-hailing services). As previously mentioned, we assume the drivers employed by an e-HSP are cooperative towards the company’s objective. We further assume a given number of vehicles of the e-hailing companies, denoted as \( N^m \) for e-HSP type \( m \). Since we focus on one-hour study period, the total e-HSP\(_m\) vehicle service time is \( N^m \times 1 = N^m \) vehicle-hours. This is used in the following equation (6) to balance e-HSP vehicle’s occupied times, vacant times, and waiting times. We then introduce the waiting time of an e-HSP vehicle serving an OD pair \( k \), denoted as \( \hat{w}^m_k \). 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:
\[
R_{jk}^m \triangleq \hat{R}_{jk}^m - \beta_3^m \hat{w}^m_k,
\]
(5)
where

\[
\hat{R}_{jk}^m = F_{O_k}^m - \beta_1^m (t_{jO_k} + t_{O_kD_k}) - \beta_2^m (d_{jO_k} + d_{O_kD_k}) + \alpha_1^m (t_{O_kD_k} - f_{O_kD_k}^0) + \alpha_2^m d_{O_kD_k}.
\]

\begin{align*}
\text{travel time} & \quad \text{based cost} \\
\text{travel distance} & \quad \text{travel based cost} \\
\text{time based revenue} & \quad \text{distance based revenue}
\end{align*}

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 \( \hat{w}_{jk}^m \) satisfy the constraints [26]:

\[
\sum_{k \in K} \left[ \sum_{j \in D} z_{jk}^m \right] \hat{w}_{jk}^m = N_k^m - \sum_{k \in K} \sum_{j \in D} z_{jk}^m t_{jO_k} - \sum_{k \in K} Q_k^m t_{O_kD_k}, \quad \hat{w}_{jk}^m \geq 0. \tag{6}
\]

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 \( \hat{w}_{jk}^m \) 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. Notice here that we model the e-HSP vehicle waiting time \( \hat{w}_{jk}^m \) endogenously in the model, similarly to how the taxi vehicle waiting times were modeled in [26].

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 $m \in M$,

$$\begin{align*}
\text{maximize} & \quad \sum_{k \in K} \sum_{j \in D} \hat{P}_{jk}^m z_{jk}^m - \beta_m^3 \left[ N^m - \sum_{k \in K} \sum_{j \in D} z_{jk}^m t_{jO_k} - \sum_{k \in K} Q_k^m t_{O_k D_k} \right] \\
\text{subject to} & \quad \sum_{j \in D} z_{jk}^m = \sum_{k' : j = D_{k'}} Q_{k'}^m \quad \text{for all } j \in D \\
& \quad \sum_{j \in D} \sum_{k \in K} z_{jk}^m \geq Q_k^m \quad \text{for all } k \in K \\
& \quad \sum_{k \in K} \sum_{j \in D} z_{jk}^m t_{jO_k} + \sum_{k \in K} Q_k^m t_{O_k D_k} \leq N^m.
\end{align*}$$

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$_m$ vehicles supply equation, the constraint

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

(8)

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 [26] 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

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

(9)

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 $N^m$. Exogenous to the above e-HSP module, the OD demands $Q_k^m$, and travel times $t_{jO_k}$ and $t_{O_k D_k}$ are determined in the customer choice module and the network congestion module, respectively, to be described below; the profits $\hat{P}_{jk}^m$ are also exogenous and are determined by (7). The customer choice and network congestion modules provide key linkages among the different components in the TGEM-ets.
Let $\phi^m_j$ denote the un-signed dual variable of the e-HSP vehicle supply equation (8). Let $\lambda^m_k$ denote the signed dual variable of the e-HSP demand constraint (9). Let $\mu^m_k$ 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 (7) are given by the following complementarity conditions:

$$\begin{cases}
0 \leq z_{jk}^m \perp -\hat{R}_{jk}^m - \beta_j^m t_{jO_k} - \phi_j^m - \lambda_k^m + t_{jO_k} \mu^m \geq 0, & \forall (j, k) \in D \times K \\
\phi_j^m \text{ free,} & \sum_{k \in K} z_{jk}^m = \sum_{k' : j = D_k} Q_{k'}^m, & \forall j \in D \\
0 \leq \lambda_k^m \perp \sum_{j \in D} z_{jk}^m - Q_k^m \geq 0, & \forall k \in K \\
0 \leq \mu^m_k \perp N^m - \left[ \sum_{k \in K} \sum_{j \in D} z_{jk}^m t_{jO_k} + \sum_{k \in K} Q_k^m t_{O_k D_k} \right] \geq 0
\end{cases}$$

(10)

The dual variable $\lambda_k^m$ can be viewed as to be proportional to the customer’s matching time with the e-HSP vehicles; for more details see Subsection [11.1]. The right-hand side of the third complementarity in (10) represents the surplus of e-HSP vehicles that plan to serve an OD pair $k$. For some OD pairs, when there is no surplus of e-HSP vehicles for the OD pair, the customers’ matching time with e-HSP vehicles can be positive, i.e., $\lambda_k^m > 0$ or zero (special case). Otherwise, if the surplus is positive, a customer’s matching time will be zero ($\lambda_k^m = 0$). Relating the dual variable $\lambda_k^m$ to (be proportional to) the customer’s matching time 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. However, it resembles some of the basic features of customer waiting times by relating $\lambda_k^m$ with the vehicle surplus as discussed above. In practice, the matching time is larger if there are more customers than the vehicles, which 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 complementary relation between the matching time $\lambda_k^m$ and the vehicle surplus. In a recent study [33], 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 numbers) as we use here.

4 Customer Choice Module

A customer traveling for an OD pair $k$ will need to wait for the e-HSP vehicle and pay the fare if choosing to ride with an e-HSP vehicle, or pay the cost of driving his/her own car. 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:

\[
U^m_k = F_{O_k}^m + \alpha^m_1 (t_{O_kD_k} - f_{O_kD_k}^0) + \alpha^m_2 d_{O_kD_k} + \gamma^m_1 t_{O_kD_k} + \gamma^m_2 t_{jO_k} + \gamma^m_3 t_{jO_k} + \gamma^m_4 t_{jO_k} + \gamma^m_5 t_{jO_k} + w^m_k .
\]  (11)

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

\[
U^0_k = \gamma^0_1 t_{O_kD_k} + \gamma^0_2 d_{O_kD_k} .
\]  (12)

Here the disutility is expressed in monetary values. The first three terms on the right-hand side of equation (11) 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 \( w^m_k \) for \( m \in \mathcal{M} \) due to waiting will be discussed in Section 4.1. If solo driving is selected, the customer disutility includes both distance-based and travel time based costs, as shown in (12).

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 \( f_{O_kD_k}^0 \) are fixed as those of the minimum-free-flow-travel-time paths, while travel time variables \( t_{O_kD_k} \) and \( t_{jO_k} \) 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 PATH that we have employed in the numerical experiments).

We model a customer’s behavior by a simple, aggregate disutility minimization problem with \( U^m_k \) as exogenous variables to determine the travel demands \( Q^m_k \) for \((m,k) \in \mathcal{M} \times \mathcal{K}\).
Customer choice module:

\[
\begin{align*}
\text{minimize} & \quad \sum_{m \in M} \sum_{k \in K} U^m_k Q^m_k \\
\text{subject to} & \quad \sum_{m \in M} Q^m_k = Q_k, \quad \text{for all } k \in K.
\end{align*}
\]

Letting \( u_k \) for \( k \in K \) be the multipliers of the equality constraints, the optimality conditions of the above linear program in \( Q^m_k \) can be cast as the following mixed complementarity conditions:

\[
0 \leq Q^m_k \perp U^m_k - u_k \geq 0 \quad \forall (k, m) \in K \times M_+ \\
u_k \text{ free,} \quad \sum_{m \in M_+} Q^m_k = Q_k \quad \forall k \in K,
\]

which have the same interpretation as 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 \( Q^m_k \) is positive for \( m \in M \)) only if the associated disutility \( U^m_k \) is minimum across all e-HSP choices. This is analogous to the shortest-path route choice in well-known Wardrop user equilibrium models; see e.g. [6, Section 1.4.5].

We remark that the constraints in (7) are different from those in (14). Nevertheless, besides the explicit presence of the model variables \( Q^m_k \) in both these problems, exogenously in (7) and endogenously in (14), as will be seen in expression (15) of the customer waiting time, the dual variable \( \lambda^m_k \) of the e-HSP demand constraints (9) appears in the definition of the waiting disutility \( w^m_k \), and thus will have an impact on the travel demands \( Q^m_k \) satisfying (14). 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 models decouple according to OD pairs \( k \in K \).

4.1 Customer waiting times

The e-HSP customers’ waiting time has two components: one is the travel time for the e-HSP vehicles to come from their previous waiting locations to the pick-up locations; the other component is the customer’ matching time with e-HSP vehicles, which is assumed to be proportional to \( \lambda^m_k \) in (10). We model the former travel time as the average of the travel times of e-HSP vehicles from all possible locations to the origin of an OD pair. Thus we have

\[
w^m_k = \gamma_2^m \frac{\sum_{j \in D} \gamma_j^m t_{jO_k}}{\sum_{j \in D} \gamma_j^m} + \gamma_3^m \lambda^m_k, \quad \text{for all } m = 1, \cdots, M.
\]

Here \( \gamma_3^m \) converts \( \lambda^m_k \) to disutility. Note that the first term in the above equation does not imply that the average e-HSP vehicle travel time to the origin of an OD pair is the average of travel times from all possible destinations to the origin: those destinations that do not have e-HSP vehicles willing to serve the
OD pair will not be counted. Since e-HSP vehicles aim to maximize their profits, they have the tendency to pick up nearby customers if at all possible. In practice, a platform may only inform drivers within a certain distance of a customer for the service request. In this case, equation (15) can be easily revised to include only those \( j \)'s that are within that distance of the origin of \( k \).

In order to properly handle the fraction \( \frac{\sum_{j \in D} z^m_{jk}}{\sum_{j \in D} z^m_{jk}} \) which is not well-defined when the denominator is equal to zero, we let \( \theta^m_{jk} \) 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 then define:

\[
w^m_k = \gamma_2^m \sum_{j \in D} \theta^m_{jk} t_j O_k + \gamma_3^m \lambda^m_k.
\]

The advantage of the scalar \( \theta^m_{jk} \) as defined is due to the following minimization description:

\[
\theta^m_{jk} \in \arg\min_{\theta \in [0,1]} \left\{-z^m_{jk} \theta + \frac{1}{2} \left[ \sum_{j' \in D} z^m_{j'k} \right] \theta^2 \right\}; \quad (16)
\]

in turn, (16) is a simple univariable bounded-constrained convex quadratic program that is characterized by the complementarity conditions:

\[
0 \leq \theta^m_{jk} \perp \left[ \sum_{j' \in D} z^m_{j'k} \right] \theta^m_{jk} - z^m_{jk} + \zeta^m_{jk} \geq 0 \quad (17)
\]

Thus bypassing the derived variable \( w^m_k \), we can express the disutility \( U^m_k \) as: for \( m > 0 \),

\[
U^m_k = F^m_{O_k} + \alpha_1^m \left(t_{O_k D_k} - f^0_{O_k D_k}\right) + \alpha_2^m d_{O_k D_k} + \gamma_1^m t_{O_k D_k} + \gamma_2^m \sum_{j \in D} \theta^m_{jk} t_j O_k + \gamma_3^m \lambda^m_k,
\]

where \( \theta^m_{jk} \) along with \( \zeta^m_{jk} \) satisfy (17).

5 Network Congestion Effects

In the network, 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 \( k \); 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 Model) 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, 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 that balances the flows and demands. The overall congestion module of the TGEM-ets is as follows: for all \((i, j) \in N \times N\),

\[
\sum_{p \in P_{ij}} h_p = \sum_{k \in K} \delta_{ijk} Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ijkl} \sum_{m \in M} z_m^m
\]

(19)

\[
0 \leq h_p \quad \downarrow \quad C_p(h) - t_{ij} \geq 0, \quad \text{for all } p \in P_{ij}.
\]

We 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. Further, as it is known in the literature (see e.g. [6]), the above network congestion module is not necessarily the optimality condition of a minimization problem with a non-additive path cost structure.

The path flow equation for nodes \((i, j) \in N \times N\) 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 \(O \cup D\), we take \(t_{ij} = 0\) and \(h_p = 0\) for all paths \(p \in P_{ij}\). 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. [6, Proposition 1.5.6].

**Proposition 1.** Suppose that the path costs \(C_p(h)\) satisfy \([4]\). Then \([19]\) holds if and only if for all \((i, j) \in N \times N\),

\[
0 \leq t_{ij} \quad \downarrow \sum_{p \in P_{ij}} h_p - \left[ \sum_{k \in K} \delta_{ijk} Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ijkl} \sum_{m \in M} z_m^m \right] \geq 0
\]

(20)

\[
0 \leq h_p \quad \downarrow \quad C_p(h) - t_{ij} \geq 0, \quad \text{for all } p \in P_{ij}.
\]

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

\[
\left[ \sum_{p \in P_{ij}} h_p C_p(h) = 0, \quad h \geq 0 \right] \Rightarrow [h_p = 0 \ \forall \ p \in P_{ij}].
\]

(21)

It suffices to show that \([20]\) implies the first equation in \([19]\). Assume not; then for some \((i, j) \in N \times N\), we have

\[
\sum_{p \in P_{ij}} h_p > \sum_{k \in K} \delta_{ijk} Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ijkl} \sum_{m \in M} z_m^m \geq 0.
\]
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. \( \square \)

For the analysis of the overall model, to be summarized in the next section, it would be convenient to use the complementarity formulation (20) to describe the network congestion, restricting the travel time variables \( t_{ij} \) to be nonnegative in particular.

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,
\[
t_{O_k D_k} = \min_{p \in \mathcal{P}_{ij}} C_p(h) \geq f_{O_k D_k}^0.
\]
This inequality yields that the disutility \( U_m^k \) as given by (18) is always nonnegative provided that the travel times \( t_{O_k D_k} \) are determined according to (19).

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 \( \sum_{k \in \mathcal{K}} z_{jk}^m = \sum_{k' \in \mathcal{K}} z_{jk}^{d_{k'}} Q_k^{m_{k'}} \), it follows that \( \sum_{k \in \mathcal{K}} z_{jk}^m \leq \sum_{k' \in \mathcal{K}} Q_k^{m_{k'}} \) for all \((j, m) \in \mathcal{D} \times \mathcal{M} \). Let
\[
\bar{Q} > \sum_{k \in \mathcal{K}} Q_k \quad \text{and} \quad Q_{\max} > \max_{k \in \mathcal{K}} Q_k.
\]
(22)

It follows from the path flow equation in (19) that for all \( p' \in \mathcal{P}_{ij} \),
\[
h_{p'} \leq \sum_{p \in \mathcal{P}_{ij}} h_p \leq \sum_{k \in \mathcal{K}} \delta_{OD}^{i, j, k} Q_k + \sum_{(k, \ell) \in \mathcal{K} \times \mathcal{K}} \delta_{ij k}^{\ell-\text{HSP}} \sum_{m \in \mathcal{M}} z_{m i \ell} \leq c \sum_{k \in \mathcal{K}} Q_k
\]
for some constant \( c > 0 \). Thus, for a constant \( \bar{h} > c \sum_{k \in \mathcal{K}} Q_k \), we have
\[
h_p < \bar{h}, \quad \forall p \in \mathcal{P}.
\]
(23)
Moreover, we have
\[
t_{i'j'} \leq \max_{(i, j) \in \mathcal{N}} \frac{\max}{p \in \mathcal{P}_{ij}} \frac{\max}{0 \leq h \leq h_1} C_p(h), \quad \text{for all} \quad (i', j') \in \mathcal{N} \times \mathcal{N}.
\]
Let \( \bar{t} \) be greater than the right-hand constant in the last bound. Similar to (23), we have the following bound on the travel times; namely,
\[
t_{i'j'} \leq \max_{(i, j) \in \mathcal{N}} \frac{\max}{p \in \mathcal{P}_{ij}} \frac{\max}{0 \leq h \leq h_1} C_p(h) < \bar{t}, \quad \text{for all} \quad (i', j') \in \mathcal{N} \times \mathcal{N}.
\]
(24)
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:

\[
\begin{align*}
h &\triangleq \{ h_p : p \in \mathcal{P} \}, \\
t &\triangleq \{ t_{ij} : (i,j) \in \mathcal{N} \times \mathcal{N} \}
\end{align*}
\]

satisfies (19) if and only if it is a solution of a variational inequality defined on the bounded rectangle \( \mathcal{H} \triangleq \{ (h, t) \geq 0 : h \leq \bar{h}, t \leq \bar{t} \} \). This assertion is similar to that employed to prove the existence of a solution to the classical traffic equilibrium problem; cf. [6, 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 \( h \) and \( t \), we let

\[
\begin{align*}
z &\triangleq \{ z_{jk}^m : (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \} \\
Q &\triangleq \{ Q_k^m : (k,m) \in \mathcal{K} \times \mathcal{M}_+ \} ,
\end{align*}
\]

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 \( t \) via the network congestion conditions (19), or equivalently their complementarity formulation (20).

6 The Overall Equilibrium Model

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

- with profits given by (5), the e-HSP optimization problems (7) for \( m = 1, \cdots, M \), or their optimality conditions (10);
- with waiting times given by (15) and disutilities given by (18), the customer choice optimization module (13), or its equivalent complementarity formulation (14);
- network congestion modeled by the equilibrium conditions (19).

As mentioned before, the e-HSP\(_m\) waiting times \( \hat{w}_k^m \) can be recovered from expression (6) after a TGEM-ets equilibrium solution is obtained. More about this recovery and its relation to the customer waiting times \( w_k^m \) will be said later.

**TGEM-ets Equilibrium.** By definition, this is a tuple \( \mathbf{w} \triangleq \{ \mathbf{z}, \mathbf{Q}, \mathbf{u}, \mathbf{h}, \mathbf{t} \} \) with \( \mathbf{u} \triangleq \{ u_k : k \in \mathcal{K} \} \), for which there exist

\[
\begin{align*}
\lambda &\triangleq \{ \lambda_k^m : (k,m) \in \mathcal{K} \times \mathcal{M} \} \\
\theta &\triangleq \{ \theta_{jk}^m : (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \}
\end{align*}
\]

such that for each \( m \in \mathcal{M} \), \( \{ z_{jk}^m : (j,k) \in \mathcal{D} \times \mathcal{K} \} \) is an optimal solution of (7) with each \( R_{jk}^m \) given by (5) and \( \{ \lambda_k^m : k \in \mathcal{K} \} \) is a tuple of multipliers of the e-HSP demand constraints; the pair \( (U_k^m, u_k) \) satisfies the complementarity condition in (14) for all \( (k,m) \in \mathcal{K} \times \mathcal{M}_+ \) with \( U_k^m \) given by (18), \( (\mathbf{h}, \mathbf{t}) \) satisfies (19), or equivalently (20), and \( \theta_{jk}^m \) satisfies (16) for all \( (j,k,m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M} \). \( \square \)

6.1 Relation of vehicle and customer waiting times

So far, the e-HSP vehicle waiting times \( \hat{w}_k^m \) have not been included in the TGEM-ets model. Our contention is that these times can be recovered from the vehicle-hours balance constraints (6) for each e-HSP
type; thus in our approach, these times are determined subsequently to the solution of the equilibrium model. This modeling approach is different from that in the literature that postulates a certain Cobb-Douglas type relation between the customer-taxi waiting times together with other components of the (taxi) model; see [25, 26, 27, 28]. However, it is interesting to test if such a relation exists for the vehicle and customer waiting times modeled in this paper. In other words, we aim to see if these two sets of waiting times is postulated to satisfy the following equations [25]:

\[
(w_k^m)^n (\hat{w}_k^m)^{n_2} = \frac{1}{A_k} \left( Q_k^m \right)^{\frac{1}{2} - \eta_1} \left( \sum_{j \in D} z_{jk}^m + \varepsilon \right)^{1 - \eta_2}, \quad \forall m > 0,
\]

for some positive constant exponents \(\eta_1\) and \(\eta_2\) 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 \(w_k^m\) actually represents the customer’s waiting cost, which however should have the same trend with the customer’s waiting time. Therefore, \(w_k^m\) is used in (25) and later the fitting model (27).

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 \(w_k^m\), 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 \(\eta_1\) and \(\eta_2\) (properly restricted), and e-HSP waiting times \(\hat{w}_k^m\) such that (6) holds and

\[
-\eta_1 \log (w_k^m + \varepsilon) - \eta_2 \log \hat{w}_k^m + \left( \frac{1}{2} - \eta_1 \right) \log (Q_k^m) + \left( \frac{1}{2} - \eta_2 \right) \log \left( \sum_{j \in D} z_{jk}^m + \varepsilon \right) - A_k
\]

\[
\approx \text{an OD-dependent constant } A_k, \quad \forall m > 0.
\]

We address this question by considering the following constrained nonlinear least-squares minimization problem for each \(m\) and lower and upper bounding the constants \(\eta_1\) and \(\eta_2\) by 0 and 1 respectively:

\[
\begin{align*}
\text{minimize} & \quad \sum_{k \in K} \left( -\eta_1 \log (w_k^m + \varepsilon) - \eta_2 \log \hat{w}_k^m + \left( \frac{1}{2} - \eta_1 \right) \log (Q_k^m) + \left( \frac{1}{2} - \eta_2 \right) \log \left( \sum_{j \in D} z_{jk}^m + \varepsilon \right) - A_k \right)^2 \\
\text{subject to} & \quad (6) \text{ on the waiting times } \hat{w}_k^m \\
& \quad 0 \leq \eta_1 \leq 1, \quad \text{and} \quad 0 \leq \eta_2 \leq 1.
\end{align*}
\]

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 the numerical section, we show that such a power-law relation roughly holds for the vehicle and customer waiting times calculated from the proposed model, with noticeable differences compared with the waiting times of traditional taxi services; see Section 8.2 for more details.
6.2 Existence proof of an equilibrium: A preview

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 does not ease the proof of existence of an equilibrium because other important bottlenecks are still present.

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 (7) contains the customers’ variables $Q^m_k$ which are decision variables in (13), the resulting non-cooperative game is of the generalized type [4, 5]. 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 [24], or as a “Multiple Optimization Problem with Equilibrium Constraints” (MOPEC) as it is termed in [8]. 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) [6] 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 nonlinear coupling of variables in the constraints;
- the lack of explicit bounds on some of the model variables, particularly the derived variables: the customer waiting time cost $w^m_{O_k}$, the marginal prices $\lambda^m_k$, and the minimum disutilities $u_k$; thus, the basic requirement of a self-map from a compact convex set into itself in a fixed-point theorem cannot be satisfied;
- 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.

7 Existence of an Equilibrium

There are several main steps in the proof. First, we obtain a mixed complementarity formulation of the model. 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. Relying on a feasibility assumption and a couple key lemmas, 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. 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.

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 mixed complementarity conditions:
There are two main complicating factors in the above complementarity formulation: one is the bilinear (thus nonlinear) term \( t_{jO}k \) in the \( z \)-complementarity condition; the second is the lack of symmetry of the multipliers \( \phi^m_j \) and \( \mu^m \) which are present in the \( z^m_{jk} \)-complementarity condition but not in the \( Q^m_k \)-complementarity condition. The latter feature is the result of the generalized nature of the game; i.e., the coupling constraints 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 \( \rho \). Note that while the variables \( z^m_{jk} \) and \( Q^m_k \) are also coupled in the e-HSP demand constraints (9), 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 \( \alpha \), let \( \alpha_+ \triangleq \max(0, \alpha) \) be its nonnegative part. Define, for each \( \rho > 0 \), the vector function (note the multiplicative factor \( \gamma^m_j \) in front of \( R^m_{jk} \) which was not in the original complementarity

\[
0 \leq z^m_{jk} \quad \downarrow -\gamma^m_{jk} - \beta^m_{jk} t_{jO_k} - \phi^m_{jk} - \lambda^m_j + t_{jO_k} \mu^m \geq 0 \quad \forall (j, k, m) \in D \times K \times M
\]

\[
\phi^m_j \text{ free} \quad \sum_{k \in K} z^m_{jk} = \sum_{k' : j = D_k} Q^m_{k'} \quad \forall (j, m) \in D \times M
\]

\[
0 \leq \lambda^m_k \quad \downarrow \sum_{j \in D} z^m_{jk} - Q^m_k \geq 0 \quad \forall (k, m) \in K \times M
\]

\[
0 \leq \mu^m \quad \downarrow N^m - \left[ \sum_{k \in K} \sum_{j \in D} z^m_{jk} t_{jO_k} + \sum_{k \in K} Q^m_k t_{O_k D_k} \right] \geq 0 \quad \forall m \in M
\]

\[
0 \leq Q^0_k \quad \downarrow V^0_k - u_k \geq 0 \quad \forall (k, m) \in K \times M
\]

\[
0 \leq Q^m_k \quad \downarrow V^m_k - u_k + \gamma^m_{jk} \lambda^m_k \geq 0 \quad \forall (k, m) \in K \times M
\]

\[
u_k \text{ free}, \quad \sum_{m \in M_+} Q^m_k = Q_k \quad \forall k \in K
\]

\[
0 \leq t_{ij} \quad \downarrow \sum_{p \in P_{ij}} h_p - \left[ \sum_{k \in K} \sum_{j \in D} \delta^o_{ijk} Q_k + \sum_{(k, \ell) \in K \times K} \delta^o_{ijk} \sum_{m \in M} z^m_{k\ell} \right] \geq 0 \quad \forall (i, j) \in N \times N
\]

\[
0 \leq h_p \quad \downarrow C_{p(i)} - t_{ij} \geq 0 \quad \forall p \in P_{ij}, \forall (i, j) \in N \times N
\]

\[
0 \leq \vartheta^m_{jk} \quad \downarrow \left[ \sum_{j' \in D} z^m_{j'k} \right] \vartheta^m_{jk} - z^m_{jk} \geq 0 \quad \forall (j, k, m) \in D \times K \times N
\]

\[
0 \leq \zeta^m_{jk} \quad \downarrow 1 - \vartheta^m_{jk} \geq 0 \quad \forall (j, k, m) \in D \times K \times M
\]
formulation; this is another manipulation that is needed in the proof):  

\[ F^\rho(w) \triangleq \left\{ \begin{array}{l}
\left( -\gamma^m_j \left( \hat{R}_{jk}^m + \beta^m_{jO_k} \right) + 
\rho \left( \sum_{k \in K} z^m_j k - \sum_{k' j \in D'_k} Q^m_{k'} + t_{jO_k}^m \sum_{k' j \in D'_k} z^m_j k' + \sum_{k' j \in D'_k} Q^m_{k'} t_{O_k D_k} - N^m \right) \right) \\
\end{array} \right\}
\]

penalty term

\[ \begin{align*}
V_k^m & = (k, m) \in K \times M_+ \\
C_p(h) - t_{ij} & = (i, j) \in N \times N \\\n\sum_{p \in P_{ij}} h_p - \sum_{k \in K} \delta_{ij}^OD_k - \sum_{(k, l) \in K \times K} \delta_{ij}^e_{k, l} \sum_{m \in M} z^m_{j k} & = (j, k, m) \in D \times K \times M
\end{align*} \]

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

\[ \begin{align*}
\sum_{k \in K} z^m_{j k} & \leq \sum_{k' \in K} Q^m_{k'} \quad \forall (j, m) \in D \times M \\
\sum_{j \in D} z^m_{j k} & \geq Q^m_k \quad \forall (k, m) \in K \times M \\
\sum_{m \in M_k} Q^m_k & = Q_k \quad \forall k \in K \\
\theta^m_{j k} & \leq 1 \quad \forall (j, k, m) \in D \times K \times M
\end{align*} \]

(28)

Note that in the set \( W \), the only coupling between the e-HSP service variables \( z^m_{j k} \) and e-HSP demand variables \( Q^m_k \) is through the e-HSP demand constraints; the e-HSP vehicle supply equations and the vehicle-hours bound constraints are transferred to the function \( F^\rho \) via a penalization. The first constraint in (28) is a relaxation of the former omitted e-HSP demand equation and is included in \( W \) to ensure the boundedness of the \( z^m_{j k} \) variables. Our goal is to show that the omitted constraints will be satisfied in the limit when the penalty parameter \( \rho \) tends to \( \infty \), 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 \((F^\rho; W)\). The result is immediate since \( W \) is a compact polyhedron; no proof is needed.

**Lemma 1.** If the path cost functions \( C_p(h) \) are continuous, and \( W \) is nonempty, then the VI \((F^\rho; W)\) 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 mixed complementarity formulation of the VI ($F^p; W$) as follows:

$$0 \leq z_{jk}^m \perp -\gamma_3^m \left( R_{jk}^m + \beta_3^m t_{ij}^m O_k \right) - \lambda_k^m + \eta_j^m + \rho \left[ \sum_{k^* \in K} z_{j,k^*}^m - \sum_{k^* : j = D_{k^*}} Q_{k^*}^m \right] + t_{j}^m O_k - \sum_{j' \in D} \sum_{k^* \in K} z_{j',k^*}^m t_{j',O_k} + \sum_{k^* \in K} Q_{k^*}^m t_{O_k,D_{k^*}} - N^m \right] + + 0 \quad \forall (j, k, m) \in D \times K \times M$$

$$0 \leq \eta_j^m \perp \sum_{k^* \in K} Q_{k^*} - \sum_{k^* \in K} z_{j,k^*}^m \geq 0 \quad \forall (j, k, m) \in D \times K \times M$$

$$0 \leq \lambda_k^m \perp \sum_{j \in D} z_{j,k}^m - Q_{k}^m \geq 0 \quad \forall (k, m) \in K \times M$$

$$0 \leq Q_k^m \perp V_k^0 - u_k \geq 0$$

$$u_k \text{ free,} \quad \sum_{m \in M_+} Q_k^m = Q_k \quad \forall k \in K$$

$$0 \leq h_p \perp C_p(h) - t_{ij} + v_p \geq 0 \quad \forall p \in P_{ij}, \forall (i, j) \in N \times N$$

$$0 \leq s_{ij} \perp \bar{t} - t_{ij} \geq 0 \quad \forall (i, j) \in N \times N$$

$$0 \leq v_p \perp \bar{h} - h_p \geq 0 \quad \forall p \in P$$

$$0 \leq \gamma_{jk}^m \perp \sum_{j \in D} z_{j,k}^m \theta_{jk}^m - z_{j,k}^m + c_{jk}^m \geq 0 \quad \forall (j, k, m) \in D \times K \times N$$

$$0 \leq \eta_{jk}^m \perp 1 - \theta_{jk}^m \geq 0 \quad \forall (j, k, m) \in D \times K \times M$$

$$0 \leq t_{ij} \perp \sum_{p \in P_{ij}} h_p - \left[ \sum_{k \in K} \delta_{ijk}^O Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ijk \ell}^e - HSP \sum_{m \in M} z_{i \ell}^m \right] + s_{ij} \geq 0 \quad \forall (i, j) \in N \times N.$$  

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

$$\bar{h} > \sum_{k \in K} \delta_{ijk}^O Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ijk \ell}^e - HSP \sum_{m \in M} z_{i \ell}^m, \quad \forall \text{ nonnegative } (z, Q) \text{ satisfying (28)}$$

and

$$\bar{t} > \max_{(i, j) \in N \times N} \max_{p \in P_{ij}} \max_{0 \leq h \leq \bar{h}} C_p(h).$$

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

**Proof.** Assume for contradiction that $v_p > 0$ for some $p \in P_{ij}$. We then have $h_p = \bar{h} > 0$. This implies $0 = C_p(h) - t_{ij} + v_p > C_p(h) - t_{ij}.$

27
Thus \( t_{ij} > C_p(h) \geq 0 \). Hence,

\[
 h_{\bar{p}} \leq \sum_{p \in P_{ij}} h_{p} + s_{ij} = \left[ \sum_{k \in K} \delta_{ij \ell}^O D Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ij \ell k}^\epsilon HSP \sum_{m \in M} z_{\ell m}^m \right] < \bar{h},
\]

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 \( \bar{p} \in P_{ij} \),

\[
 h_{\bar{p}} < \sum_{p \in P_{ij}} h_{p} + s_{ij} = \left[ \sum_{k \in K} \delta_{ij \ell}^O D Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ij \ell k}^\epsilon HSP \sum_{m \in M} z_{\ell m}^m \right] < \bar{h},
\]

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

Letting for \( m > 0 \),

\[
 \phi_{j}^{\rho, m} \triangleq -\frac{\rho}{\gamma_m^3} \left[ \sum_{k \in K} z_{jk}^m - \sum_{k' \in \mathcal{K}_k} Q_{k'}^m + \eta_j^m \right], \quad \lambda_k^m \triangleq \frac{\bar{\lambda}_k^m}{\gamma_m^3},
\]

and \( \varphi_{j}^{\rho, m} \triangleq \frac{\rho}{\gamma_m^3} \left[ \sum_{k' \in \mathcal{K}_{j' \in D_k}} z_{jk'}^m t_{j' \in \mathcal{D}_k} - \sum_{k' \in \mathcal{K}} Q_{k'}^m t_{O_k \in \mathcal{D}_k} - N_m^m \right], \tag{29} \]

and recalling \( U_k^m \triangleq V_k^m + \gamma_m^3 \lambda_k^m \), we can write the above complementarity conditions of the VI \((P^\rho; W)\) by dropping the complementarity involving the variables \( \eta_j^m \); this results in the following conditions where all the variables are dependent on the parameter \( \rho \) (through the dependence on \( \phi_{j}^{\rho, m} \) and \( \varphi_{j}^{\rho, m} \)):

\[
0 \leq z_{jk}^m \quad \perp \quad -\bar{R}_{jk}^m - \beta_m^3 t_{O_k} - \phi_{j}^{\rho, m} - \lambda_k^m + \bar{t}_{O_k} \varphi_{j}^{\rho, m} \geq 0 \quad \forall (j, k, m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M}
\]

\[
0 \leq \lambda_k^m \quad \perp \quad \sum_{j \in \mathcal{D}} z_{jk}^m - Q_k^m \geq 0 \quad \forall (k, m) \in \mathcal{K} \times \mathcal{M}
\]

\[
0 \leq Q_k^m \quad \perp \quad U_k^m - u_k \geq 0 \quad \forall (k, m) \in \mathcal{K} \times \mathcal{M}_+
\]

\( u_k \) free,

\[
 \sum_{m \in \mathcal{M}_+} Q_k^m = Q_k \quad \forall k \in \mathcal{K}
\]

\[
0 \leq h_{p} \quad \perp \quad C_p(h) - t_{ij} \geq 0 \quad \forall p \in P_{ij}, \forall (i, j) \in \mathcal{N} \times \mathcal{N}
\]

\[
0 \leq \theta_{jk}^m \quad \perp \quad \sum_{j' \in \mathcal{D}} z_{jk}^m \theta_{j' \in \mathcal{D}_k} - z_{jk}^m + \zeta_{jk}^m \geq 0 \quad \forall (j, k, m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M}
\]

\[
0 \leq \zeta_{jk}^m \quad \perp \quad 1 - \theta_{jk}^m \geq 0 \quad \forall (j, k, m) \in \mathcal{D} \times \mathcal{K} \times \mathcal{M}
\]

\[
0 \leq t_{ij} \quad \perp \quad \sum_{p \in P_{ij}} h_{p} - \left[ \sum_{k \in K} \delta_{ij \ell}^O D Q_k + \sum_{(k, \ell) \in K \times K} \delta_{ij \ell k}^\epsilon HSP \sum_{m \in M} z_{\ell m}^m \right] \geq 0 \quad \forall (i, j) \in \mathcal{N} \times \mathcal{N}.
\]

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 \( \infty \). For each \( \nu \), let \( w_\nu \) be a corresponding solution of the VI \( (F^{\rho_\nu}; W) \). Without loss of generality, we may assume that the sequence \( \{ w_\nu \} \) converges to a limit \( w_\infty \triangleq (z_\infty, Q_\infty, h_\infty, t_\infty, \theta_\infty) \) which must belong to \( W \). 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 \( w_\infty \) satisfies

\[
\sum_{\ell \in K} z_{j_\ell}^{\infty;m} = \sum_{k': j=K_{k'}} Q_{k'}^{\infty;m} \quad \forall (j, m) \in D \times M; \tag{30}
\]

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

\[
N^m - \left[ \sum_{k \in K} \sum_{j \in D} z_{j_\ell}^{\infty;m} t_{jO_k}^\infty + \sum_{k \in K} Q_{k'}^{\infty;m} t_{O_kD_k}^\infty \right] \geq 0; \tag{31}
\]

and (c) the same variables \( \{ z_{j_\ell}^{\infty;m} \} \) 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:

\[
\phi_{j}^{\nu;m}; \lambda_{k}^{\nu;m}; \varphi_{j_\ell}^{\nu;m}; \zeta_{j_\ell}^{\nu;m}; \text{ and } u_{k}^{\nu}; \tag{32}
\]

which are the short-hands of the respective variables that are dependent on \( \rho_\nu \). These variables prevent us from directly passing to the limit \( \nu \to \infty \) in the complementarity conditions corresponding to \( \rho_\nu \). Among the variables in (32), the ones that require the most attention are \( \varphi_{j_\ell}^{\nu;m} \) which appear in the bilinear (thus nonlinear) term: \( t_{jO_k}^{\nu;m} \varphi_{j_\ell}^{\nu;m} \). The other variables in (32) 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 wait 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 (33) involving the positive scalar \( \varepsilon \) is a type of a Slater condition that is needed to deal with the nonlinearity of the products: \( z_{j_\ell}^m t_{jO_k}^m \) and \( Q_k^m t_{O_kD_k}^m \) of the unknowns. Let \( T \) be the set of travel times induced by pairs of vehicle-demand profiles \((z, Q)\) satisfying the first three conditions in (28) that define the polyhedron \( W \).

**Lemma 3.** Suppose that there exists a scalar \( \varepsilon > 0 \) such that for each nonnegative tuple \( Q \) satisfying

\[
\sum_{m \in M_+} Q_k^m = Q_k \text{ for all } k \in K \text{ and for each tuple } t \in T, \text{ there exists } z \text{ such that the tuple } (z, Q, t) \text{ satisfies}
\]

29
the constraints in (7) for all \(m > 0\) with

\[
N^m - \left[ \sum_{k \in K} \sum_{j \in D} z_{jk}^m t_{jO_k} + \sum_{k \in K} Q_k^m t_{O_kD_k} \right] \geq \varepsilon \quad \forall m \in M.
\]

(33)

The following two statements hold:

(a) the limiting tuple \((z^\infty, Q^\infty, t^\infty)\) satisfies (30) and (31);

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

**Proof.** Corresponding to the pair \((Q^\nu, t^\nu)\), let \(\bar{z}^\nu\) be such that the tuple \((\bar{z}^\nu, Q^\nu, t^\nu)\) satisfies the constraints in (7) for all \(m > 0\) with

\[
N^m - \left[ \sum_{k \in K} \sum_{j \in D} \bar{z}_{jk}^\nu t_{jO_k} + \sum_{k \in K} Q_k^\nu t_{O_kD_k} \right] \geq \varepsilon \quad \forall m \in M.
\]

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

\[
F^{\nu}(w^\nu)^T (w - w^\nu) \geq 0, \quad \forall w \in W;
\]

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

\[
0 \leq F^{\nu}(w^\nu)^T (\bar{w}^\nu - w^\nu)
\]

\[
= \sum_{m \in M} \sum_{j \in D} \sum_{k \in K} \left[ -\gamma_3^m \left( \hat{R}_{jk}^m + \beta_3^m t_{jO_k}^\nu \right) + \rho_\nu \left( \sum_{k' \in K} \sum_{j : k' \in D} Q_k^m + \sum_{k \in K} Q_k^m \right) \right] (\bar{z}_{jk}^\nu - z_{jk}^\nu) +
\]

\[
\rho_\nu \sum_{m \in M} \sum_{j \in D} \sum_{k \in K} t_{jO_k}^\nu \left[ \sum_{k' \in K} \sum_{j : k' \in D} z_j^\nu t_{jO_k}^\nu + \sum_{k \in K} Q_k^m t_{O_kD_k}^\nu - N^m \right] (\bar{z}_{jk}^\nu - z_{jk}^\nu) +
\]

\[
= -\gamma_3^m \sum_{m \in M} \sum_{j \in D} \sum_{k \in K} \left( \hat{R}_{jk}^m + \beta_3^m t_{jO_k}^\nu \right) (\bar{z}_{jk}^\nu - z_{jk}^\nu) - \rho_\nu \sum_{m \in M} \sum_{j \in D} \left( \sum_{k' \in K} z_j^\nu + \sum_{k \in K} Q_k^m \right)^2 \text{ bounded}
\]

\[
\rho_\nu \sum_{m \in M} \left[ \sum_{k' \in K} \sum_{j' \in D} z_{j'}^\nu t_{j'O_k}^\nu + \sum_{k \in K} Q_k^m t_{O_kD_k}^\nu - N^m \right] \sum_{j \in D} \sum_{k \in K} t_{jO_k}^\nu (\bar{z}_{jk}^\nu - z_{jk}^\nu)
\]

using the condition that

\[
\sum_{k' \in K} z_j^\nu = \sum_{k' \in K} Q_k^m \leq \sum_{k' \in K} Q_k.
\]

Moreover, since

\[
N^m - \left[ \sum_{k \in K} \sum_{j \in D} z_{jk}^\nu t_{jO_k} + \sum_{k \in K} Q_k^m t_{O_kD_k} \right] \geq \varepsilon
\]
by the choice of $z^\nu$, we deduce
\[
\sum_{j \in \mathcal{D}} \sum_{k \in \mathcal{K}} t^\nu_{j,k} \left( z^\nu_{j,k} - z^\nu_{j,k}^\prime \right) \leq - \left[ \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z^\nu_{j',k'} t^\nu_{j',O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} - N^m \right] - \varepsilon.
\]

Since $t_{+}t \geq (t_{+})^2$ for any scalar $t$, we deduce
\[
0 \leq \text{a bounded term} - \rho^\nu \sum_{m \in \mathcal{M}} \left( \sum_{t \in \mathcal{K}} z^\nu_{j,k}^\prime t^\nu_{j,O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} - N^m \right)^2
\]
\[
- \rho^\nu \sum_{m \in \mathcal{M}} \left[ \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z^\nu_{j',k'} t^\nu_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} - N^m \right] + \varepsilon \sum_{m \in \mathcal{M}} \rho^\nu \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z^\nu_{j',k'} t^\nu_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} - N^m \right] \leq \gamma^m \phi^\nu_{j,m}
\]

This shows that
\[\lim_{\nu \to \infty} \left( \sum_{t \in \mathcal{K}} z^\nu_{j,k}^\prime - \sum_{k' : j = D_k} Q^\nu_{k',j} \right) = 0 \text{ for all } m \in \mathcal{M} \text{ and } j \in \mathcal{D};\]
\[\lim_{\nu \to \infty} \left[ \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z^\nu_{j',k'} t^\nu_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} - N^m \right] = 0; \text{ and}
\]

set as $\hat{e}^\nu_{j,m} = \frac{\phi^\nu_{j,m}}{\rho^\nu}$

- the sequence $\{\phi^\nu_{j,m}\}$ is bounded for all $m \in \mathcal{M}$.

These conclusions establish the two claims of the lemma. □

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

We may augment the complementarity conditions of the VI $(F^\nu; \mathcal{W})$ by the conditions:
\[
\phi^\rho_{j,m} \text{ free, } \sum_{t \in \mathcal{K}} z^\nu_{j,k} - \sum_{k' : j = D_k} Q^\nu_{k',j} = 0, \quad \forall (j, m) \in \mathcal{D} \times \mathcal{M}
\]
\[
0 \leq \phi^\nu_{j,m} - \sum_{k' \in \mathcal{K}} \sum_{j' \in \mathcal{D}} z^\nu_{j',k'} t^\nu_{j'O_{k'}} + \sum_{k' \in \mathcal{K}} Q^\nu_{k',j'} t^\nu_{O_{k'},D_k} \geq 0,
\]

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

\[ 0 \leq z_{jk}^{\nu:m} \perp -\hat{R}_{jk}^{\nu:m} - \beta^{m}t_{jO_k}^{\nu} - \phi_{j}^{\nu:m} - \lambda_{k}^{\nu:m} + t_{jO_k}^{\nu} \varphi^{\nu:m} \geq 0 \quad \forall (j,k,m) \in D \times K \times M \]
\[ 0 \leq Q_{k}^{\nu:0} \perp V_{k}^{\nu:0} - u_{k}^{\nu} \geq 0 \]
\[ 0 \leq Q_{k}^{\nu:m} \perp V_{k}^{\nu:m} - u_{k}^{\nu} + \gamma^{m} \lambda_{k}^{\nu:m} \geq 0 \quad \forall (k,m) \in K \times M \]
\[ 0 \leq \varphi^{\nu:m} \perp \bar{e}^{\nu:m} + N^{m} - \left[ \sum_{k' \in K} \sum_{j' \in D} \zeta_{j'k'}^{\nu:m} t_{j'k'}^{\nu} + \sum_{k' \in K} Q_{k'}^{\nu:m} t_{k'k}^{\nu} \right] \geq 0 \quad \forall m \in M \]
\[ 0 \leq t_{ij}^{\nu} \perp \sum_{p \in P_{ij}} h_{p}^{\nu} - \left[ \sum_{k \in K} \delta^{OD}_{ij} Q_{k} + \sum_{(k',k) \in K \times K} \delta^{HSP}_{ij} \sum_{m \in M} z_{ij}^{\nu:m} \right] \geq 0 \quad \forall (i,j) \in N \times N \]
\[ 0 \leq h_{p}^{\nu} \perp C_{p}(h^{\nu}) - t_{ij}^{\nu} \geq 0 \quad \forall p \in P_{ij}, \forall (i,j) \in N \times N \]
\[ 0 \leq \theta_{jk}^{\nu:m} \perp \sum_{j \in D} z_{jk}^{\nu:m} \left[ \theta_{jk}^{\nu:m} - z_{jk}^{\nu:m} + \zeta_{jk}^{\nu:m} \right] \geq 0 \quad \forall (j,k,m) \in D \times K \times N \]

---

\[ \phi_{j}^{\nu:m} \text{ free,} \]

\[ \sum_{k \in K} z_{jk}^{\nu:m} - \sum_{k': j = D_{k'}} Q_{k'}^{\nu:m} = e_{j}^{\nu:m} \quad \forall (j,m) \in D \times M \]

\[ u_{k}^{\nu} \text{ free,} \]

\[ \sum_{m \in M_{+}} Q_{k}^{\nu:m} = Q_{k} \quad \forall k \in K \]

---

\[ 0 \leq \lambda_{k}^{\nu:m} \perp \sum_{j \in D} z_{jk}^{\nu:m} - Q_{k}^{\nu:m} \geq 0 \quad \forall (k,m) \in K \times M \]

\[ 0 \leq \zeta_{jk}^{\nu:m} \perp 1 - \theta_{jk}^{\nu:m} \geq 0 \quad \forall (j,k,m) \in D \times K \times M, \]

where the primary variables \((z_{jk}^{\nu:m}, Q_{k}^{\nu:m}, t_{jO_k}^{\nu}, \phi_{j}^{\nu:m})\); 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 \((\phi_{j}^{\nu:m}, u_{k}^{\nu})\) 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 [3].

**Lemma 4.** Consider the following mixed complementarity conditions:

\[ 0 \leq x \perp G(x) + A^{T}\mu + B^{T}\lambda \geq 0 \]
\[ \mu \text{ free} \quad Cx - e = 0 \]
\[ 0 \leq \lambda \perp Dx - f \geq 0, \]

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 \((e^\infty, f^\infty)\). \(\Box\)

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

**Theorem 1.** Under the assumptions of Lemmas \([1, 2, 3]\) a TGEM-ets equilibrium exists, provided that \(\gamma_3^m > 0\) for all \(m \in M\). \(\Box\)

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 \(\rho\) tends to \(\infty\), the theorem establishes that an equilibrium solution exists. It is important to note, however, that such constraints are not shown to hold for finite \(\rho\). 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 [6, 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 \([1, 2, 15]\) set up to optimize the secondary objective among all equilibrium solutions. For instance, one such objective could be the deadhead miles given by \([1]\). Details of this issue are best left for future investigation.

### 8 Numerical Experiments

Formulated as a mixed complementarity problem (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 \([7, 11]\), 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 complex engineering equilibrium problems. 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 [6] 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 1. 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 1: Illustration of the small network](image)

<table>
<thead>
<tr>
<th>Links</th>
<th>From node</th>
<th>To node</th>
<th>Free flow travel time (h)</th>
<th>Length (mile)</th>
<th>Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.3</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0.5</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0.4</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0.4</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>4</td>
<td>0.3</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1.0</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0.4</td>
<td>15</td>
<td>50</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>1</td>
<td>0.4</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>2</td>
<td>0.5</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

Parameters related to the e-HSP vehicle fees and the customer disutility are provided in Table 2. They
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 are based on those from the City of Seattle [18, 17],
with some modifications. The customer’s value of time (i.e., $\gamma^m_i$) by solo driving should be much larger
than that using e-HSP vehicles (since e-HSP customers can do something productive when riding an
e-HSP car). Between riding taxi and TNC cars, customers using taxi services should have smaller value
of time since they can be both productive and feel safer (compared with riding a TNC vehicle). We conduct sensitivity analysis later in this section regarding how different choices of some of the parameters \((\alpha^m_1, \alpha^m_2, \beta^m_2, \gamma^m_1)\) may change the customers’ choice of modes and the total system VMT and VHT. Other parameters are not tested since either the model results are not very sensitive to these parameters or they are the same for taxi and TNC services. In practice, \(\gamma^m_3\), \(m = 1, 2\) can be small because customers when waiting for matching with e-HSP vehicles can potentially do something that is productive rather than wasting their time. In this paper, we set \(\gamma^1_3 = \gamma^2_3 = 0.0001\) in all the numerical experiments.

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

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Notation</th>
<th>Small network</th>
<th>Sioux-Falls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed fare ($)</td>
<td>(F^m(m = 1, 2))</td>
<td>3, 2</td>
<td>3, 2</td>
</tr>
<tr>
<td>Time-based fare rate ($/hr)</td>
<td>(\alpha^m_1(m = 1, 2))</td>
<td>20, 15</td>
<td>15, 10</td>
</tr>
<tr>
<td>Distance-based fare rate ($/mile)</td>
<td>(\alpha^m_2(m = 1, 2))</td>
<td>2, 1.5</td>
<td>2.25, 2</td>
</tr>
<tr>
<td>Conversion factor (from travel time to cost, $/hr)</td>
<td>(\beta^m_1(m = 1, 2))</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>Conversion factor (from travel distance to cost, $/mile)</td>
<td>(\beta^m_2(m = 0, 1, 2))</td>
<td>0.95, 0.55, 0.9</td>
<td>2, 0.5, 2</td>
</tr>
<tr>
<td>Conversion factor (from OSP vehicle waiting time to cost)</td>
<td>(\beta^m_3(m = 1, 2))</td>
<td>0.2, 0.1</td>
<td>2, 1</td>
</tr>
<tr>
<td>Value of time of customer (while traveling, $/hr)</td>
<td>(\gamma^m_1(m = 0, 1, 2))</td>
<td>40, 7, 18</td>
<td>50, 3, 15</td>
</tr>
<tr>
<td>Value of time of customer (while waiting, $/hr)</td>
<td>(\gamma^m_2(m = 0, 1, 2))</td>
<td>0, 3, 2</td>
<td>0, 1, 0.5</td>
</tr>
<tr>
<td>The number of e-HSP vehicles</td>
<td>(N^m(m = 1, 2))</td>
<td>400, 400</td>
<td>40000, 40000</td>
</tr>
</tbody>
</table>

The results of solving TGEM-ets verify that the equilibria are reached for both customers’ mode choices (in terms of choosing solo driving, TNC, or taxi) 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. The deadhead miles are 2683.53 vehicle-miles, and VMT and vehicle hours traveled (VHT) are 5463.47 vehicle-miles and 3.71 vehicle-hours, respectively. 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).

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 taxi-based and TNC-based VHT, respectively, in these tables. Four set of parameters are selected for this purpose: time-based fare rate \((\alpha^m_1, m = 1, 2)\), distance-based fare rate \((\alpha^m_2, m = 1, 2)\), distance-based conversation factor \((\beta^m_2, m = 0, 1, 2)\), and value of time of customers \((\gamma^m_1, m = 1, 2)\). 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 TNC, i.e., \(\beta^0_2 = \beta^2_2\). When changing customers’ value of time parameters, we set the parameter for solo driving fixed as the base case \((\gamma^0_1 = 40)\), and change those for TNC and taxi 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.
Table 3: Results of the small network (base case)

<table>
<thead>
<tr>
<th>OD Pair</th>
<th>Mode Choice</th>
<th>Customer Disutility</th>
<th>Num of Used Paths</th>
<th>Min Path TT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 -&gt; 2</td>
<td>Solo: 4.499</td>
<td>Solo: 44.962</td>
<td>1</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td>Taxi: 0</td>
<td>Taxi: 44.974</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TNC: 45.501</td>
<td>TNC: 44.962</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 -&gt; 3</td>
<td>Solo: 40</td>
<td>Solo: 58.657</td>
<td>1</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>Taxi: 0</td>
<td>Taxi: 61.779</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TNC: 0</td>
<td>TNC: 59.293</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 -&gt; 4</td>
<td>Solo: 0</td>
<td>Solo: 70.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Taxi: 44.165</td>
<td>Taxi: 67.317</td>
<td>2</td>
<td>1.297</td>
</tr>
<tr>
<td></td>
<td>TNC: 5.835</td>
<td>TNC: 67.317</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

will decrease (or at least remain the same). The results show such a trend for most of the parameters as shown in the tables.

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

<table>
<thead>
<tr>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>Solo%</th>
<th>Taxi%</th>
<th>TNC%</th>
<th>VMT</th>
<th>VHT (Taxi)</th>
<th>VHT (TNC)</th>
<th>VHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>8.6</td>
<td>28.57</td>
<td>0.00</td>
<td>71.43</td>
<td>5529.94</td>
<td>0.00</td>
<td>3.10</td>
<td>3.76</td>
</tr>
<tr>
<td>20</td>
<td>12.6</td>
<td>28.57</td>
<td>35.72</td>
<td>35.71</td>
<td>5529.94</td>
<td>1.12</td>
<td>1.97</td>
<td>3.76</td>
</tr>
<tr>
<td>20</td>
<td>14.5</td>
<td>28.57</td>
<td>44.62</td>
<td>26.81</td>
<td>5529.94</td>
<td>2.16</td>
<td>0.94</td>
<td>3.76</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>0.00</td>
<td>35.71</td>
<td>64.29</td>
<td>6329.94</td>
<td>1.12</td>
<td>2.91</td>
<td>4.03</td>
</tr>
<tr>
<td>20.5</td>
<td>10</td>
<td>28.57</td>
<td>0.00</td>
<td>71.43</td>
<td>5529.94</td>
<td>0.00</td>
<td>3.10</td>
<td>3.76</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>31.74</td>
<td>31.51</td>
<td>36.75</td>
<td>5463.47</td>
<td>1.74</td>
<td>1.24</td>
<td>3.71</td>
</tr>
<tr>
<td>32</td>
<td>15</td>
<td>64.29</td>
<td>0.00</td>
<td>35.71</td>
<td>4779.94</td>
<td>0.00</td>
<td>1.97</td>
<td>3.37</td>
</tr>
<tr>
<td>40</td>
<td>30</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2779.94</td>
<td>0.00</td>
<td>0.00</td>
<td>2.48</td>
</tr>
</tbody>
</table>

The four tables show that once the e-HSP usage (by either TNC or taxi) increases, the total system VMT/VHT also increases and in many cases fairly dramatically. For example, Table 3 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 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
Table 5: Results of changing distance-based fare rate ($\alpha_m^m, m = 1, 2$)

<table>
<thead>
<tr>
<th>$\alpha_1^2$</th>
<th>$\alpha_2^2$</th>
<th>Solo%</th>
<th>Taxi%</th>
<th>TNC%</th>
<th>VMT</th>
<th>VHT (Taxi)</th>
<th>VHT (TNC)</th>
<th>VHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>1.2</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>6329.94</td>
<td>4.03</td>
<td>0.00</td>
<td>4.03</td>
</tr>
<tr>
<td>1.6</td>
<td>1.2</td>
<td>0.00</td>
<td>71.43</td>
<td>28.57</td>
<td>6329.94</td>
<td>3.09</td>
<td>0.93</td>
<td>4.03</td>
</tr>
<tr>
<td>1.71</td>
<td>1.2</td>
<td>0.00</td>
<td>58.00</td>
<td>42.00</td>
<td>6329.94</td>
<td>2.35</td>
<td>1.68</td>
<td>4.03</td>
</tr>
<tr>
<td>1.84</td>
<td>1.2</td>
<td>0.00</td>
<td>35.71</td>
<td>64.29</td>
<td>6329.94</td>
<td>1.12</td>
<td>2.91</td>
<td>4.03</td>
</tr>
<tr>
<td>1.9</td>
<td>1.4</td>
<td>0.00</td>
<td>71.43</td>
<td>28.57</td>
<td>6329.94</td>
<td>3.10</td>
<td>0.93</td>
<td>4.03</td>
</tr>
<tr>
<td>1.96</td>
<td>1.4</td>
<td>0.00</td>
<td>56.48</td>
<td>43.52</td>
<td>5579.94</td>
<td>2.98</td>
<td>1.05</td>
<td>4.03</td>
</tr>
<tr>
<td>2.01</td>
<td>1.4</td>
<td>20.79</td>
<td>35.71</td>
<td>43.50</td>
<td>5747.90</td>
<td>1.12</td>
<td>2.23</td>
<td>3.83</td>
</tr>
<tr>
<td>1.9</td>
<td>1.5</td>
<td>28.57</td>
<td>71.43</td>
<td>0.00</td>
<td>5529.94</td>
<td>3.10</td>
<td>0.00</td>
<td>3.76</td>
</tr>
<tr>
<td>2.03</td>
<td>1.5</td>
<td>64.29</td>
<td>35.71</td>
<td>0.00</td>
<td>4779.94</td>
<td>1.97</td>
<td>0.00</td>
<td>3.37</td>
</tr>
<tr>
<td>2.23</td>
<td>1.5</td>
<td>64.29</td>
<td>0.00</td>
<td>35.71</td>
<td>4779.99</td>
<td>0.00</td>
<td>1.97</td>
<td>3.37</td>
</tr>
<tr>
<td>2</td>
<td>1.8</td>
<td>28.57</td>
<td>71.43</td>
<td>0.00</td>
<td>5529.98</td>
<td>3.10</td>
<td>0.00</td>
<td>3.76</td>
</tr>
<tr>
<td>2.2</td>
<td>1.8</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2779.94</td>
<td>0.00</td>
<td>0.00</td>
<td>2.48</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$\beta_0^2$</th>
<th>$\beta_1^2$</th>
<th>Solo%</th>
<th>Taxi%</th>
<th>TNC%</th>
<th>VMT</th>
<th>VHT (Taxi)</th>
<th>VHT (TNC)</th>
<th>VHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.92</td>
<td>0.7</td>
<td>28.57</td>
<td>71.43</td>
<td>0.00</td>
<td>5529.94</td>
<td>3.10</td>
<td>0.00</td>
<td>3.76</td>
</tr>
<tr>
<td>0.92</td>
<td>0.82</td>
<td>28.57</td>
<td>35.85</td>
<td>35.58</td>
<td>5529.94</td>
<td>1.13</td>
<td>1.97</td>
<td>3.37</td>
</tr>
<tr>
<td>0.92</td>
<td>0.88</td>
<td>64.29</td>
<td>35.71</td>
<td>0.00</td>
<td>4779.94</td>
<td>0.00</td>
<td>1.97</td>
<td>3.37</td>
</tr>
<tr>
<td>0.66</td>
<td>0.25</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2779.93</td>
<td>0.00</td>
<td>0.00</td>
<td>2.48</td>
</tr>
<tr>
<td>0.76</td>
<td>0.25</td>
<td>85.14</td>
<td>14.86</td>
<td>0.00</td>
<td>3611.94</td>
<td>0.80</td>
<td>0.00</td>
<td>2.83</td>
</tr>
<tr>
<td>0.94</td>
<td>0.25</td>
<td>42.35</td>
<td>34.25</td>
<td>23.39</td>
<td>5240.51</td>
<td>1.89</td>
<td>0.74</td>
<td>3.58</td>
</tr>
<tr>
<td>0.76</td>
<td>0.15</td>
<td>82.55</td>
<td>17.45</td>
<td>0.00</td>
<td>3757.21</td>
<td>0.94</td>
<td>0.00</td>
<td>2.89</td>
</tr>
<tr>
<td>0.77</td>
<td>0.15</td>
<td>66.16</td>
<td>33.84</td>
<td>0.00</td>
<td>4674.84</td>
<td>1.86</td>
<td>0.00</td>
<td>3.32</td>
</tr>
</tbody>
</table>

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 2. 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 2 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 (the case with all customers choosing solo driving) with respect to the percentage of customers who choose e-HSP modes (taxi or TNC). 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.
Table 7: Results of changing customer value of time ($\gamma_m^1, m = 1, 2$)

<table>
<thead>
<tr>
<th>$\gamma_1^1$</th>
<th>$\gamma_2^1$</th>
<th>Solo%</th>
<th>Taxi%</th>
<th>TNC%</th>
<th>VMT</th>
<th>VHT (Taxi)</th>
<th>VHT (TNC)</th>
<th>VHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>18</td>
<td>0.00</td>
<td>100.00</td>
<td>0.00</td>
<td>6329.99</td>
<td>4.03</td>
<td>0.00</td>
<td>4.03</td>
</tr>
<tr>
<td>4</td>
<td>18</td>
<td>28.57</td>
<td>71.43</td>
<td>0.00</td>
<td>5529.94</td>
<td>3.10</td>
<td>0.00</td>
<td>3.76</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>64.29</td>
<td>0.00</td>
<td>35.71</td>
<td>4779.94</td>
<td>0.00</td>
<td>1.97</td>
<td>3.37</td>
</tr>
<tr>
<td>7</td>
<td>17.3</td>
<td>28.57</td>
<td>66.11</td>
<td>5.32</td>
<td>5529.94</td>
<td>2.88</td>
<td>0.21</td>
<td>3.76</td>
</tr>
<tr>
<td>7</td>
<td>16.3</td>
<td>0.00</td>
<td>35.71</td>
<td>64.29</td>
<td>6329.94</td>
<td>1.12</td>
<td>2.91</td>
<td>4.03</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>6329.94</td>
<td>0.00</td>
<td>4.03</td>
<td>4.03</td>
</tr>
<tr>
<td>8</td>
<td>17.3</td>
<td>28.57</td>
<td>35.71</td>
<td>35.71</td>
<td>5529.94</td>
<td>1.12</td>
<td>1.97</td>
<td>3.76</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>0.00</td>
<td>0.00</td>
<td>100.00</td>
<td>6329.98</td>
<td>0.00</td>
<td>4.03</td>
<td>4.03</td>
</tr>
</tbody>
</table>

8.2 Numerical results of waiting times

We use the small network in Figure 1 to test the relation of the waiting times of e-HSP vehicles and their customers, using the fitting model (27). For this, 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 (27). Figure 3 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
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 3: Distribution of the objective values (small network)

Solving the fitting model (27) for each TGEM-ets problem will result in a coefficient pair ($\eta_1$ and $\eta_2$). Figure 4 and Figure 5 show, respectively, the $\eta_1$ and $\eta_2$ parameters in the waiting time function (25) for e-hailing taxis and TNCs. Interestingly, these values are clustered around a few critical values (e.g., for e-hailing taxi, they are 0, 0.2, 0.4 for $\eta_1$ and 0, 0.1, 0.3 for $\eta_2$). While Figure 3 - Figure 5 indicate a rather good fitting of the expression (25), there are some noticeable differences from the relationship established in the past for traditional street-hailing taxis. For example, $\eta_1$ and $\eta_2$ 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 e-hailing services and the street-hailing taxi services. Further investigation of such a relation is important and deserves to be pursued in future research.

8.3 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 [19]. 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 [2]) 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 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 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 PATH solver can solve the Sioux-Falls network problem successfully, sometimes using different starting points from 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 6 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 2, 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.

It can also be observed 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. We can also see from the figure that when demand symmetry is small and the e-HSP usage is large, it is less likely that the VMT increase is zero, i.e., there are always significant deadhead miles. Also, at very high e-HSP usage levels the deadhead miles can go down depending on the network topology (e.g., high demand symmetry) since it is likely that consecutive service requests are close to one another.

The above 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 has been actually observed recently in practice. For example, in New York City, it was reported that e-hailing services have created more congestion during peak periods than non-peak periods [16]. One 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-HSP usage and the potential VMT decrease due to the reduction of searching for parking traffic, 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 ridesharing (such as Uber pool), 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). The proposed
model can potentially serve as a tool by both transportation planners and e-HSPs for their planning.

![Change of VMT with e-HSP Usage](image)

Figure 6: VMT changes with percentage of customers choosing e-HSP modes (Sioux-Falls Network)

## 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 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 a customer.
(i.e., street-hailing). The model developed in this paper is one of the first attempts in understanding the congestion impacted by e-hailing transportation service providers which more closely resemble TNCs such as Uber and Lyft type of providers. 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 (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 modeled. The second assumption can be relaxed by explicitly modeling the matching between e-HSP vehicles and customers, for which the methods in [25] for traditional taxis and more recently in [33] 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 include a model to reflect how the driver selects a provider. The fifth assumption can be relaxed by specially modeling e-HSP platform and drivers’ behaviors related to surcharging, trip cancellation, etc.

In the proposed model, we also assume TNC vehicles operate similarly as taxis, i.e., they only serve a single group of customers for a given OD pair, i.e., no ridesharing is modeled. Although this is largely the case in the current situation (the use of ridesharing-type of services is pretty low in current TNCs), when TNCs become the dominant transportation mode in a network (e.g., if automated vehicles are mature and become dominant in the market) in the future, ridesharing will be probably the norm of urban transportation. To capture this, the proposed model needs to be extended to integrate ridesharing network models, e.g., the one in [23]. This paper also focuses on identifying a user equilibrium solution. However, transportation planners may be more interested in identifying a system wide optimal solution and the required incentives for the adoption of the system optimal solution. Furthermore, although we show there exists an equilibrium solution to the model, there may be multiple such solutions. Therefore, it would be interesting to study if there might be some conditions for the model to possess a unique equilibrium solution. In addition, we assume 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 between taxi or TNC and transit also need to be explicitly modeled. Last but not least, the proposed models need to be tested and further developed (e.g., to relax some of the assumptions as discussed above) using real-world 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.

Acknowledgements

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


