Congestion and environmental toll schemes for the morning commute with heterogeneous users and parallel routes

by

Jiancheng Long
Professor
School of Automotive and Transportation Engineering
Hefei University of Technology
Hefei 230009, China
Tel: +86-551-62901960
E-mail: jianchenglong@hfut.edu.cn

W.Y. Szeto
Associate Professor
Department of Civil Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong
Tel: +852-28578552
Email: ceszeto@hku.hk

Abstract

We design a congestion and environmental toll (CET) scheme for the morning commute with heterogeneous users in a single OD network with parallel routes. The designed toll scheme relies upon the concept of marginal-cost pricing and is anonymous. The Henderson approach is used to model road congestion and the tolling problem to examine commuter's arrival time and route choice at the CET equilibrium (CETE). Linear interpolation is applied to approximate the emission cost function and the resulting CETE problem is formulated as an unconstrained optimization problem, which is solved by the modified Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. Unlike the existing approach, this novel approach does not require that the arrival of each group of commuters at the destination at the equilibrium follows a predetermined order, and can handle non-monotone (emission) cost function. As two special cases, no-toll equilibrium (NTE) and the congestion toll equilibrium (CTE) are also examined, and the two resultant equilibrium problems are formulated and solved by the same approach. This approach is shown to be more efficient than the existing approach. Bi-level programming models are proposed to formulate the optimal congestion toll and CET design problems, in which the CTE and CETE problems are the corresponding lower level problem. These models are solved by the double BFGS method, which uses a classical BFGS method to solve the upper level model and the proposed BFGS method to solve the lower level model. Finally, numerical examples are provided to illustrate the properties of the models and the efficiency of the proposed solution algorithms.

Keywords: Morning commute; Road pricing; Air pollution; General heterogeneity; Unconstrained optimization problem; BFGS method.

1. Introduction

With the development of the economy and the increasing demand for urban transportation, traffic congestion during the morning peak hour is widespread in major cities around the world. The traffic environmental problems associated with traffic congestion have also received more and more attention. In the past few decades, studies of morning commute can be broadly classified into two types: (1) modeling peak hour congestion (e.g., Vickrey, 1969; Arnott et al., 1994; Lindsey, 2004; Liu et al., 2015; Xiao et al., 2015; Guo et al., 2018; Liu et al., 2018) and (2) peak hour congestion management (e.g., Arnott et al., 1990; Laih, 1994; Zhang et al., 2008, 2011; Lindsey et al., 2012; Chen et al., 2015; Xiao et al., 2016; Ma et al., 2017). Among them, the former mainly focuses on analyzing how travelers choose their travel modes, departure times, routes, etc. to finish their commutes, obtaining the equilibrium traffic flows, and evaluating the performance of transportation systems. The latter is aimed at managing the level of traffic demand and its spatial and temporal distributions and inducing travelers to change travel options so that traffic is spread over time and space, thereby alleviating traffic congestion.

1.1. User heterogeneity in peak hour congestion problems

In modeling morning peak hour congestion, commuters can be assumed either (1) homogeneous (e.g., Vickrey, 1969; Henderson, 1974, 1977; Chu, 1995; Xiao et al., 2015; Guo et al., 2018; Coria and Zhang, 2017) or (2) heterogeneous (e.g., Arnott et al., 1994; Lindsey, 2004; Chen et al., 2015; Liu et al., 2015; Wu and Huang, 2015) with respect to the value of time (VOT) (including the unit cost of travel time and the unit cost of schedule delay early or late). The congestion models with homogenous commuters usually have analytical solutions, and hence are widely used to evaluate the performance of traffic management measures. Compared with the assumption that commuters are homogeneous, the assumption that commuters are heterogeneous is more in line with the actual situation. This is because commuters are heterogeneous in nature. However, if user heterogeneity is considered, analytical solutions to the congestion models cannot be obtained except for some special cases, such as when restrictions are imposed on commuters' VOT (including the unit cost of travel time and the unit cost of schedule delay) in the bottleneck model.

To solve the congestion model with general user heterogeneity, we can discretize time and the VOT of commuters, and treat the problem as a dynamic traffic assignment (DTA) problem. However, since the travel cost function highly depends on the network loading model used and may not be monotonic (Ghali and Smith, 1993; Long et al., 2015, 2016), many existing algorithms that rely on the monotone property of the cost function for guaranteeing convergence do not often give a convergent solution for the resultant DTA problem (Long et al., 2015, 2016; Guo et al., 2018). Instead, simulation-based DTA models have been gaining attention in recent years and are often used in the situation where analytical solutions are not attainable (e.g., Florian et al., 2008; Shafiei et al., 2018). The simulation-based models focus on enabling practical deployment for realistic large-scale networks. However, the solution properties of these models, such as existence and uniqueness, are not guaranteed and cannot be determined in advance.

Recently, a semi-analytical approach has been developed to solve bottleneck models with general user heterogeneity (e.g., Chen et al., 2015; Liu et al., 2015). Chen et al. (2015) formulated the step-tolled bottleneck model for a corridor with general user heterogeneity as a capacity-constrained static traffic assignment problem. Liu et al. (2015) formulated the bottleneck model for a single-OD network with multiple parallel routes and general user heterogeneity as a static traffic assignment problem. Due to the asymmetric link cost functions, the two static traffic assignment problems were formulated as variational inequality (VI) problems, and solved by the Gauss-Seidel decomposition (GSD) method (Pang, 1985).

1.2. Road pricing for peak hour congestion management

The main function of peak hour congestion management is to redistribute traffic in time and space so as to improve the efficiency of the transportation system and achieve specific goals, such as mitigating traffic congestion, improving safety, saving energy, reducing roadside emissions, etc. Road pricing is a commonly used measure of peak hour congestion management. As a means of economic regulation, it has been long recognized as an effective measure to manage traffic congestion. So far, road pricing has been successfully implemented in several cities, such as London, Singapore, Milan, and Stockholm. In the literature, various road pricing schemes have been developed and can be roughly classified into two categories: the first-best (e.g., Vickrey, 1969; Yildirim and Hearn, 2005; Sumalee and Xu, 2011; Carey and Watling, 2012; Coria and Zhang, 2017; Ma et al., 2017) and the second-best (e.g., Laih, 1994; Lindsey et al., 2012; Chen et al., 2015; Ren et al., 2016; Li et al., 2018).

The first-best pricing schemes include the marginal-cost pricing scheme. The marginal-cost toll equals the difference between the marginal social cost and the marginal private cost (Yang and Huang, 2005). The first-best or marginal-cost pricing schemes can achieve system optimum in terms of minimum total system cost or maximum social surplus, but is cumbersome for practical implementation because travelers cannot know their toll charges in advance and the toll collection cost is high (Lindsey et al., 2012). Owing to the imperfection of the first-best pricing schemes, the second-best counterparts consider not only the improvement of the road system but also the issues of practical implementation. Compared with the first-best pricing schemes, the second-best counterparts are easier to implement and have a lower operating cost, but they are usually harder to design.

Two types of objectives are mainly considered in road pricing problems: (1) congestion (e.g., Arnott et al., 1990; Laih, 1994; Sumalee and Xu, 2011; Carey and Watling, 2012; Lindsey et al., 2012; Yang and Huang, 2005; Ren et al., 2016) and (2) environmental (e.g., Yin and Lawphongpanich, 2006; Sharma and Mishra, 2013; Aziz et al., 2017; Ma et al., 2017; Li et al., 2018). If only the congestion objective is considered, the corresponding toll scheme is referred to as the congestion toll (CT) scheme. If both the congestion and environmental objectives are considered, the corresponding toll scheme is referred to as the congestion and environmental toll (CET) scheme. The congestion objective function has better mathematical properties than the environmental objective function, such as monotonicity. This results in the road pricing problems with the congestion objective easier to be solved than those with the environmental objective. In the literature, most of the road pricing problems only consider the congestion objective. To incorporate the environmental objective into these problems, there are mainly two approaches: (1) reformulate the objective as a side constraint (e.g., Sharma and Mishra, 2013; Aziz et al., 2017) and (2) combine the environmental objective with the congestion objective using the weighted-sum method (e.g., Yin and Lawphongpanich, 2006; Li et al., 2018). Generally, adding environmental constraints may deteriorate the nice properties of the feasible solution set of the original problem, leading to the resultant optimization problem difficult to solve. On the contrast, the second approach maintains the same feasible solution set as the road pricing problems with only the congestion objective.

1.3. Approaches to modeling peak hour congestion and road pricing problems

There are two common approaches to modeling peak hour congestion and road pricing problems: (1) the Vickrey approach (e.g., Vickrey, 1969; Arnott et al., 1994; Lindsey, 2004; Chen et al., 2015; Liu et al., 2015; Xiao et al., 2015; Guo et al., 2018) and (2) the Henderson approach (e.g., Henderson, 1974, 1977; Chu, 1995; Coria and Zhang, 2017). The Vickrey approach assumes that there is a bottleneck with a fixed capacity at the end of a road and models congestion as queuing behind the bottleneck (Vickrey, 1969). The Henderson approach models congestion by a speed-flow function at each instant, where the speed for any commuter is constant throughout the journey and his/her travel time is solely determined by the flow at the moment of departure (Henderson, 1974, 1977). The Henderson approach was reformulated by Chu (1995). In the reformulated model, it is assumed that commuters' travel speed depends on the density of the traffic at the moment of arrival. Chu (1995) solved the model with closed-form solutions for the relationship between speed and traffic flow, which facilitate the derivation of analytical results and the analysis of economic implications of road pricing (Coria and Zhang, 2017). In the literature, the Henderson approach has been used to model peak hour congestion and road congestion pricing problems with homogenous users. Whether this approach can be applied to model those with heterogeneous users and environmental (or in particular vehicle emissions) considerations is questionable. Moreover, there are the two main challenges of modeling these two features in the problems and solving the resultant problems: (1) Empirical results show that the emission rate function is usually non-monotone with respect to travel speed (Penic and Upchurch, 1992). The non-monotonicity leads to the challenge of formulating the marginal social cost. (2) The arrival of each group of commuters at the destination at the CET equilibrium (CETE) may not follow a predetermined order. Furthermore, it is unclear whether there are efficient solution methods to solve the resulting problems.

1.4. Objectives and contributions

In this study, we adopt the Henderson approach to model peak hour congestion and road pricing problems with environmental considerations and heterogeneous users in a single origin-destination (OD) network with multiple parallel roads. We examine the CETE, in which the CET scheme relies upon the concept of marginal-cost pricing. To determine marginal social costs, we apply linear interpolation to approximate the emission cost function, which can be non-monotone with respect to travel speed. The CETE problem is formulated as an unconstrained optimization problem. The equivalence between the proposed formulation and the equilibrium condition is proved. The CETE problem is solved by the Broyden-Fletcher-Goldfarb-Shanno (BFGS)-based method.

We also investigate the no-toll equilibrium (NTE) and the CT equilibrium (CTE) as special cases. Unlike Coria and Zhang (2017), we consider heterogeneous instead of homogenous users. Unlike Liu et al. (2015) and Chen et al. (2015), we used different modeling (Henderson vs Vickrey), formulation (unconstrained optimization vs VI), and solution (BFGS-based vs GSD) approaches. We compare the efficiency of the two solution approaches. We demonstrate the applicability of the proposed methodology using the empirical data on a segment of California State Route 91.

To determine an optimal CET scheme, we develop a bilevel program, which is solved by the proposed double BFGS method. We also determine an optimal CT scheme as a special case and compare the results of the two schemes.

The main contributions of our research are as follows.

First, we propose the CETE problem in a single-OD network with heterogeneous users and parallel routes and formulate it as an unconstrained optimization problem. This novel formulation approach can also be used to express the corresponding NTE and CTE problems, which can be formulated as static traffic assignment problems using the traditional approach. However, unlike the traditional approach, this novel approach does not require that the arrival of each group of commuters at the destination at the equilibrium follows a predetermined order, and can handle non-monotone (emission) cost function.

Second, we are the pioneer to use and modify the BFGS method to solve the CETE problem and the corresponding NTE and CTE problems. We demonstrate that the BFGS-based method is more efficient than the traditional GSD method for solving the NTE and CTE problems.

Third, we provide a theoretical analysis of the CETE problem.

Fourth, we propose a novel optimal CET design problem for a single-OD network with general user heterogeneity and parallel routes. The problem is formulated as a bi-level program, and solved by the double BFGS method. The proposed methodology can be directly extended for solving the optimal CT counterpart.

The remainder of this paper is organized as follows. In Section 2, two semi-analytical approaches are introduced to solve the NTE problem. In Section 3, the CETE problem is formulated as an unconstrained optimization problem and the solution method for this problem, i.e., the BFGS-based method, is presented. In Section 4, the optimal toll scheme design problems are formulated as bi-level programming problems and the double BFGS method is proposed to solve them. Numerical examples are given in Section 5, and finally, conclusions are provided in Section 6.

2. No-toll equilibrium

2.1. Notations

We consider $\underline{R}$ parallel roads connecting a residential district to the CBD. Let R be the set of the parallel roads (routes), and L_r, $v_r^{\max}$, c_r^0, and τ_r^0 be the length, the maximum travel speed, the fixed travel cost, and the free-flow travel time of road $r \in R$. Consider a fixed number of commuters, *D*, who travel from home to the CBD through the parallel roads during the morning rush hour. Due to traffic congestion, commuters may experience schedule delay early or late. Let $S = \{1, 2\}$ be the set of two types of schedule delays, and $s = 1$ and $s = 2$ denote schedule delay early and late, respectively. To consider user heterogeneity in terms of VOT, the commuters are divided into *n* groups. Let G be the set of groups, D_g, α_g, and β_{gs} be the traffic demand, the unit cost of travel time, and the unit cost of the *s*-th type schedule delay of commuters of group $g \in G$, respectively. For simplicity, all groups are assumed to have an identical desired arrival time t^*.

Using the Henderson approach, the arrival flows of the commuters experiencing schedule delay early are independent of those experiencing schedule delay late and the formulation related to the two types of schedule delays are similar (Chu, 1995). To simplify the model formulation, we propose the concept "relative arrival time", which is defined as follows:

Definition 1 (Relative arrival time). Let t be the arrival time of a commuter. If $t \le t^*$, then the relative arrival time is defined as $\ell = t - t^*$; otherwise, the relative arrival time is defined as $\ell = t^* - t$.

The following notations are adopted to formulate the NTE problem:

$f_{rs}(\ell)$ — flow rate of commuters using road r, experiencing the *s*-th type schedule delay, and having relative arrival time ℓ.

$f_{grs}(\ell)$ — flow rate of commuters of group *g* using road r, experiencing the *s*-th type schedule delay, and having relative arrival time ℓ.

$\overline{f}_{grs}(\ell)$ — flow rate with respect to the isocost curve of commuters of group *g* who use road r, experience the *s*-th type schedule delay, and have relative arrival time ℓ.

ℓ_{grs} — the smallest relative arrival time of commuters of group *g* who use road r and experience the *s*-th type schedule delay.

N_{grs} — number of commuters of group *g* who use road r and experience the *s*-th type schedule delay.

D — vector of traffic demands $[D_q, g \in G]$.

$C_{grs}(\ell)$ — generalized trip cost of commuters of group g using road r, experiencing the *s*-th type schedule delay, and having relative arrival time ℓ.

π_q equilibrium generalized trip cost of commuters of group g.

$\boldsymbol{\pi}$ vector of equilibrium generalized trip costs $[\pi_q, g \in G]$.

$\boldsymbol{\alpha}$ vector of unit costs of travel time $[\alpha_q, g \in G]$.

$\boldsymbol{\beta}$ vector of unit costs of schedule delay $[\beta_{qs}, g \in G, s \in S]$.

$\tau_{rs}(\ell)$ travel time of commuters who use road r, experience the *s*-th type schedule delay, and have relative arrival time ℓ.

$\overline{\tau}_{grs}(\ell)$ travel time with respect to the isocost curve of commuters of group *g* who use road r, experience the *s*-th type schedule delay, and have relative arrival time ℓ.

$v_{rs}(\ell)$ travel speed of commuters who use road r, experience the *s*-th type schedule delay, and have relative arrival time ℓ.

Following Chu (1995) and Coria and Zhang (2017), we assume that traffic flow has a zero group velocity and commuters' travel speed on each road is determined by the arrival flow rate through a power speed-flow function, and

$$\frac{1}{v_{rs}(\ell)} = \frac{1}{v_r^{\max}} + \left[\frac{f_{rs}(\ell)}{\zeta_r Q_r}\right]^{\kappa}, \forall r \in R, s \in S, \ell \le 0, \tag{1}$$

where ζ_r is a positive parameter, Q_r is the capacity of road *r*, and the parameter κ is the elasticity of the travel delay with respect to the arrival flow rate.

By definition, we have $\tau_{rs}(\ell) = L_r / v_{rs}(\ell)$ and $\tau_r^0 = L_r / v_r^{\max}$. Equivalently, they correspond to $1/v_{rs}(\ell) = \tau_{rs}(\ell)/L_r$ and $1/v_r^{\max} = \tau_r^0 / L_r$. Substituting these two equations into the resultant equation obtained after rearranging Eq. (1), we have

$$f_{rs}(\ell) = \zeta_r Q_r \left(\frac{1}{v_{rs}(\ell)} - \frac{1}{v_r^{\max}}\right)^{\frac{1}{\kappa}} = \zeta_r Q_r \left(\frac{\tau_{rs}(\ell) - \tau_r^0}{L_r}\right)^{\frac{1}{\kappa}}, \forall r \in R, s \in S, \ell \le 0. \tag{2}$$

By definition, we also have

$$f_{rs}(\ell) = \sum_{g \in G} f_{grs}(\ell), \forall r \in R, s \in S, \ell \le 0. \tag{3}$$

Based on Eq. (1), we define the following travel time function in terms of the arrival flow rate f for any road $r \in R$:

$$\tau_r(f) = L_r \left[\frac{f}{\zeta_r Q_r}\right]^{\kappa} + \tau_r^0. \tag{4}$$

Based on travel time function (4), we have

$$\tau_{rs}(\ell) = \tau_r(f_{rs}(\ell)) = L_r \left[\frac{f_{rs}(\ell)}{\zeta_r Q_r}\right]^{\kappa} + \tau_r^0, \forall r \in R, s \in S, \ell \le 0. \tag{5}$$

The generalized trip cost of each commuter consists of three components: a fixed travel cost (e.g., parking cost), in-vehicle travel time cost, and a penalty cost associated with schedule delay early or late. The generalized trip cost of commuters with respect to arrival time *t* can be formulated as follows:

$$C_{grs}(\ell) = c_r^0 + \alpha_g \tau_{rs}(\ell) - \beta_{gs}\ell,\ \forall g \in G, r \in R, s \in S, \ell \le 0, \tag{6}$$

where the penalty cost associated with schedule delay early (i.e., $-\beta_{g1}\ell$) represents commuters' disutility of arriving at their destination early. The generalized trip cost function was introduced by Vickrey (1969), and widely used to model morning peak hour congestion (e.g., Arnott et al., 1994; Lindsey, 2004; Liu et al., 2015; Xiao et al., 2015; Guo et al., 2018). Small (1982) conducted an empirical study to estimate this function.

Following Liu et al. (2015), we adopt the following two assumptions on commuters' VOTs:

Assumption A1: $\beta_{g1} < \alpha_g, \forall g \in G$.

Assumption A2: $\beta_{gs} / \alpha_g \ne \beta_{g's} / \alpha_{g'}$, $\forall g \in G, g' \in G \setminus \{g\}, s \in S$.

Assumption A1 states that the unit cost of schedule delay early must be less than that of travel time. This assumption is required for the existence of a meaningful equilibrium (Lindsey, 2004; Liu et al., 2015). Assumption A2 leads to a unique arrival order for each commute group (Liu et al., 2015).

2.2. The no-toll equilibrium condition

By definition, arrival flow rates are non-negative and satisfy the flow conservation condition. Hence, we have

$$f_{grs}(\ell) \ge 0, \forall g \in G, r \in R, s \in S, \ell \le 0 \ \text{ and} \tag{7}$$

$$\sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{grs}(\ell)d\ell = D_g, \forall g \in G . \tag{8}$$

The NTE condition for commuters' route and arrival time choice can be described as follows: No commuter can reduce his/her generalized trip cost by unilaterally altering his/her arrival time and route. The NTE conditions can be mathematically expressed as follows:

$$C_{grs}(\ell)\begin{cases} = \pi_g, & \text{if } f_{grs}(\ell) > 0, \\ \ge \pi_g, & \text{if } f_{grs}(\ell) = 0, \end{cases} \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{9}$$

Definition 2 (NTE flow vector). If the flow vector $\mathbf{f} = [f_{grs}(\ell), g \in G, r \in R, s \in S, \ell \le 0]$ satisfies conditions (7)-(9), then $\mathbf{f}$ is defined as an NTE flow vector.

The NTE solution can be illustrated and analyzed graphically using isocost curves (Arnott et al., 1994; Lindsey, 2004; Chen et al., 2015). As shown in Fig. 1, the isocost curves in a three-group NTE problem for travelers who use road r and experience the *s*-th type schedule delay are presented. The isocost curve of each group represents the group's willingness to pay for each relative arrival time slot. At the NTE, a relative arrival time slot is always assigned to the group with the highest willingness to pay for it. This implies that commuters should always stay on the upper envelope of all the isocost curves at the NTE.

2.3. The arrival order of commuter groups

The same as the bottleneck model with general user heterogeneity in the studies of Arnott et al. (1994) and Liu et al. (2015), the arrival order of group *g* under the NTE condition (9) is determined by the relative cost of schedule delay to travel time, i.e., β_{gs}/α_g. Commuters make a trade-off between their travel time and the corresponding schedule delay when they choose their routes and arrival times. The farther from the center of the rush hour a commuter travels, the lower travel time and the higher schedule delay he/she experiences. Therefore, at the NTE, the commuters with a higher relative cost of schedule delay to travel time (i.e., β_{gs}/α_g) prefer to travel closer to the peak, i.e., have larger relative arrival time. In other words, the arrival order of commuter groups in terms of relative arrival time follows the ascending order of β_{gs}/α_g. Let $g_s(i)$ be the group ID of the *i*-th arrival group experiencing the *s*-th type schedule delay. By definition, $[g_s(i), i \in I, s \in S]$ provides a mapping from arrival order IDs to group IDs, where I is the set of arrival order IDs and $I = \{1, 2, \cdots, n\}$.

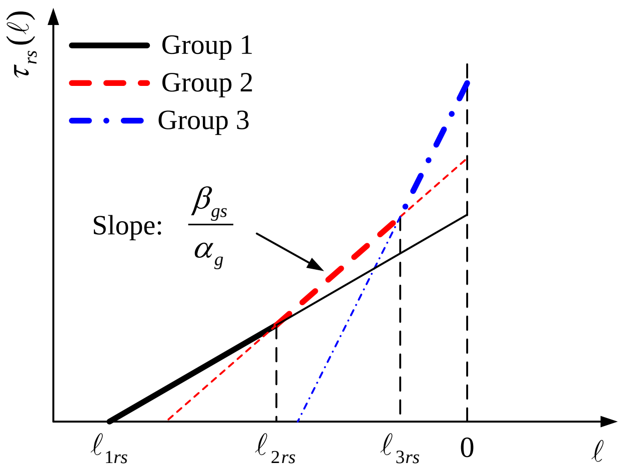

Fig. 1. An illustration of isocost curves at the NTE for travelers who use road r and experience the *s*-th type schedule delay (Thick lines depict the upper envelope of the isocost curves).

2.4. The static traffic assignment problem approach

2.4.1. The basic setting

Following Liu et al. (2015), we formulate the NTE problem with general user heterogeneity as the following static traffic assignment problem, although we use a different approach to model congestion. We consider a network with a set of $n+1$ nodes and a set of $2n\underline{R}$ routes (i.e., the combination of *n* commuter groups, $\underline{R}$ roads, and two types of schedule delays). As shown in Fig. 2, there are *n* origin-destination (OD) pairs and commuters entering the network at node *g* always exit at node $g+1$. The demand between origin g and destination $g+1$ is D_g. There are $2\underline{R}$ routes connecting each OD pair and commuters between each OD pair choose between the $2\underline{R}$ routes. Let (g, r, s) be the combination of indices representing the route connecting OD pair $(g, g+1)$ on which the commuters use road r and experience the s-th type schedule delay. By definition, N_{grs} is the flow of route (g, r, s). All commuters on route (g, r, s) have the same commute cost, denoted by C_{grs}. The detailed derivation of the commute cost of each route is given in Section 2.4.2.

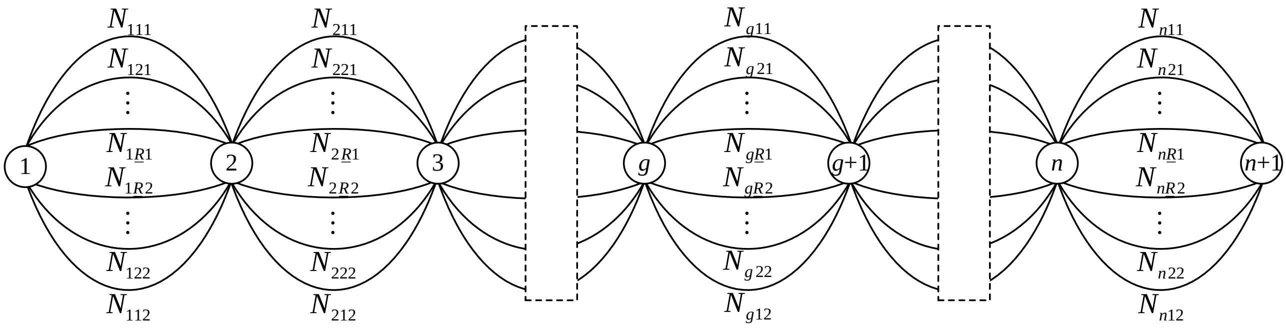

Fig. 2. The equivalent static traffic assignment problem.

2.4.2. The model formulation

As mentioned in Appendix A, using the vector of route flows $\mathbf{N} = [N_{grs}, g \in G, r \in R, s \in S]$, we can retrieve route costs, i.e., $\mathbf{C}(\mathbf{N}) = [C_{grs}, g \in G, r \in R, s \in S]$, and the arrival flow rates of each group, i.e., $\mathbf{f}(\mathbf{N}) = [f_{grs}(\ell), g \in G, r \in R, s \in S, \ell \leq 0]$. The NTE problem can be formulated as the following VI problem:

Theorem 1. Let $\mathbf{N}^* \in \Omega$ solve the following VI problem:

$$\left\langle \mathbf{C}(\mathbf{N}^*), \mathbf{N} - \mathbf{N}^* \right\rangle \geq 0, \forall \mathbf{N} \in \Omega, \tag{10}$$

where Ω is a closed convex set and

$$\Omega = \left\{ \mathbf{N} \geq \mathbf{0} : \sum_{r \in R} \sum_{s \in S} N_{grs} = D_g, \forall g \in G \right\}.$$

Then, $\mathbf{f}(\mathbf{N}^*)$ is an NTE flow vector.

The proof of Theorem 1 is given in Appendix B.1.

Proposition 1: There exists at least one solution to VI problem (10).

The proof of Proposition 1 directly follows that in the study of Liu et al. (2015). Theorem 1 and Proposition 1 imply that there is at least a solution to the NTE problem with general user heterogeneity and route choice.

2.4.3. Solution algorithm: Gauss-Seidel decomposition method

VI problem (10) can be solved by any general computational techniques developed for VI problems provided that the convergence requirement is met. In this study, the GSD method (Pang, 1985; Liu et al., 2015) was adopted to solve VI problem (10). The following gap function was used to evaluate the quality of the computed solutions to the NTE problem:

$$G_1(\mathbf{N}) = 1 - \frac{\sum_{g \in G} N_g C_g^{\min}}{\sum_{g \in G} \sum_{r \in R} \sum_{s \in S} N_{grs} C_{grs}}, \tag{11}$$

where $C_g^{\min} = \min_{r \in R, s \in S} \{C_{grs}\}$. If $\mathbf{N}$ is a solution to VI problem (10), then $G_1(\mathbf{N})$ equals zero. Otherwise, the gap value is positive.

Let $\mathbf{N}_g = [N_{grs}, r \in R, s \in S]$ and $\mathbf{N}_{g-} = [N_{g'rs}, g' \in G \setminus \{g\}, r \in R, s \in S]$. By definition, we have $\mathbf{N} = (\mathbf{N}_g, \mathbf{N}_{g-})$ for all $g \in G$. We define the function $\mathbf{C}_g(\mathbf{N}_g, \mathbf{N}_{g-}) = [C_{grs}, r \in R, s \in S]$, and the following VI problem, denoted by $VI\left(\Omega_g, \mathbf{C}_g(\mathbf{N}_g, \mathbf{N}_{g-})\right)$:

$$\left\langle \mathbf{C}_g(\mathbf{N}_g^*, \mathbf{N}_{g-}), \mathbf{N}_g - \mathbf{N}_g^* \right\rangle \geq 0, \forall \mathbf{N}_g \in \Omega_g, \tag{12}$$

where $\mathbf{N}_{g-}$ is an input to VI problem (12), and

$$\Omega_g = \left\{ \mathbf{N}_g \geq \mathbf{0}, \sum_{r \in R} \sum_{s \in S} N_{grs} = D_g \right\}.$$

Proposition 2: VI problem (12) can be equivalently formulated as the following optimization problem:

$$\min_{\mathbf{N}_g \in \Omega_g} \sum_{r \in R} \sum_{s \in S} \int_0^{N_{grs}} C_{grs}(x, \mathbf{N}_{g-}) dx, \tag{13}$$

where $C_{grs}(x, \mathbf{N}_{g-})$ is defined by Eqs. (A.11) and (A.13) by replacing N_{grs} with *x* in Eq. (A.11).

The proof is given in Appendix B.2.

The GSD method presented in Algorithm 1 was adopted to solve VI problem (10). In Algorithm 1, we set the initial solution $\mathbf{N}_0 = [N_{grs} = D_g / (2\underline{R}), \forall g \in G, r \in R, s \in S]$, i.e., the traffic demand of each group is evenly assigned to each route. According to Proposition 2, the sub-problem $VI\left(\Omega_g, \mathbf{C}_g(\mathbf{N}_g, \mathbf{N}_{g-})\right)$ can be reformulated as optimization problem (13), which was solved by the algorithm proposed by Dial (2006).

Algorithm 1 The GSD method for the NTE problem

Inputs: A feasible initial solution $\mathbf{N}_0$ and the convergence tolerance ε_1.

```
Set N = N_0.
while G_1(N) ≥ ε_1 do
    for each s ∈ S do
        for i = 1, 2, ⋯ n do
            Set g = g_s(i).
            Solve the problem VI(Ω_g, C_g(N_g, N_g−)) and obtain its solution N_g*.
            Set N_g = N_g*.
        end for
    end for
end while
```

Output: $\mathbf{N}$.

2.5. The unconstrained optimization problem approach

2.5.1. Overview

We initially set a given vector of equilibrium generalized trip costs $\boldsymbol{\pi}$ for a given demand vector $\mathbf{D}$ at NTE. For any group $g \in G$, ignoring other groups, we have an NTE problem with homogeneous users, i.e., users of group *g*, and obtain an equilibrium travel time curve and an equilibrium arrival flow rate curve for each type of schedule delay and each road $r \in R$ (see Chu (1995) for details). We then obtain the upper envelope of all the equilibrium arrival flow rate/travel time curves. At the NTE, a relative arrival time slot is always assigned to the group with the highest equilibrium arrival flow rate, or equivalently the group with the highest travel time (i.e., the highest willingness to pay for the relative arrival time slot). This implies that commuters should always stay on the upper envelope of all the equilibrium arrival flow rate curves at the NTE. Therefore, using the upper envelope of all the equilibrium arrival flow rate curves, we retrieve the number of commuters of the group who use each route. Taking the summation of the retrieved number of commuters of each group on all routes, we retrieve the traffic demand of each group. Let $\tilde{D}_g(\boldsymbol{\pi})$ be the retrieved traffic demand of group *g* from the vector of equilibrium generalized trip costs $\boldsymbol{\pi}$ and $\tilde{\mathbf{D}}(\boldsymbol{\pi}) = [\tilde{D}_g(\boldsymbol{\pi}), g \in G]$ be the vector function of retrieved traffic demands. If $\left[\tilde{\mathbf{D}}(\boldsymbol{\pi}) - \mathbf{D}\right]^T \left[\tilde{\mathbf{D}}(\boldsymbol{\pi}) - \mathbf{D}\right]$ is equal or close to zero (meaning that $\tilde{\mathbf{D}}(\boldsymbol{\pi}) \approx \mathbf{D}$), the equilibrium generalized trip costs for the NTE problem with the demand vector $\mathbf{D}$ is approximately $\boldsymbol{\pi}$, and we obtain the equilibrium flow pattern from $\boldsymbol{\pi}$. Otherwise, we update the vector $\boldsymbol{\pi}$ by the BFGS-based method and repeat the above procedure until $\tilde{\mathbf{D}}(\boldsymbol{\pi}) \approx \mathbf{D}$.

2.5.2. Retrieving traffic demands from equilibrium generalized trip costs

By definition, the equilibrium travel time curve of each group can be formulated as follows:

$$\overline{\tau}_{grs}(\ell) = \frac{\pi_g - c_r^0 + \beta_{gs}\ell}{\alpha_g}, \ \forall g \in G, r \in R, s \in S, \ell \le 0\,. \tag{14}$$

Let $\overline{\ell}_{grs}$ be the relative arrival time at which the equilibrium travel time of commuters of group *g* who use road r and experience the *s*-th type schedule delay equals the free flow travel time of road *r*. According to this definition, we have $\overline{\tau}_{grs}(\overline{\ell}_{grs}) = \tau_r^0$ and

$$\overline{\ell}_{grs} = -(\pi_g - c_r^0 - \alpha_g \tau_r^0) / \beta_{gs}\,. \tag{15}$$

As commuters' travel time on a road is not less than the free flow travel time of this road, we have $\overline{\tau}_{grs}(\ell) \ge \tau_r^0$. Substituting Eqs. (14) and (15) into this inequality and simplifying the resultant inequality, we have $\ell \ge \overline{\ell}_{grs}$. Thus, the equilibrium travel time curve for the commuters of group g entering road r, experiencing the *s*-th type schedule delay, and having relative arrival time ℓ is well defined on $\ell \in [\overline{\ell}_{grs}, 0]$.

The equilibrium travel time curve and the equilibrium arrival flow rate curve can be related by travel time function (4), and hence we have

$$\bar{\tau}_{grs}(\ell) = \tau_r(\bar{f}_{grs}(\ell)) = L_r \left[\frac{\bar{f}_{grs}(\ell)}{\zeta_r Q_r} \right]^{\kappa} + \tau_r^0, \forall r \in R, s \in S, \ell \in [\bar{\ell}_{grs}, 0]. \tag{16}$$

For all $\ell \in (-\infty, -(\pi_g - c_r^0 - \alpha_g \tau_r^0)/\beta_{gs})$, based on Eq. (6), we have

$$C_{grs}(\ell) = c_r^0 + \alpha_g \tau_{rs}(\ell) - \beta_{gs}\ell > c_r^0 + \alpha_g \tau_r^0 - \beta_{gs}\bar{\ell}_{grs} > \pi_g, \forall r \in R, s \in S, \ell \in (-\infty, \bar{\ell}_{grs}). \tag{17}$$

Based on Eq. (16) and inequality (17), we can obtain the equilibrium arrival flow rate curve of each group as follows:

$$\bar{f}_{grs}(\ell) = \begin{cases} \zeta_r Q_r \left(\dfrac{\pi_g - c_r^0 + \beta_{gs}\ell}{\alpha_g L_r} - \dfrac{\tau_r^0}{L_r} \right)^{\frac{1}{\kappa}}, & \text{if } \ell > \bar{\ell}_{grs}, \\ 0, & \text{if } \ell \leq \bar{\ell}_{grs}, \end{cases} \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{18}$$

Substituting $f_{rs}(\ell) = \bar{f}_{grs}(\ell)$ and Eq. (18) into Eq. (5), and substituting the resultant equation into Eq. (6), we have $C_{grs}(\ell) = \pi_g$ if $\bar{\ell}_{grs} \leq \ell \leq 0$, and $C_{grs}(\ell) > \pi_g$ otherwise. This verifies that the function $\bar{f}_{grs}(\ell)$ forms an equilibrium arrival flow rate curve of group *g*. Using Eq. (18), we can retrieve the overall arrival flow rate for each road and schedule delay type, i.e., the upper envelope of equilibrium arrival flow rate curves of all groups, as follows:

$$f_{rs}(\ell) = \max_{g \in G}\{\bar{f}_{grs}(\ell)\}, \forall r \in R, s \in S, \ell \leq 0. \tag{19}$$

Theorem 2. The following arrival flow pattern satisfies NTE condition (9):

$$f_{grs}(\ell) = \begin{cases} \bar{f}_{grs}(\ell), & \text{if } f_{rs}(\ell) = \bar{f}_{grs}(\ell), \\ 0, & \text{otherwise}, \end{cases} \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{20}$$

The proof is given in Appendix B.3.

Using Eq. (20), we can retrieve the number of commuters of each group on each route and the traffic demand of each group from the vector of equilibrium generalized trip costs, respectively:

$$N_{grs}(\boldsymbol{\pi}) = \int_{-\infty}^{0} f_{grs}(\ell)\, d\ell, \forall g \in G, r \in R, s \in S \text{ and} \tag{21}$$

$$\tilde{D}_g(\boldsymbol{\pi}) = \sum_{r \in R}\sum_{s \in S}\int_{-\infty}^{0} f_{grs}(t)\, dt = \sum_{r \in R}\sum_{s \in S} N_{grs}(\boldsymbol{\pi}), \forall g \in G. \tag{22}$$

The detailed procedure for computing $\tilde{D}_g(\boldsymbol{\pi})$ is provided in Appendix C.1.

2.5.3. The model formulation

Based on the preceding discussion, the NTE problem with general user heterogeneity and route choice can be formulated as a system of nonlinear equations as follows:

$$\mathbf{Z}(\boldsymbol{\pi}) = \tilde{\mathbf{D}}(\boldsymbol{\pi}) - \mathbf{D} = \mathbf{0}. \tag{23}$$

Theorem 3. Let $\boldsymbol{\pi}^* = [\pi_g^*, g \in G]$ be a solution to the system of nonlinear equations (23). Then

$\tilde{\mathbf{f}}(\boldsymbol{\pi}^*) = [f_{grs}(\ell), g \in G, r \in R, s \in S, \ell \le 0]$ whose elements are defined by Eq. (20) is an NTE flow vector.

The proof is given in Appendix B.4.

Theorem 4. Let $\mathbf{f}^*$ be an NTE flow vector and $\boldsymbol{\pi}^*$ be the corresponding vector of equilibrium generalized trip costs. Then we have $\mathbf{f}^* = \tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$ and $\mathbf{Z}(\boldsymbol{\pi}^*) = \mathbf{0}$.

The proof is given in Appendix B.5.

Theorems 3 and 4 imply that the system of nonlinear equations (23) is an alternative way to formulate the NTE problem with general user heterogeneity and route choice. Theorem 1 and Proposition 1 guarantees the existence of a solution to the NTE problem with general user heterogeneity and route choice. Hence, the system of nonlinear equations (23) must have a solution.

Theorem 5. Let $\boldsymbol{\pi}^* = [\pi_g^*, g \in G]$ be an optimal solution to the following unconstrained optimization problem:

$$\min_{\boldsymbol{\pi}} \eta(\boldsymbol{\pi}) = \mathbf{Z}(\boldsymbol{\pi})^{\mathrm{T}} \mathbf{Z}(\boldsymbol{\pi}). \tag{24}$$

Then $\boldsymbol{\pi}^*$ is a solution to the system of nonlinear equations (23) and the flow vector $\tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$ is an NTE flow vector.

The proof is given in Appendix B.6.

2.5.4. The elimination of zero retrieved traffic demand

For a given generalized trip cost vector $\boldsymbol{\pi}$, if there exists a group g such that its positive equilibrium arrival flow rate is strictly lower than the highest equilibrium arrival flow rate on all roads for the whole studied period, i.e., $\overline{f}_{grs}(\ell) < f_{rs}(\ell)$ is satisfied for all $r \in R$, $s \in S$, and $\ell \in \{\ell' | \overline{f}_{grs}(\ell') > 0\}$. According to Eq. (20), we have $f_{grs}(\ell) = 0$ for all $r \in R$, $s \in S$, and $\ell \le 0$. This implies that the retrieved traffic demand $\tilde{D}_g(\boldsymbol{\pi}) = 0$. According to Eq. (18), $\overline{f}_{grs}(\ell)$ is continuous and monotonically non-decreasing with respect to π_g. This implies that the positive equilibrium arrival flow rate of group g is still strictly lower than the highest equilibrium arrival flow rate on all roads for the whole studied period and the retrieved traffic demand of this group still equals zero if the increase in π_g is not large enough. Hence, we have $\partial \eta(\boldsymbol{\pi}) / \partial \pi_g = 0$. This implies that π_g cannot be updated by gradient descent algorithms. Hence, zero retrieved traffic demand should be eliminated during the solution procedures. We redefine the function $\tilde{D}_g(\boldsymbol{\pi})$ by the function $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-})$, where $\boldsymbol{\pi}_{g-} = [\pi_{g'}, g' \in G \setminus \{\mathrm{g}\}]$. We have the following proposition:

Proposition 3: $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-})$ is monotonically increasing with respect to π_g.

The proof is given in Appendix B.7.

Algorithm 2 presents the pseudocode of the method for eliminating zero retrieved traffic demand. The one-dimensional search problem in Line 3 of Algorithm 2 can be solved by the bi-section or interpolation method.

Algorithm 2 The elimination of zero retrieved traffic demand

Input: A given solution $\boldsymbol{\pi}$.

while $\tilde{\mathbf{D}}(\boldsymbol{\pi}) > \mathbf{0}$ is not satisfied **do**
    for each $g \in G$ **do**
        Solve the one-dimensional search problem $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-}) = D_g$ and obtain its solution π_g^*.
        Set $\pi_g = \pi_g^*$.
    end for
end while

Output: The updated $\boldsymbol{\pi}$.

2.5.5. Solution algorithm: The BFGS-based method

In this study, unconstrained optimization problem (24) was solved by the proposed method developed based on the framework of the BFGS method. The BFGS method is one of the most popular quasi-Newton methods. The advantages of the BFGS method over a Newton method are that the convergence of the former is superlinear and fast and that its computational complexity is significantly lower than that of the Newton method. The proposed BFGS-based method for solving unconstrained optimization problem (24) is summarized by Algorithm 3.

Algorithm 3 The BFGS-based method for the NTE problem

Inputs: A feasible initial solution $\boldsymbol{\pi}_0$, the convergence tolerance ε_2, the constants σ and $\rho \in (0,1)$, and an initial symmetric positive definite matrix $\mathbf{H}_0$.

Set $\boldsymbol{\pi} = \boldsymbol{\pi}_0$, $\mathbf{H} = \mathbf{H}_0$.
while $\|\mathbf{Z}(\boldsymbol{\pi})\| \geq \varepsilon_2$ **do**
    Implement Algorithm 2 to eliminate zero traffic demand and update $\boldsymbol{\pi}$.
    Set $\mathbf{d} = -\mathbf{H}\nabla\eta(\boldsymbol{\pi})$.
    Find the smallest non-negative integer *m* such that $\eta(\boldsymbol{\pi}) - \eta(\boldsymbol{\pi} + \rho^m \mathbf{d}) \geq -\sigma\rho^m \nabla\eta(\boldsymbol{\pi})^T \mathbf{d}$.
    Set $\mathbf{s} = \rho^m \mathbf{d}$, $\mathbf{y} = \nabla\eta(\boldsymbol{\pi} + \rho^m \mathbf{d}) - \nabla\eta(\boldsymbol{\pi})$, and $\boldsymbol{\pi} = \boldsymbol{\pi} + \rho^m \mathbf{d}$.
    if $\mathbf{s}^T\mathbf{y} > 0$ then
        Set $\mathbf{H} = \mathbf{H} + \dfrac{(\mathbf{s}^T\mathbf{y} + \mathbf{y}^T\mathbf{H}\mathbf{y})(\mathbf{s}\mathbf{s}^T)}{(\mathbf{s}^T\mathbf{y})^2} - \dfrac{\mathbf{H}\mathbf{y}\mathbf{s}^T + \mathbf{s}\mathbf{y}^T\mathbf{H}}{\mathbf{s}^T\mathbf{y}}$.
    end if
end while

Output: $\boldsymbol{\pi}$.

There are four basic components for using the BFGS method to solve an unconstrained optimization problem: (i) providing an initial solution, (ii) checking the termination criterion, (iii) computing the gradient of the objective function, and (iv) using a line search to find a proper step size in each iteration. The detail setting of the four basic components in Algorithm 3 is as follows:

(i) Any vector $\boldsymbol{\pi}_0$ can be used as the initial solution for Algorithm 3. However, in our BFGS-based method, we further need to eliminate zero retrieved traffic demand. This step is handled by Algorithm 2. Note that implementing Algorithm 2 to eliminate zero retrieved traffic demand can be time-consuming when $\boldsymbol{\pi}_0$ is far away from an optimal solution to unconstrained optimization problem (24). This is because there can be many zero demands. In this study, we used Algorithm 1 to get an approximate solution to VI problem (10) and used the corresponding average generalized trip cost of each group as its initial equilibrium generalized trip cost in Algorithm 3.

(ii) The termination criterion $\|\mathbf{Z}(\boldsymbol{\pi})\| < \varepsilon_2$ was used in this study. If this was not satisfied, Lines 3-8 would be executed. To check the termination criterion, the value of $\|\mathbf{Z}(\boldsymbol{\pi})\|$ is determined beforehand, which in turn requires determining $\tilde{\mathbf{D}}(\boldsymbol{\pi})$. In this study, Procedure C.1 in Appendix C.1 was used to obtain $\tilde{\mathbf{D}}(\boldsymbol{\pi})$ and Eq. (23) was used to compute $\mathbf{Z}(\boldsymbol{\pi})$ so as to calculate $\|\mathbf{Z}(\boldsymbol{\pi})\|$.

(iii) The gradient of the objective function of unconstrained optimization problem (24), which is required in Lines 4-6, were obtained by Procedure C.2 in Appendix C.2.

(iv) For the line search (Line 5 of Algorithm 3), the Armijo Rule was applied to determine the step size (Shi and Shen, 2005).

3. Congestion and environmental toll equilibrium

3.1. The approximation of the emission cost function

We consider that $\underline{H}$ types of pollutants are generated by the vehicles on each road $r \in R$. Let $H = \{1, 2, \cdots, \underline{H}\}$ be the set of pollutants. It is assumed that the emission rate of each type of pollutant from a vehicle is a function of the average speed of the vehicle. The average speed of a vehicle on road $r \in R$ that arrives at the destination at time *t* is a function of the corresponding travel time, i.e., $\tau_{rs}(\ell) = L_r / v_{rs}(\ell)$, and hence the emission rate of each type of pollutant from a vehicle is a function of the travel time of the vehicle. Let $\varphi_r^h(\tau)$ be the emission rate of pollutant *h* from the vehicles with a travel time of τ on road $r \in R$. The total emission cost of all pollutants per vehicle with a travel time of τ on this road can be formulated as

$$E_r(\tau) = \sum_{h \in H} \lambda_h \varphi_r^h(\tau), \tag{25}$$

where λ_h is the unit emission cost of pollutant *h* and in \$/g, $\varphi_r^h(\tau)$ is in g/veh, and $E_r(\tau)$ is in \$/veh.

We assume that the emission cost function $E_r(\tau)$ is convex, i.e., $d^2 E_r(\tau) / d\tau^2 > 0$. As shown in Fig. 3, linear interpolation is applied to approximate the emission cost function $E_r(\tau)$. Let $\underline{K}$ be the number of linear interpolants and $K = \{1, 2, \cdots, \underline{K}\}$, and $\tau_{r,k}$ be the start (smallest) value of τ on the *k*-th linear interpolant. By definition, we have $\tau_{r,k} < \tau_{r,k+1}$. As $\tau_{rs}(\ell) \geq \tau_r^0$, we set $\tau_{r,1} = \tau_r^0$. With the linear interpolation, $E_r(\tau)$ can be approximated by the following function:

$$\tilde{E}_r(\tau) = a_{r,k}\tau + b_{r,k}, \quad \forall r \in R, k \in K, \tau \in [\tau_{r,k}, \tau_{r,k+1}], \tag{26}$$

where

$$a_{r,k} = \frac{E_r(\tau_{r,k+1}) - E_r(\tau_{r,k})}{\tau_{r,k+1} - \tau_{r,k}} \text{ and } b_{r,k} = \frac{\tau_{r,k+1}E_r(\tau_{r,k}) - \tau_{r,k}E_r(\tau_{r,k+1})}{\tau_{r,k+1} - \tau_{r,k}}.$$

3.2. A second-best congestion and environmental toll scheme

We add the superscript "CETE" for the notations associated with the NTE to represent the CETE counterparts, at which no commuters can reduce his/her generalized trip cost (including the CET) by unilaterally altering his/her arrival time. We also use the superscript "CET" in some notations to mean that the corresponding variables or functions depend on the CET scheme. Using the approximated emission cost function, we can express the total system cost (TSC) of commuting by cars, i.e., the sum of total system travel cost (TSTC) and total system environmental cost (TSEC), as follows:

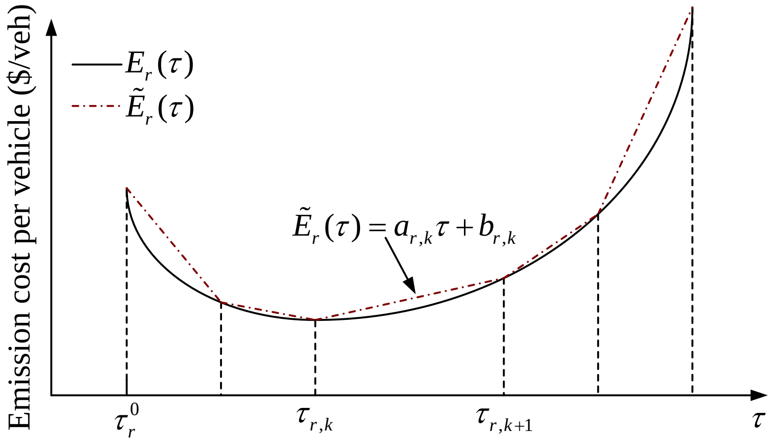

Fig. 3. The linear interpolation of an emission cost function on road *r*.

$$\tilde{\eta} = \sum_{g\in G}\sum_{r\in R}\sum_{s\in S}\int_{\ell_s}^{0} f_{grs}(\ell)[C_{grs}(\ell) + \tilde{E}_r(\tau_{rs}(\ell))]d\ell. \tag{27}$$

Substituting Eqs. (3), (5), (6), and (26) into Eq. (27) and taking the derivative of the resultant expression, we can obtain the marginal social cost $MC_{grs}^{CET}(\ell)$ as follows:

$$MC_{grs}^{CET}(\ell) = \frac{\partial\tilde{\eta}}{\partial f_{grs}(\ell)} = C_{grs}(\ell) + [\alpha_g\kappa + a_{r,k}(\kappa+1)]\tau_{rs}(\ell) + b_{r,k} - (\alpha_g + a_{r,k})\kappa\tau_r^0,$$
$$\forall g\in G, r\in R, s\in S, k\in K, \ell\in[\vec{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}], \tag{28}$$

where $\vec{\ell}_{rs,k} = \max\{\ell \mid \tau_{rs}(\ell) = \tau_{r,k}\}$ and $\overleftarrow{\ell}_{rs,k} = \min\{\ell \mid \tau_{rs}(\ell) = \tau_{r,k+1}\}$. By definition, we have $\tau_{rs}(\overleftarrow{\ell}_{rs,k}) = \tau_{rs}(\vec{\ell}_{rs,k+1}) = \tau_{r,k+1}$.

Based on Eq. (28), we have

$$MC_{grs}^{CET}(\ell) - C_{grs}(\ell) = [\alpha_g\kappa + a_{r,k}(\kappa+1)]\tau_{rs}(\ell) + b_{r,k} - (\alpha_g + a_{r,k})\kappa\tau_r^0,$$
$$\forall g\in G, r\in R, s\in S, k\in K, \ell\in[\vec{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]. \tag{29}$$

Eq. (29) implies that the difference between the marginal social cost $MC_{grs}^{CET}(\ell)$ and the marginal private cost $C_{grs}(\ell)$ depends on commuters' VOT. To obtain an anonymous toll scheme, we design the following

CET scheme:

$$p_{rs}^{CET}(\ell)=\begin{cases}[\alpha_r^{CET}\kappa+a_{r,k}(\kappa+1)]\tau_{rs}(\ell)+b_{r,k}-(\alpha_r^{CET}+a_{r,k})\kappa\tau_r^0, & \text{if } \ell\in[\overrightarrow{\ell}_{rs,k},\overleftarrow{\ell}_{rs,k}],\\ \max_{g\in G}\{\pi_g^{CET}-c_r^0-\alpha_g\tau_{k+1}+\beta_{gs}\ell\}, & \text{if } \ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1}),\end{cases} \tag{30}$$
$$\forall r\in R, s\in S, k\in K, \ell\le 0,$$

where α_r^{CET} is a parameter associated with the CET scheme for road r and the vector $\boldsymbol{\alpha}^{CET}=[\alpha_r^{CET},\forall r\in R]$ is required to be optimized for the scheme. For the CETE problem in this section, $\boldsymbol{\alpha}^{CET}$ is assumed to be given. However, an optimal vector of $\boldsymbol{\alpha}^{CET}$ is obtained by a bi-level program presented in Section 4.1.

In the proposed CET scheme (30), the toll charge $p_{rs}^{CET}(\ell)$ for $\ell\in[\overrightarrow{\ell}_{rs,k},\overleftarrow{\ell}_{rs,k}]$ is derived from the concept of marginal-cost pricing. Under this marginal-cost-based pricing, we have $\tau_{rs}(\overleftarrow{\ell}_{rs,k})=\tau_{rs}(\overrightarrow{\ell}_{rs,k+1})=\tau_{r,k+1}$. The toll charge $p_{rs}^{CET}(\ell)$ for $\ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1})$ is designed to maintain this constant travel time during this period. The price $\pi_g^{CETE}-c_r^0-\alpha_g\tau_{k+1}+\beta_{gs}\ell$ can maintain this constant travel time for group g. In the proposed CET scheme, the largest price for $\ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1})$ is adopted to achieve the CETE.

The generalized trip cost of commuters after implementing the CET scheme can be expressed as follows:

$$C_{grs}^{CETE}(\ell)=c_r^0+\alpha_g\tau_{rs}(\ell)-\beta_{gs}\ell+p_{rs}^{CET}(\ell),\ \forall g\in G, r\in R, s\in S, \ell\le 0. \tag{31}$$

Substituting the first condition of Eq. (30) and $\tau_{rs}^{CETE}(\ell)=\tau_{k+1}$ for all $\ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1})$ into Eq. (31), we have

$$C_{grs}^{CETE}(\ell)=\begin{cases}c_{r,k}^0+\alpha_{gr,k}\tau_{rs}(\ell)-\beta_{gs}\ell, & \text{if } \ell\in[\overrightarrow{\ell}_{rs,k},\overleftarrow{\ell}_{rs,k}],\\ \alpha_g\tau_{k+1}-\beta_{gs}\ell+\max_{g\in G}\{\pi_g^{CETE}-\alpha_g\tau_{k+1}+\beta_{gs}\ell\}, & \text{if } \ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1}),\end{cases} \tag{32}$$
$$\forall g\in G, r\in R, s\in S, k\in K, \ell\le 0,$$

where $c_{r,k}^0=c_r^0-(\alpha_r^{CET}+a_{r,k})\kappa\tau_r^0+b_{r,k}$ and $\alpha_{gr,k}=\alpha_g+\alpha_r^{CET}\kappa+a_{r,k}(\kappa+1)$.

3.3. The congestion and environmental toll equilibrium condition

By definition, arrival flow rates are non-negative and satisfy the flow conservation condition. Hence, we have

$$f_{grs}^{CETE}(\ell)\ge 0,\forall g\in G, r\in R, s\in S, \ell\le 0 \ \text{ and} \tag{33}$$

$$\sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{grs}^{CETE}(\ell)d\ell=D_g,\forall g\in G. \tag{34}$$

The CETE condition for commuters' route and arrival time choice can be mathematically expressed as

$$C_{grs}^{CETE}(\ell)\begin{cases}=\pi_g^{CETE}, & \text{if } f_{grs}^{CETE}(\ell)>0,\\ \ge\pi_g^{CETE}, & \text{if } f_{grs}^{CETE}(\ell)=0,\end{cases}\forall g\in G, r\in R, s\in S, \ell\le 0. \tag{35}$$

Definition 3 (CETE flow vector). If the flow vector $\mathbf{f}^{CETE}=[f_{grs}^{CETE}(\ell), g\in G, r\in R, s\in S, \ell\le 0]$ satisfies conditions (33)-(35), then $\mathbf{f}^{CETE}$ is defined as a CETE flow vector.

At the CETE, the arrival of each group at the destination in the whole studied period does not follow a predetermined order. Therefore, the CETE problem cannot be formulated as a static traffic assignment problem in the same way as the NTE problem. However, similar to solving the NTE problem, the proposed unconstrained optimization problem approach can be extended to solve the CETE problem, which involves retrieving traffic demand from equilibrium generalized trip cost discussed next.

3.4. Retrieving traffic demand from equilibrium generalized trip cost

Let $\vec{\ell}_{grs,k}$ and $\overleftarrow{\ell}_{grs,k}$ be two relative arrival times such that $\tau_{grs}(\vec{\ell}_{grs,k}) = \tau_{r,k}$ and $\tau_{grs}(\overleftarrow{\ell}_{grs,k}) = \tau_{p,k+1}$, respectively. Based on $C_{grs}^{CETE}(\vec{\ell}_{grs,k}) = C_{grs}^{CETE}(\overleftarrow{\ell}_{grs,k}) = \pi_g^{CETE}$ and using Eq. (31), we have

$$\vec{\ell}_{grs,k} = -\frac{\pi_g^{CETE} - c_{r,k}^0 - \alpha_{gr,k}\tau_{r,k}}{\beta_{gs}} \quad \text{and} \quad \overleftarrow{\ell}_{grs,k} = -\frac{\pi_g^{CETE} - c_{r,k}^0 - \alpha_{gr,k}\tau_{r,k+1}}{\beta_{gs}}. \tag{36}$$

For all group $g \in G$, road $r \in R$, and type of schedule delay $s \in S$, we define the following two functions of relative arrival time ℓ on $\ell \in (-\infty, 0]$:

$$\overline{f}_{grs}^{CETE}(\ell) = \begin{cases} \zeta_r Q_r \left(\dfrac{\pi_g^{CETE} - c_{r,k}^0 + \beta_{gs}\ell - \alpha_{gr,k}\tau_r^0}{\alpha_{gr,k} L_r} \right)^{\frac{1}{\kappa}}, & \text{if there exists } k \in K \text{ such that } \ell \in [\vec{\ell}_{grs,k}, \overleftarrow{\ell}_{grs,k}], \\ \zeta_r Q_r \left(\dfrac{\tau_{r,k+1} - \tau_r^0}{L_r} \right)^{\frac{1}{\kappa}}, & \text{if there exists } k \in K \text{ such that } \ell \in (\overleftarrow{\ell}_{grs,k}, \vec{\ell}_{grs,k+1}), \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{37}$$

$$\overline{p}_{grs}^{CET}(\ell) = \begin{cases} [\alpha_r^{CET}\kappa + a_{r,k}(\kappa+1)]\tau_{rs}(\ell) + b_{r,k} - (\alpha_r^{CET} + a_{r,k})\kappa\tau_r^0, & \text{if there exists } k \in K \text{ such that } \ell \in [\vec{\ell}_{grs,k}, \overleftarrow{\ell}_{grs,k}], \\ \pi_g^{CETE} - c_r^0 - \alpha_g \tau_{r,k+1} + \beta_{gs}\ell, & \text{if there exists } k \in K \text{ such that } \ell \in (\overleftarrow{\ell}_{grs,k}, \vec{\ell}_{grs,k+1}), \\ 0, & \text{otherwise.} \end{cases} \tag{38}$$

Using Eqs. (37) and (38), we can retrieve the CET scheme and the arrival flow rate for each road at the CETE as follows, respectively:

$$f_{rs}^{CETE}(\ell) = \max_{g \in G}\{\overline{f}_{grs}^{CETE}(\ell)\}, \forall r \in R, s \in S, \ell \le 0 \quad \text{and} \tag{39}$$

$$p_{rs}^{CET}(\ell) = \max_{g\in G}\{\overline{p}_{grs}^{CET}(\ell)\}, \forall r\in R, s\in S, \ell \le 0 . \quad (40)$$

Theorem 6. If CET scheme (40) is implemented, the following arrival flow pattern satisfies CETE condition (35):

$$f_{grs}^{CETE}(\ell) = \begin{cases} \overline{f}_{grs}^{CETE}(\ell), & \text{if } f_{rs}^{CETE}(\ell) = \overline{f}_{grs}^{CETE}(\ell) \text{ and } p_{rs}^{CET}(\ell) = \overline{p}_{grs}^{CET}(\ell), \\ 0, & \text{otherwise,} \end{cases} \forall g\in G, r\in R, s\in S, \ell \le 0. \quad (41)$$

The proof is given in Appendix B.8.

Using Eq. (41), we can retrieve the number of commuters of each group by the equilibrium generalized trip cost vector:

$$N_{grs}^{CETE}(\boldsymbol{\pi}^{CETE}) = \int_{-\infty}^{0} f_{grs}^{CETE}(\ell)d\ell, \forall g\in G, r\in R, s\in S \quad \text{and} \quad (42)$$

$$\tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE}) = \sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{grs}^{CETE}(\ell)d\ell = \sum_{r\in R}\sum_{s\in S} N_{grs}^{CETE}(\boldsymbol{\pi}^{CETE}), \forall g\in G . \quad (43)$$

The detailed procedure for computing $\tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE})$ is given in Appendix D.1.

3.5. The model formulation

Based on the preceding discussion, the CETE problem with general user heterogeneity and route choice can be formulated as a system of nonlinear equations as follows:

$$\mathbf{Z}^{CETE}(\boldsymbol{\pi}^{CETE}) = \tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE}) - \mathbf{D} = \mathbf{0} , \quad (44)$$

where $\tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE}) = [\tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE}), g\in G]$ denotes the vector mapping function of the system of equations in which $\tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE})$ is defined by Eq. (43).

Theorem 7. Let $\boldsymbol{\pi}^{CETE*}$ be a solution to the system of nonlinear equations (44). The corresponding vector $\mathbf{f}^{CETE}(\boldsymbol{\pi}^{CETE*})$ is a CETE flow vector, where $\mathbf{f}^{CETE}(\boldsymbol{\pi}^{CETE}) = [f_{grs}^{CETE}(\ell), g\in G, r\in R, s\in S, \ell\le 0]$ is a vector function of $\boldsymbol{\pi}^{CETE}$ and defined by Eq. (41).

The proof is similar to that of Theorem 3.

Theorem 8. Let $\mathbf{f}^{CETE*}$ be a CETE flow vector and $\boldsymbol{\pi}^{CETE*}$ be the corresponding vector of equilibrium generalized trip costs. Then we have $\mathbf{f}^{CETE*} = \tilde{\mathbf{f}}(\boldsymbol{\pi}^{CETE*})$ and $\mathbf{Z}^{CETE}(\boldsymbol{\pi}^{CETE*}) = \mathbf{0}$, where $\tilde{\mathbf{f}}(\boldsymbol{\pi}^{CETE*}) = [f_{grs}^{CETE}(\ell), g\in G, r\in R, s\in S, \ell\le 0]$ and its elements are defined by Eq. (41).

The proof is similar to that of Theorem 4. Theorems 7 and 8 imply that the system of nonlinear equations (44) can be used to formulate the CETE problem with general user heterogeneity and route choice.

Theorem 9. Let $\boldsymbol{\pi}^{CETE*}$ be an optimal solution to the following unconstrained optimization problem:

$$\min_{\boldsymbol{\pi}^{CETE}} \eta^{CETE}(\boldsymbol{\pi}^{CETE}) = \mathbf{Z}^{CETE}(\boldsymbol{\pi}^{CETE})^{\mathrm{T}}\mathbf{Z}^{CETE}(\boldsymbol{\pi}^{CETE}) . \quad (45)$$

Then $\boldsymbol{\pi}^{CETE*}$ is a solution to the system of nonlinear equations (44) and the corresponding vector $\mathbf{f}^{CETE}(\boldsymbol{\pi}^{CETE*})$ is a CETE flow vector.

The proof is similar to that of Theorem 5.

3.6. The elimination of zero retrieved traffic demand

Similar to the case of solving NTE, zero retrieved traffic demand should be eliminated during the solution procedure. We redefine the function $\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})$ by $\tilde{D}_g^{CETE}(\pi_g^{CETE}, \boldsymbol{\pi}_{g-}^{CETE})$, where $\boldsymbol{\pi}_{g-}^{CETE} = [\pi_{g'}^{CETE}, g' \in G \setminus \{g\}]$. We have the following proposition:

Proposition 8: $\tilde{D}_g^{CETE}(\pi_g^{CETE}, \boldsymbol{\pi}_{g-}^{CETE})$ is monotonically increasing with respect to π_g^{CETE} .

The proof is similar to that of Proposition 7.

Algorithm 2 can be extended to eliminate zero retrieved traffic demand for the CETE problem by replacing $\tilde{\mathbf{D}}(\boldsymbol{\pi}) > \mathbf{0}$ with $\tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE}) > \mathbf{0}$ in Line 1 and replacing $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-}) = D_g$ with $\tilde{D}_g^{CETE}(\pi_g^{CETE}, \boldsymbol{\pi}_{g-}^{CETE}) = D_g$ in Line 3.

3.7. Solution algorithm: The BFGS-based method

Similar to Algorithm 3, unconstrained optimization problem (45) was solved by the BFGS-based method. Here, the details of the algorithm are omitted. By numerical tests, we found that the solutions to the CETE problem and the CTE problem (see Appendix E) were very close. Therefore, the initial solution to unconstrained optimization problem (45) was estimated by the solution to the CTE problem. The detailed derivation of the gradient of the objective function of problem (45) is provided in Appendix D.2.

4. The optimal toll schemes

4.1. The bi-level programming problems

The optimal CET scheme design problem can be modeled as a bi-level programming problem. The upper level problem is aimed at reducing TSC by optimizing the CET scheme, i.e., optimizing the vector $\boldsymbol{\alpha}^{CET}$. The lower level problem is the CETE problem, which is used to evaluate the performance of any given vector $\boldsymbol{\alpha}^{CET}$. The upper level problem can be formulated as the following optimization problem:

$$\begin{aligned} \min \psi^{CET}(\boldsymbol{\alpha}^{CET}) = & \sum_{g\in G}\sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{grs}^{CETE}(\ell)\left[\alpha_g \tau_r(f_{rs}^{CETE}(\ell)) + \beta_{gs}\ell\right]d\ell \\ & + \sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{rs}^{CETE}(\ell)\tilde{E}_r(\tau_r(f_{rs}^{CETE}(\ell)))d\ell, \end{aligned} \tag{46}$$

where $[f_{rs}^{CETE}(\ell), r \in R, s \in S, \ell \le 0]$ and $[f_{grs}^{CETE}(\ell), g \in G, r \in R, s \in S, \ell \le 0]$ are functions of $\boldsymbol{\alpha}^{CET}$ and can be determined by Eqs. (39) and (41) at the CETE, respectively. The two terms on the right hand side of Eq. (46) are TSTC and TSEC, respectively.

Similar to the optimal CET scheme design problem, the optimal CT scheme design problem can also be modeled as a bi-level programming problem. Differently, the upper level problem is aimed at reducing TSTC by optimizing the CT scheme, i.e., optimizing the vector $\boldsymbol{\alpha}^{CT} = [\alpha_r^{CT}, \forall r \in R]$, where α_r^{CT} is a parameter associated with the CT scheme for road *r*. The lower level problem is the CTE problem, which can be formulated as an NTE problem (see Appendix E for details) and can be solved either by the GSD or BFGS method.

4.2. Solution algorithm: the double BFGS method

The upper level problem (46) can be viewed as unconstrained nonlinear optimization problems and can be solved by the classical BFGS method with the use of approximated gradients. We used the average VOT of all commuters as the initial solution to optimization model (46). The gradient of the objective function in (46) was estimated by the central difference method. Let $\boldsymbol{\alpha}_{r-}^{CET} = [\alpha_{r'}^{CET}, r' \in R \setminus \{r\}]$. $\psi^{CET}(\boldsymbol{\alpha}^{CET})$ can then be redefined by $\psi^{CET}(\alpha_r^{CET}, \boldsymbol{\alpha}_{r-}^{CET})$, respectively. The estimated gradients can be expressed as follows:

$$\frac{\partial \psi^{CET}(\boldsymbol{\alpha}^{CET})}{\partial \alpha_r^{CET}} \approx \frac{\psi^{CET}(\alpha_r^{CET} + \Delta, \boldsymbol{\alpha}_{r-}^{CET}) - \psi^{CET}(\alpha_r^{CET} - \Delta, \boldsymbol{\alpha}_{r-}^{CET})}{2\Delta}, \forall r \in R, \tag{47}$$

where Δ is small but non-infinitesimal. The same as Algorithm 3, the Armijo Rule was used to determine the step size. $\left\| \nabla \psi^{CET}(\boldsymbol{\alpha}^{CET}) \right\|$ was adopted as the gap function for the optimal CET scheme design problem.

The objective function value of $\psi^{CET}(\boldsymbol{\alpha}^{CET})$ in (46) was obtained in two steps. Firstly, with $\boldsymbol{\alpha}^{CET}$ as an input, the BFGS-based method (see Sections 2.5. and 3.7 for details) was used to solve the corresponding CETE problem for its equilibrium flow rates. Secondly, $\psi^{CET}(\boldsymbol{\alpha}^{CET})$ was computed using the corresponding equilibrium flow rates obtained in the last step and the corresponding objective function in (46).

The solution algorithm involves two BFGS methods. One is for solving the upper level problem and forms the outer loop of the procedure. The other one is for solving the lower level problem and forms the inner loop in each iteration of the outer loop. To save space, the procedure of the solution algorithm is omitted here. Note that the proposed double BFGS method can be directly extended to solve the optimal CT scheme design problem.

5. Numerical examples

In this section, four examples are given to illustrate the properties of the proposed models and the performance of the proposed solution algorithms. All of the experiments were run on a computer with an Intel Core i5-2540 2.60-GHz CPU with 8.00 GB of RAM.

5.1. Setting of the emission cost function

For all pollutants $h \in H$, we adopted the following emission rate function modified from the study of Penic and Upchurch (1992):

$$\varphi_r^h(\tau) = \frac{\tau^2}{L_r} A_h \exp\left\{\frac{B_h L_r}{\tau}\right\}, \tag{48}$$

where A_h and B_h are constants with respect to pollutant *h*.

In this study, we consider three types of pollutants: nitrogen oxides (NOx), volatile organic compounds (VOC), and carbon monoxide (CO). The parameters of the emission rate function (48) and the emission cost function (25) are given in Table 1. The setting of the parameters of the emission rate function directly follows the study of Penic and Upchurch (1992) and the setting of the parameter λ_h was modified from the study of Mayeres et al. (1996).

Table 1. Parameters of emission rate functions and monetary valuation of each type of pollutant.

Type of pollutant *h*	NO_x	VOC	CO
A_h	0.5658	1.0023	12.2267
B_h	0.0371	0.0137	0.0133
λ_h	0.49404	0.10560	0.00036

5.2. Arrival order of commuter groups and the effect of the number of interpolants on system performance

We consider an example with a single road and two user classes to illustrate the effect of the number of interpolants on system performance and verify that commuters do not have a fixed arrival order if the CET scheme is adopted. The length and the free-flow travel time of the road are 15 km and 0.25 h, respectively. The two parameters for the power speed-flow function (4) are ζ_1=1.0, Q_1 = 3817 veh/h, and $\kappa = 4.08$. The desired arrival time t^* = 8 am. The input parameter vectors for the two user classes are as follows: $\boldsymbol{\alpha}$ = [3.9, 4.1] ($), $\boldsymbol{\beta}$ = [2.2, 2.24, 10, 10.2] ($), $\boldsymbol{\alpha}^{CT} = \boldsymbol{\alpha}^{CET}$ = 4 ($), and $\mathbf{D}$ = [250, 1770]. The parameter of the GSD method is $\varepsilon_1 = 1.0\times10^{-12}$, except that it took a value of 100.0 when the GSD method was used to generate the initial solutions to the BFGS-based method for solving the NTE and CTE problems. The parameters of the BFGS-based algorithm for solving the NTE and CTE problems are as follows: $\varepsilon_2 = 1.0\times10^{-8}$, $\sigma = 0.4$, and $\rho = 0.1$.

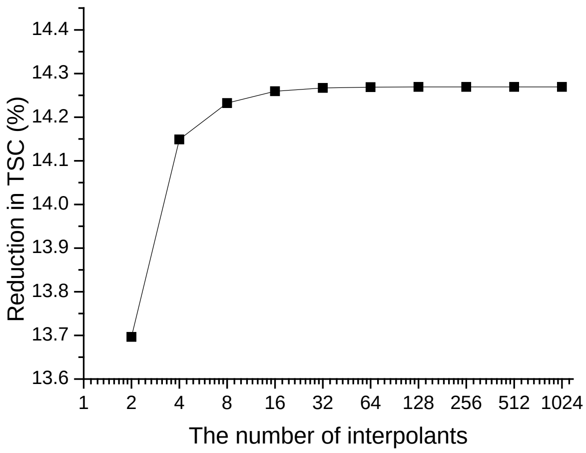

Fig. 4. The effect of the number of interpolants on system performance.

We set different numbers of linear interpolants, solved the optimal CET problem, and obtained the reduction in TSC as revealed in Fig. 4. It is observed that the reduction in TSC is monotonically increasing as the number of interpolants increases. This is because when the more interpolants are adopted, the approximation of the emission cost function is more accurate, and the obtained CET scheme can achieve better performance. We can also observe that the reduction in TSC is almost unchanged when the number of interpolants is larger than 16. In the following experiments, we fixed the number of interpolants to be 20.

We solved the NTE and CTE problems by both the GSD method and the BFGS-based method, and solved the CETE problems by the BFGS-based method. The obtained equilibrium flows and the equilibrium arrival flow rates are provided in Table 2 and Fig. 5, respectively. We found that both the GSD method and the BFGS-based method could obtain the same equilibrium flows and the same equilibrium arrival flow rates for the NTE and CTE problems.

Table 2. NTE, CTE, and CETE flows.

	NTE		CTE		CETE	
Class	Early	Late	Early	Late	Early	Late
1	318.69	211.31	144.11	385.89	136.23	393.77
2	1341.45	388.55	1509.96	220.04	1517.73	212.27

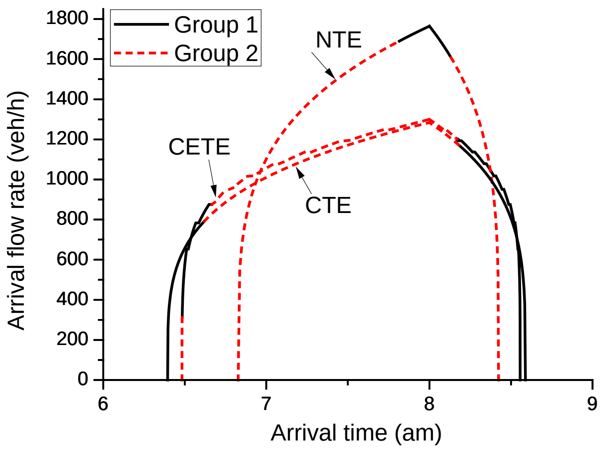

Fig. 5. The equilibrium arrival flow rate of each group.

At the NTE, the arrival order is flow-independent and determined by the relative cost β_{gs} / α_g for the s-th type schedule delay. As shown in Fig. 5, the commuters of group 1 prefer to travel closer to the peak at the NTE. This is because group 1 has a higher relative cost of schedule delay to travel time than group 2. However, the commuters of group 2 prefer to travel closer to the peak at the CTE. This is because that the arrival order is determined by the relative cost $\beta_g / (\alpha_g + \alpha_1^{CT}\kappa)$ and group 2 has a higher relative cost than group 1. Different from the commuters at the NTE and the CTE, the commuters at the CETE do not have a predetermined arrival order. We can observe from Fig. 5 that some commuters of group 2 commuters arrive earlier than the early arrival commuters of group 1 but other early arrival commuters of group 2 arrive later than them. This is because the relative cost of schedule delay to travel time (i.e., $\beta_{gs} / \alpha_{g1,k}$) is flow-dependent.

5.3. A comparison of system performance under different toll schemes

We constructed an example with a single road and four user classes to compare system performance under different toll schemes. The input parameter vectors for the four classes are the same as those in the study of Liu et al. (2015): $\boldsymbol{\alpha} = [6.4, 2.5, 2.0, 1.7]$, $\boldsymbol{\beta} = [3.9, 1.95, 1.8, 1.5, 15.21, 4.5, 3.5, 5]$, and $\mathbf{D} = [1500, 1140, 800, 500]$. The parameters of the road and the BFGS-based method for the lower level problems are the same as those in Section 5.2. The parameters of the BFGS-based algorithm for solving the upper level problem of both optimal CT and CET scheme design problems are as follows: $\varepsilon_2 = 0.01$, $\sigma = 0.4$, and $\rho = 0.4$.

We solved the NTE problem and the optimal toll scheme problems, and got the optimal value of α_1^{CT} and system performance as shown in Table 3. We can observe that the optimal CET scheme outperforms the optimal CT scheme in terms of both TSTC and TSEC. It is not surprising that the optimal CET scheme is better than the optimal CT scheme in terms of TSEC as the former does consider TSEC while the latter does not. The reason for the optimal CET scheme outperforming the optimal CT scheme in terms of TSTC is that the proposed optimal CT scheme is anonymous and cannot achieve system optimum in terms of TSTC.

We obtained the equilibrium individual generalized trip costs as revealed in Table 4. We can observe that only the commuters of group 1 benefit from the optimal CT scheme—after implementing the optimal CT scheme, the commuters of group 1 have a lower equilibrium individual generalized trip cost but the commuters of other groups have a higher equilibrium individual generalized trip cost. This implies that the implementation of the optimal CT scheme may raise the fairness issue. Different from the optimal CT scheme, the optimal CET scheme can lead to a larger equilibrium individual generalized trip cost for all commuters.

Table 3. The performance of the proposed toll schemes.

Performance	NT	CT	CET
α_1^{CT} and α_1^{CET}	0	4.66	4.47
TSTC	18852.78	13987.85	13954.35
TSEC	2471.57	2226.91	2179.78
TSC	21324.35	16214.76	16134.13
Reduction in TSC (%)	0	23.96	24.34

Table 4. Equilibrium generalized trip costs under the optimal toll schemes.

Group	NT	CT	CET
1	8.45	**8.23**	8.75
2	3.67	5.19	5.66
3	2.92	4.70	5.17
4	2.56	4.23	4.71

5.4. The applicability of the proposed method

We constructed an example based on empirical data on a segment of California State Route 91 (SR-91) to demonstrate the applicability of the proposed method as a numerical analysis tool for real-world applications. This specific segment is 33 miles long and roughly represents the median commute for those living in Corona who use SR-91 (Hall, 2019). The distributions of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{D}$ are the same as those in the study of Liu et al. (2015): α_g follows log-normal distribution with a mean of \$21 and a variance of \$110, β_{g1}/α_g is distributed uniformly over [0, 0.23] with a point mass at 0.23, β_{g1}/β_{g2} equals 0.4 for all users, and the demand was discretized into 110 groups according to the cumulative density function of α_g and β_{g1}/α_g. The free-flow travel time of the road is 37 min, the capacity of the road is Q_1=1500 veh/h, users' desired arrival time t^* is 6:50 am, the total demand is 11250, and we set $\kappa = 4.08$.

Liu et al. (2015) used the road detector data from the California Department of Transportation's Performance Measurement System (PeMS) (California Department of Transportation, 1999) to estimate the actual travel times in the general-purpose lanes for each arrival time for every non-holiday weekday in 2004 for trips on SR-91. Every five-minute interval between 4:00 am and 10:00 am was used. As shown in Fig. 6, we found that the NTE travel times can well fit the actual travel times when ζ_1 = 4.07. With this parameter, the average gap between the NTE and actual travel times is 0.73 min. This result implies that the proposed method can be used as a numerical analysis tool for practical applications.

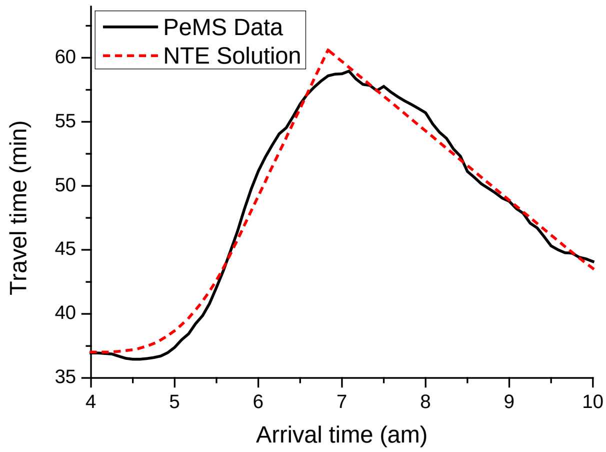

Fig. 6. NTE travel times vs. actual travel times.

5.5. The effectiveness of the proposed solution algorithms

We conducted an experiment using a network with three parallel roads and multiple user classes to illustrate the effectiveness of the proposed solution algorithms. The parameters of the three parallel roads are provided in Table 5. The number of classes *n* = 50. The parameters of each group are randomly generated: As Chen et al. (2015), we assume that (1) the unit cost of travel time (α_g) is uniformly distributed from 5 to 35 \$/h; (2) β_{g1}/α_g is uniformly distributed from 0.05 to 0.95, independent of α_g; (3) β_{g2}/α_g is uniformly distributed from 1.05 to 3.95, independent of α_g and β_{g1}/α_g; and (4) the number of travelers for each class is an independent random variable following a Normal Distribution with a mean of 3000/*n* and a standard deviation of 3000/*n*. The sampling method directly follows that of Chen et al. (2015). The parameters of the proposed solution algorithms are the same as those in Sections 5.2 and 5.3.

Table 5. The parameters of the three parallel roads.

Road	Capacity (veh/h)	Length (km)	Free flow travel time (h)
1	3817	15	0.25
2	2000	9	0.15
3	3000	12	0.2

We set all the elements in the vector $\boldsymbol{\alpha}^{CT}$ and $\boldsymbol{\alpha}^{CET}$ to be the average unit cost of the travel time of all commuters and then solved the NTE, CTE, and CETE problems. In the NTE problem, for any given vector of route flows $\mathbf{N}$ obtained by the GSD method, we can retrieve the vector of route costs $\mathbf{C}(\mathbf{N})$ (see Appendix A for detials). This vector can be further used to retrieve the average generalized trip cost of each group, denoted by $\bar{\pi}_g(\mathbf{N})$. Let $\bar{\boldsymbol{\pi}}(\mathbf{N})=[\bar{\pi}_g(\mathbf{N}), \forall g \in G]$. To fairly compare the efficiency between the proposed BFGS method and the GSD method, $\left\|\mathbf{Z}(\bar{\boldsymbol{\pi}}(\mathbf{N}))\right\|$ and $\left\|\mathbf{Z}^{CTE}(\bar{\boldsymbol{\pi}}^{CTE}(\mathbf{N}))\right\|$ were used as the gap functions for the GSD method for solving the NTE and CTE problems, respectively, where $\left\|\mathbf{Z}^{CTE}(\bar{\boldsymbol{\pi}}^{CTE}(\mathbf{N}))\right\|$ is similarly defined as $\left\|\mathbf{Z}(\bar{\boldsymbol{\pi}}(\mathbf{N}))\right\|$ for the CTE problem. The gap values are presented in Fig. 7. We can observe that the gap values decrease as the CPU time increases. The BFGS-based method is more efficient to solve the NTE and CTE problems than the proposed GSD method. Compared with the NTE problem, the CTE problem can be solved more quickly by the BFGS-based method, but the CETE problem can be solved more slowly by that method.

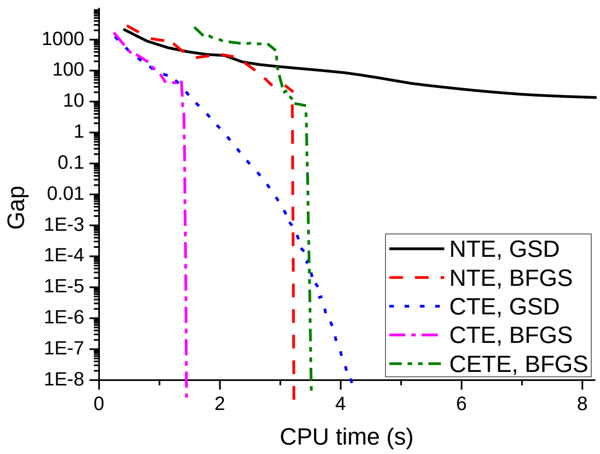

Fig. 7. The convergence of the solution algorithms for solving the equilibrium problems.

We used the double BFGS method to solve the CT and CET scheme design problems for the optimal toll schemes. The values of the gap functions $\left\|\nabla\psi^{CT}(\boldsymbol{\alpha}^{CT})\right\|$ and $\left\|\nabla\psi^{CET}(\boldsymbol{\alpha}^{CET})\right\|$ obtained during the implementation of the double BFGS method are presented in Fig. 8. We can observe that the values of the gap functions quickly decrease as the number of iterations increases. The values of the gap functions decreased to 0.01 after 5 and 6 iterations for the CT and the CET scheme design problems, respectively.

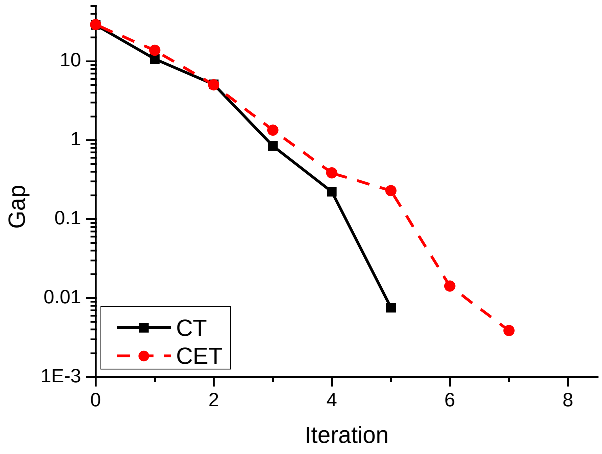

Fig. 8. The convergence of the double BFGS method for solving the optimal toll scheme problems.

The optimal vectors of $\boldsymbol{\alpha}^{CT}$ and $\boldsymbol{\alpha}^{CET}$ and the performance of the proposed toll schemes are provided in Table 6. We can observe that the optimal vector of $\boldsymbol{\alpha}^{CT}$ for the optimal CT scheme is slightly different from that of $\boldsymbol{\alpha}^{CET}$ for the optimal CET scheme, and the values of α_r^{CT} and α_r^{CET} are also different for each road. This implies that the toll charges should be different on various roads under various schemes. We can also observe from Table 6 that the optimal CET scheme outperforms the optimal CT scheme in terms of both TSTC and TSEC. This result is consistent with that in Section 5.3.

Table 6. The performance of the proposed toll schemes.

Performance	NT	CT	CET
$\boldsymbol{\alpha}^{CT}$ and $\boldsymbol{\alpha}^{CET}$	-	[21.30, 31.35, 26.28]	[21.14, 31.00, 26.12]
TSTC ($)	41936.66	33058.68	33057.17
TSEC ($)	1379.77	1817.84	1796.97
TSC ($)	43316.43	34876.52	34854.14
Reduction in TSC (%)	0	19.48	19.54

6. Conclusions

In this study, the Henderson approach was adopted to model road congestion and examine commuter's arrival time and route choice in a single OD network with heterogeneous users and parallel routes at NTE, CTE, and CETE. The last two equilibria assume that anonymous second-best CT and CET schemes are implemented. These schemes were designed based on the concept of marginal-cost pricing.

At CETE, the arrival of each group at the destination in the whole studied period does not follow a predetermined order, and the CETE problem cannot be formulated and solved using the traditional approach i.e., reformulating the problem as a static traffic assignment problem using a VI and solving it by the GSD method. Taking equilibrium generalized trip costs as decision variables, we formulated the CETE problem as an unconstrained optimization problem and solved it using the BFGS method. This approach was also used to solve NTE and CTE problems.

Bi-level programming models were proposed to formulate the optimal CT and CET design problems. The proposed models were solved by the double BFGS method. Finally, numerical examples are provided to illustrate the properties of all proposed models and the efficiency of the proposed solution algorithms. The results show that the BFGS-based method outperforms the GSD method in terms of solving the NTE and CTE problems. The optimal CET scheme outperforms the optimal CT scheme in terms of TSC.

In this paper, it is assumed that the emission cost function is convex and the rate of emissions depends on average vehicle speed. However, if vehicle speed varies appreciably, the estimated rate of emissions may not be accurate. Moreover, emission rates of some pollutants spike sharply during the acceleration phase of the driving cycle. These features cannot be captured by our models, which are formulated by the Henderson approach. In the future, we shall consider other traffic flow models that can capture these features. In this paper, the travel demand is fixed, all travelers have the same desired arrival time, and vehicles are homogeneous in terms of their passenger car equivalents and emissions characteristics. This assumption or simplification makes the analysis in the paper much less tedious. However, the modeling framework in the paper is still applicable if we consider elastic demand, different desired arrival times, and multi-class vehicles. Therefore, elastic demand (e.g., Arnott et al., 1993; Harks et al., 2018), heterogeneous desired arrival times (e.g., Henderson, 1981; Zhu et al., 2019), and heterogeneous vehicle types (e.g., Tian et al., 2013; van den Berg and Verhoef, 2016) can be considered in future research.

Acknowledgments

This work is jointly supported by the National Key R&D Program of China (2018YFB1600903), the National Natural Science Foundation of China (71431003, 71871077), a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (HKU 17201915), and the Fundamental Research Funds for the Central Universities (PA2019GDQT0020). The authors are grateful to the reviewers for their constructive comments, and to Dr. Yang Liu for the discussion on the example of California State Route 91.

Appendix A: Retrieving route costs and arrival flow rates from route flows in the NTE problem

By definition, we have

$$\int_{\ell_{g_s(i)rs}}^{\ell_{g_s(i+1)rs}} f_{g_s(i)rs}(\ell)d\ell = N_{g_s(i)rs}, \forall i \in I, r \in R, s \in S. \tag{A.1}$$

Eq. (A.1) implies that if $N_{g_s(i)rs} = 0$, then $\ell_{g_s(i+1)rs} = \ell_{g_s(i)rs}$ is satisfied; otherwise, $\ell_{g_s(i+1)rs} > \ell_{g_s(i)rs}$. All commuters of group $g_s(i)$ have the same generalized trip cost. Therefore, we have

$$C_{g_s(i)rs}(\ell) = C_{g_s(i)rs}, \forall i \in I, r \in R, s \in S, \ell \in [\ell_{g_s(i)rs}, \ell_{g_s(i+1)rs}]. \tag{A.2}$$

Substituting Eq. (A.2) into Eq. (6) and rearranging the resultant equation, we have

$$\tau_{rs}(\ell)=\frac{C_{g_s(i)rs}-c_r^0+\beta_{g_s(i)s}\ell}{\alpha_{g_s(i)}},\forall i\in I,r\in R,s\in S,\ell\in[\ell_{g_s(i)rs},\ell_{g_s(i+1)rs}]. \quad \text{(A.3)}$$

Only the commuters of group $g_s(i)$ arrive at the destination during the interval $[\ell_{g_s(i)rs},\ell_{g_s(i+1)rs}]$. Therefore, substituting Eq. (A.3) into Eq. (2), we can obtain the arrival flow rate of group $g_s(i)$:

$$f_{g_s(i)rs}(\ell)=\zeta_r Q_r\left(\frac{C_{g_s(i)rs}-c_r^0+\beta_{g_s(i)s}\ell-\alpha_{g_s(i)}\tau_r^0}{\alpha_{g_s(i)}L_r}\right)^{\frac{1}{\kappa}},\forall i\in I,r\in R,s\in S,\ell\in[\ell_{g_s(i)rs},\ell_{g_s(i+1)rs}]. \quad \text{(A.4)}$$

Substituting Eq. (A.4) into Eq. (A.1) and rearranging the resultant equation, we have

$$\left[\tau_{rs}(\ell_{g_s(i+1)rs})-\tau_r^0\right]^{\frac{\kappa+1}{\kappa}}-\left[\tau_{rs}(\ell_{g_s(i)rs})-\tau_r^0\right]^{\frac{\kappa+1}{\kappa}}=\frac{\beta_{g_s(i)s}N_{g_s(i)rs}(\kappa+1)L_r^{1/\kappa}}{\zeta_r Q_r\alpha_{g_s(i)}\kappa},\forall i\in I,r\in R,s\in S. \quad \text{(A.5)}$$

The commuters with the smallest relative arrival time (i.e., the first and the last commuters) experience free flow travel time (Chu, 1995; Coria and Zhang, 2017). Therefore, we have $\tau_{rs}(\ell_{g_s(1)rs})=\tau_r^0$ for all $r\in R$ and $s\in S$. Substituting this equation into the resultant expression obtained after taking summation on both sides of Eq. (A.5), we have

$$\left[\tau_{rs}(\ell_{g_s(i+1)rs})-\tau_r^0\right]^{\frac{\kappa+1}{\kappa}}=\sum_{j=1}^{i}\frac{\beta_{g_s(j)s}N_{g_s(j)rs}(\kappa+1)L_r^{1/\kappa}}{\alpha_{g_s(j)}\zeta_r Q_r\kappa},\forall i\in I,r\in R,s\in S. \quad \text{(A.6)}$$

Rearranging Eq. (A.6), we get

$$\tau_{rs}(\ell_{g_s(i+1)rs})=\tau_r^0+L_r\left(\sum_{j=1}^{i}\frac{\beta_{g_s(j)s}N_{g_s(j)rs}(\kappa+1)}{\alpha_{g_s(j)}\zeta_r Q_r L_r\kappa}\right)^{\frac{\kappa}{\kappa+1}},\forall i\in I,r\in R,s\in S. \quad \text{(A.7)}$$

Based on Eq. (A.3), we obtain

$$\tau_{rs}(\ell_{g_s(i)rs})=\frac{C_{g_s(i)rs}-c_r^0+\beta_{g_s(i)s}\ell_{g_s(i)rs}}{\alpha_{g_s(i)}},\forall i\in I,r\in R,s\in S\ \text{ and} \quad \text{(A.8)}$$

$$\tau_{rs}(\ell_{g_s(i+1)rs})=\frac{C_{g_s(i)rs}-c_r^0+\beta_{g_s(i)s}\ell_{g_s(i+1)rs}}{\alpha_{g_s(i)}},\forall i\in I,r\in R,s\in S. \quad \text{(A.9)}$$

Using Eqs. (A.7)-(A.9), we have

$$\ell_{g_s(i+1)rs}-\ell_{g_s(i)rs}=\frac{L_r\alpha_{g_s(i)}}{\beta_{g_s(i)s}}(\omega_{r,i+1}^s-\omega_{r,i}^s),\forall i\in I,r\in R,s\in S. \quad \text{(A.10)}$$

where

$$\omega_{r,i}^s=\left(\sum_{j'=1}^{i-1}\frac{\beta_{g_s(j')s}N_{g_s(j')rs}(\kappa+1)}{\zeta_r Q_r\alpha_{g_s(j')}L_r\kappa}\right)^{\frac{\kappa}{\kappa+1}}. \quad \text{(A.11)}$$

By definition, we have $\ell_{g_s(n+1)rs} = 0$. Substituting $\ell_{g_s(n+1)rs} = 0$ into Eq. (A.10) and taking summation on both sides of the resultant system of linear equations, we have

$$\ell_{g_s(i)rs} = -L_r \sum_{j=i}^{n} \frac{\alpha_{g_s(j)}}{\beta_{g_s(j)s}} (\omega_{r,j+1}^s - \omega_{r,j}^s), \forall i \in I, r \in R, s \in S. \quad (A.12)$$

Substituting Eqs. (A.7) and (A.12) into Eq. (A.8), and rearranging the resultant equation, we have

$$C_{g_s(i)rs} = c_r^0 + \alpha_{g_s(i)}\tau_r^0 + L_r \alpha_{g_s(i)} \omega_{i+1}^s + \beta_{g_s(i)s} L_r \sum_{j=i+1}^{n} \frac{\alpha_{g_s(j)}}{\beta_{g_s(j)s}} (\omega_{r,j+1}^s - \omega_{r,j}^s), \forall i \in I, r \in R, s \in S. \quad (A.13)$$

Using Eq. (A.12), we can retrieve the smallest relative arrival time of commuters of each group, i.e., $[\ell_{grs}, g \in G, r \in R, s \in S]$ from the vector of route flows $\mathbf{N} = [N_{grs}, g \in G, r \in R, s \in S]$. Using Eq. (A.13), we can retrieve route costs from route flows, i.e., $[C_{grs}, g \in G, r \in R, s \in S]$ from the vector of route flows $\mathbf{N}$. Using Eqs. (2), (A.3), (A.12)-(A.13), we can further retrieve the arrival flow rates of each group from $\mathbf{N}$ as follows:

$$f_{g_s(i)rs}(\ell) = \begin{cases} \zeta_r Q_r \left(\dfrac{C_{g_s(i)rs} - c_r^0 - \alpha_{g_s(i)}\tau_r^0 + \beta_{g_s(i)s}\ell}{\alpha_{g_s(i)} L_r} \right)^{\frac{1}{\kappa}}, & \text{if } \ell \in [\ell_{g_s(i)rs}, \ell_{g_s(i+1)rs}], \forall i \in I, r \in R, s \in S, \\ 0, & \text{otherwise,} \end{cases}$$

(A.14)

where $C_{g_s(i)rs}$ and $\omega_{r,j}^s$ are defined by Eqs. (A.13) and (A.11), respectively.

Appendix B: The proofs of propositions and theorems

B.1. The proof of Theorem 1

Proof. To simplify the notation, we use the superscript '*' for variables associated with $\mathbf{N}^*$. By the definition of $\mathbf{f}(\mathbf{N}^*)$, we have $\mathbf{f}(\mathbf{N}^*) \geq \mathbf{0}$ and

$$f_{g_s(i)rs}^*(\ell) = \begin{cases} \zeta_r Q_r \left(\dfrac{C_{g_s(i)rs}^* - c_r^0 - \alpha_{g_s(i)}\tau_r^0 + \beta_{g_s(i)s}\ell}{\alpha_{g_s(i)} L_r} \right)^{\frac{1}{\kappa}}, & \text{if } \ell \in [\ell_{g_s(i)rs}^*, \ell_{g_s(i+1)rs}^*], \forall i \in I, r \in R, s \in S. \\ 0, & \text{otherwise,} \end{cases} \quad (B.1)$$

Substituting Eq. (A.3) into Eq. (6), we have

$$C_{g_s(i)rs}^*(\ell) = c_r^0 + \alpha_{g_s(i)s} \frac{C_{g_s(j)rs}^* + \beta_{g_s(j)s}\ell}{\alpha_{g_s(j)}} - \beta_{g_s(i)s}\ell, \ \forall \ell \in [\ell_{g_s(j)rs}^*, \ell_{g_s(j+1)rs}^*]. \quad (B.2)$$

Differentiating Eq. (B.2) with respect to ℓ, we have

$$\frac{dC_{g_s(i)rs}^*(\ell)}{d\ell} = \frac{\alpha_{g_s(i)}\beta_{g_s(j)s}}{\alpha_{g_s(j)}} - \beta_{g_s(i)s}, \ \forall \ell \in [\ell_{g_s(j)rs}^*, \ell_{g_s(j+1)rs}^*]. \quad (B.3)$$

The arrival order of all the groups is in ascending order with respect to β_{gs}/α_g. Therefore, we have $\beta_{g_s(i)s}/\alpha_{g_s(i)} > \beta_{g_s(j)s}/\alpha_{g_s(j)}$ if $i > j$ and $\beta_{g_s(i)s}/\alpha_{g_s(i)} < \beta_{g_s(j)s}/\alpha_{g_s(j)}$ if $i < j$. Based on Eq. (B.3), we have

$$\frac{dC^*_{g_s(i)rs}(\ell)}{d\ell} < 0,\ \forall j < i, \ell \in [\ell^*_{g_s(j)rs}, \ell^*_{g_s(j+1)rs}] \text{ and} \tag{B.4}$$

$$\frac{dC^*_{g_s(i)rs}(\ell)}{d\ell} > 0,\ \forall j > i, \ell \in [\ell^*_{g_s(j)rs}, \ell^*_{g_s(j+1)rs}]. \tag{B.5}$$

Based on Eq. (A.2) and inequalities (B.4) and (B.5), we have

$$\frac{dC^*_{g_s(i)rs}(\ell)}{d\ell}\begin{cases} <0, & \text{if } \ell < \ell^*_{g_s(i)rs}, \\ =0, & \text{if } \ell^*_{g_s(i)rs} \le \ell \le \ell^*_{g_s(i+1)rs}, \forall i \in I, r \in R, s \in S, \ell \le 0. \\ >0, & \text{if } \ell > \ell^*_{g_s(i+1)rs}, \end{cases} \tag{B.6}$$

Based on Eq. (A.2) and inequality (B.6), we have

$$C^*_{g_s(i)rs}(\ell)\begin{cases} = C^*_{g_s(i)rs}, & \text{if } \ell^*_{g_s(i)rs} \le \ell \le \ell^*_{g_s(i+1)rs}, \\ > C^*_{g_s(i)rs}, & \text{otherwise,} \end{cases} \forall i \in I, r \in R, s \in S, \ell \le 0. \tag{B.7}$$

By definition, $\mathbf{N}^*$ is a solution to VI problem (10). Following Friesz et al. (1993), we can prove that the following condition is satisfied:

$$C^*_{grs}\begin{cases} = \pi^*_g, & \text{if } N^*_{grs} > 0, \\ \ge \pi^*_g, & \text{if } N^*_{grs} = 0, \end{cases} \forall g \in G, r \in R, s \in S. \tag{B.8}$$

By definition, if $N^*_{g_s(i)rs} = 0$, we have $\ell^*_{g_s(i)rs} = \ell^*_{g_s(i+1)rs}$ and $f^*_{grs}(\ell) = 0$ is satisfied for all $\ell \in (-\infty, 0]$. If $N^*_{g_s(i)rs} > 0$, we have $\ell^*_{g_s(i)rs} < \ell^*_{g_s(i+1)rs}$, $f^*_{grs}(\ell) > 0$ is satisfied for all $\ell \in (\ell^*_{g_s(i)rs}, \ell^*_{g_s(i+1)rs})$, and $f^*_{grs}(\ell) = 0$ is satisfied for all $\ell \in (-\infty, \ell^*_{g_s(i)rs}) \bigcup (\ell^*_{g_s(i+1)rs}, 0]$. Hence, based on conditions (B.7) and (B.8), we have

$$C^*_{grs}(\ell)\begin{cases} = \pi^*_g, & \text{if } f^*_{grs}(\ell) > 0, \\ \ge \pi^*_g, & \text{if } f^*_{grs}(\ell) = 0, \end{cases} \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{B.9}$$

Based on Eq. (A.1), we have

$$\int_{\ell^*_{g_s(i)rs}}^{\ell^*_{g_s(i+1)rs}} f^*_{g_s(i)rs}(\ell)d\ell = N^*_{g_s(i)rs}, \forall i \in I, r \in R, s \in S. \tag{B.10}$$

With $\mathbf{N}^* \in \Omega$, we have

$$\sum_{r \in R}\sum_{s \in S} N^*_{grs} = D_g, \forall g \in G. \tag{B.11}$$

Based on Eqs. (B.1), (B.10), and (B.11), we have

$$\sum_{r \in R}\sum_{s \in S}\int_{-\infty}^{0} f^*_{grs}(\ell)d\ell = \sum_{r \in R}\sum_{s \in S} N^*_{grs} = D_g, \forall g \in G. \tag{B.12}$$

Eqs. (B.1), (B.9), and (B.12) imply that $\mathbf{f}(\mathbf{N}^*)$ satisfies conditions (7)-(9), and hence $\mathbf{f}(\mathbf{N}^*)$ is an NTE flow vector. This completes the proof. □

B.2. The proof of Proposition 2

Proof. Taking derivatives of Eq. (A.11), we have

$$\frac{\partial \omega_{r,j}^{s}}{\partial N_{g_s(i)r's'}} = \begin{cases} \dfrac{\beta_{g_s(i)s} N_{g_s(i)rs}}{\zeta_r Q_r \alpha_{g_s(i)} L_r} \left(\displaystyle\sum_{j'=1}^{j-1} \frac{\beta_{g_s(j')s} N_{g_s(j')rs} (\kappa+1)}{\zeta_r Q_r \alpha_{g_s(j')} L_r \kappa} \right)^{-\frac{1}{\kappa+1}}, & \text{if } j > i,\ r = r' \text{ and } s = s', \\ 0, & \text{otherwise.} \end{cases} \tag{B.13}$$

Eq. (B.13) implies that $\partial \omega_{r,j}^{s} / \partial N_{g_s(i)r's'} \geq 0$ and $\partial \omega_{r,j+1}^{s} / \partial N_{g_s(i)r's'} - \partial \omega_{r,j}^{s} / \partial N_{g_s(i)r's'} \geq 0$. Based on Eqs. (A.13) and (B.13), we have

$$\frac{\partial C_{g_s(i)rs}}{\partial N_{g_s(i)r's'}} = \begin{cases} L_r \alpha_{g_s(i)} \dfrac{\partial \omega_{r,i+1}^{s}}{\partial C_{g_s(i)rs}} + \beta_{g_s(i)s} L_r \displaystyle\sum_{j=i+1}^{n} \frac{\alpha_{g_s(j)}}{\beta_{g_s(j)s}} \left(\frac{\partial \omega_{r,j+1}^{s}}{\partial C_{g_s(i)rs}} - \frac{\partial \omega_{r,j}^{s}}{\partial C_{g_s(i)rs}} \right), & \text{if } r = r' \text{ and } s = s', \\ 0, & \text{otherwise.} \end{cases} \tag{B.14}$$

Based on Eq. (B.14), we have

$$\frac{\partial C_{grs}}{\partial N_{gr's'}} = \begin{cases} \geq 0, & \text{if } r = r' \text{ and } s = s', \\ 0, & \text{otherwise.} \end{cases} \tag{B.15}$$

Eq. (B.15) implies that the Jacobian of the function $\mathbf{C}_g(N_g, \mathbf{N}_{g-})$ is a non-negative diagonal matrix, and hence VI problem (12) can be equivalently formulated as optimization problem (13) (see Theorem 1.1 in Nagurney (1999)). This completes the proof. □

B.3. The proof of Theorem 2

Proof. Substituting Eq. (5) into Eq. (6), we have

$$C_{grs}(\ell) = c_r^0 + \alpha_g \tau_r(f_{rs}(\ell)) - \beta_{gs}\ell,\ \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{B.16}$$

We define the following function:

$$\bar{C}_{grs}(\ell) = c_r^0 + \alpha_g \tau_r(\bar{f}_{grs}(\ell)) - \beta_{gs}\ell,\ \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{B.17}$$

Substituting Eq. (18) into Eq. (B.17), we can conclude that

$$\bar{C}_{grs}(\ell) \begin{cases} = \pi_g, & \text{if } \ell > \bar{\ell}_{grs}, \\ = c_r^0 + \alpha_g \tau_r^0 - \beta_{gs}\ell \geq \pi_g, & \text{if } \ell \leq \bar{\ell}_{grs}, \end{cases} \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{B.18}$$

Eq. (4) implies that $\tau_r(f)$ is strictly monotone. Hence, based on Eq. (18) and condition (B.18), we have

$$\bar{C}_{grs}(\ell) \begin{cases} = \pi_g, & \text{if } \bar{f}_{grs}(\ell) > 0, \\ \geq \pi_g, & \text{if } \bar{f}_{grs}(\ell) = 0, \end{cases} \forall g \in G, r \in R, s \in S, \ell \leq 0. \tag{B.19}$$

Based on Eqs. (19), (B.16), and (B.17), we have

$$C_{grs}(\ell)\begin{cases} =\bar{C}_{grs}(\ell), & \text{if } f_{rs}(\ell)=\bar{f}_{grs}(\ell), \\ \geq \bar{C}_{grs}(\ell), & \text{otherwise,} \end{cases} \forall g\in G, r\in R, s\in S, \ell\leq 0. \tag{B.20}$$

Combining conditions (B.19) and (B.20), we have

$$C_{grs}(\ell)\begin{cases} =\pi_g, & \text{if } f_{grs}(\ell)>0, \\ \geq \pi_g, & \text{if } f_{grs}(\ell)=0, \end{cases} \forall g\in G, r\in R, s\in S, \ell\leq 0, \tag{B.21}$$

which implies that the arrival flow pattern defined by Eq. (20) satisfies NTE condition (9). This completes the proof. □

B.4. The proof of Theorem 3

Proof. To simplify the notation, we use the superscript '*' for variables associated with $\boldsymbol{\pi}^*$. As $\boldsymbol{\pi}^*$ is a solution to the system of nonlinear equation (23), it must satisfy Eqs. (22) and (23). Hence, we have

$$\sum_{r\in R}\sum_{s\in S}\int_{-\infty}^{0} f_{gps}^*(t)dt=\tilde{D}_g(\boldsymbol{\pi}^*)=D_g, \forall g\in G. \tag{B.22}$$

This means that condition (8) is satisfied.

Based on Eqs. (18)-(20), we have $\tilde{\mathbf{f}}(\boldsymbol{\pi}^*)\geq \mathbf{0}$, which means that condition (7) is satisfied. According to Theorem 2, the flow pattern expressed as Eq. (20) satisfies condition (9).

To sum up, $\tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$ satisfies conditions (7)-(9) and hence $\tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$ is an NTE flow vector. This completes the proof. □

B.5. The proof of Theorem 4

Proof. By definition, $\mathbf{f}^*$ satisfy conditions (7)-(9). Condition (9) implies that, for any group $g\in G$, its positive flow rate (i.e., $f_{grs}^*(\ell)>0$) always stay on its own equilibrium arrival flow rate curve. We assume that there exists a group $\bar{g}\in G$, a road $\bar{r}\in R$, a type of schedule delay $\bar{s}\in S$, and an interval $(\vec{\ell},\overleftarrow{\ell})$ such that $f_{\bar{g}\bar{r}\bar{s}}^*(\ell)>0$ and $\tau_{\bar{r}\bar{s}}^*(\ell)=\bar{\tau}_{\bar{g}\bar{r}\bar{s}}^*(\ell)$ is satisfied for all $\ell\in(\vec{\ell},\overleftarrow{\ell})$ and group $\bar{g}$ does not have the highest equilibrium travel time (or equivalently the highest arrival flow rate) during this interval, where $\vec{\ell}<\overleftarrow{\ell}$.

As $\mathbf{f}^*$ is an NTE flow vector, we have $f_{g\bar{r}\bar{s}}^*(\ell)=0$ and $C_{g\bar{r}\bar{s}}^*(\ell)\geq\pi_g^*$ for all $g\in G\setminus\{\bar{g}\}$ and $\ell\in(\vec{\ell},\overleftarrow{\ell})$. Let group $\tilde{g}\in G$ be the group with the highest equilibrium travel time during the interval $(\vec{\ell},\overleftarrow{\ell})$. Equivalently, we have $\bar{\tau}_{\tilde{g}\bar{r}\bar{s}}^*(\ell)>\bar{\tau}_{\bar{g}\bar{r}\bar{s}}^*(\ell)=\tau_{\bar{r}\bar{s}}^*(\ell)$ for all $\ell\in(\vec{\ell},\overleftarrow{\ell})$. Substituting $\tau_{\bar{r}\bar{s}}^*(\ell)=\bar{\tau}_{\bar{g}\bar{r}\bar{s}}^*(\ell)$ into Eq. (6), we have

$$C_{\tilde{g}\bar{r}\bar{s}}^*(\ell)=c_{\bar{r}}^0+\alpha_{\tilde{g}}\bar{\tau}_{\bar{g}\bar{r}\bar{s}}^*(\ell)-\beta_{\tilde{g}s}\ell<c_{\bar{r}}^0+\alpha_g\bar{\tau}_{\tilde{g}\bar{r}\bar{s}}^*(\ell)-\beta_{\tilde{g}s}\ell=\pi_{\tilde{g}}^*, \forall\ell\in(\vec{\ell},\overleftarrow{\ell}). \tag{B.23}$$

Inequality (B.23) implies that the commuters of group $\tilde{g}$ can reduce their generalized travel cost by changing their relative arrival times to fall within the interval $(\vec{\ell},\overleftarrow{\ell})$. This contradicts that $\mathbf{f}^*$ is an NTE flow vector. Therefore, the commuters must always stay on the upper envelope of all the equilibrium arrival flow rate curves, i.e., $\mathbf{f}^* = \tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$.

$\mathbf{f}^*$ satisfies condition (8), which implies that $\tilde{D}_g(\boldsymbol{\pi}^*) = D_g$ is satisfied for all $g \in G$, i.e., $\tilde{\mathbf{D}}(\boldsymbol{\pi}^*) = \mathbf{D}$. Therefore, we have $\mathbf{Z}(\boldsymbol{\pi}^*) = \tilde{\mathbf{D}}(\boldsymbol{\pi}^*) - \mathbf{D} = \mathbf{0}$. This completes the proof. □

B.6. The proof of Theorem 5

Proof. By Theorem 1 and Proposition 5, we can conclude that the NTE problem with general user heterogeneity and route choice must have a solution. Let $\tilde{\mathbf{f}}$ be an NTE flow vector (i.e., a solution to the NTE problem), and $\tilde{\boldsymbol{\pi}}$ be the corresponding vector of equilibrium generalized trip costs. Based on Theorem 4, we have $\mathbf{Z}(\tilde{\boldsymbol{\pi}}) = \mathbf{0}$. By definition, we have $0 \le \eta(\boldsymbol{\pi}^*) \le \eta(\tilde{\boldsymbol{\pi}}) = \mathbf{Z}(\tilde{\boldsymbol{\pi}})^{\mathrm{T}} \mathbf{Z}(\tilde{\boldsymbol{\pi}}) = 0$. This inequality implies $\eta(\boldsymbol{\pi}^*) = 0$, and hence we have $\mathbf{Z}(\boldsymbol{\pi}^*) = \mathbf{0}$. Therefore, $\boldsymbol{\pi}^*$ is also a solution to the system of nonlinear equations (23). According to Theorem 3, $\tilde{\mathbf{f}}(\boldsymbol{\pi}^*)$ is an NTE flow vector. This completes the proof.□

B.7. The proof of Proposition 3

Proof. According to Eq. (18), $\overline{f}_{grs}(\ell)$ is monotonically increasing with respect to π_g. This together with Eqs. (19) and (20) implies that $f_{rs}(\ell)$ and $f_{grs}(\ell)$ are monotonically increasing with respect to π_g. This together with both $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-}) = \tilde{D}_g(\boldsymbol{\pi})$ and Eq. (22) implies that $\tilde{D}_g(\pi_g, \boldsymbol{\pi}_{g-})$ is monotonically increasing with respect to π_g. This completes the proof.□

B.8. The proof of Theorem 6

Proof. By definition, we have $\tau_{rs}^{CETE}(\ell) = \tau_r(f_{rs}^{CETE}(\ell))$. Substituting this equation into Eq. (31), we have

$$C_{grs}^{CETE}(\ell) = c_r^0 + \alpha_g \tau_r(f_{rs}^{CETE}(\ell)) - \beta_{gs}\ell + p_{rs}^{CET}(\ell), \ \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{B.24}$$

We define the following function:

$$\overline{C}_{grs}^{CETE}(\ell) = c_r^0 + \alpha_g \tau_r(\overline{f}_{grs}^{CETE}(\ell)) - \beta_{gs}\ell + \overline{p}_{grs}^{CET}(\ell), \ \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{B.25}$$

By definition, we have $\vec{\ell}_{grs,1} \le \vec{\ell}_{grs,k} < \overleftarrow{\ell}_{grs,k}$ for all $k \in K$. Based on Eqs. (37) and (38), if $\vec{\ell}_{grs,1} < \ell \le 0$, we have $\overline{f}_{grs}^{CETE}(\ell) > 0$ and $\overline{p}_{grs}^{CET}(\ell) > 0$; otherwise, we have $\overline{f}_{grs}^{CETE}(\ell) = 0$ and $\overline{p}_{grs}^{CET}(\ell) = 0$. Substituting Eqs. (37) and (38) into Eq. (B.25), we can conclude that

$$\overline{C}_{grs}^{CETE}(\ell) \begin{cases} = \pi_g^{CETE}, & \text{if } \ell > \vec{\ell}_{grs,1}, \\ = c_{r,1}^0 + \alpha_{gr,1}\tau_r^0 - \beta_{gs}\ell \ge \pi_g^{CETE}, & \text{if } \ell \le \vec{\ell}_{grs,1}, \end{cases} \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{B.26}$$

$\tau_r(f)$ is strictly monotone. Hence, based on condition (B.26) and Eq. (37), we have

$$\overline{C}_{grs}^{CETE}(\ell) \begin{cases} = \pi_g^{CETE}, & \text{if } \overline{f}_{grs}^{CETE}(\ell) > 0, \\ \ge \pi_g^{CETE}, & \text{if } \overline{f}_{grs}^{CETE}(\ell) = 0, \end{cases} \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{B.27}$$

Based on Eqs. (37)-(41), we have $f_{rs}^{CETE}(\ell) \ge \overline{f}_{grs}^{CETE}(\ell)$ and $p_{rs}^{CET}(\ell) \ge \overline{p}_{grs}^{CET}(\ell)$ for all $\ell \le 0$. Based on Eqs. (B.24) and (B.25), we have $C_{grs}^{CETE}(\ell) \ge \overline{C}_{grs}^{CETE}(\ell)$ for all $\ell \le 0$. Based on Eqs. (37)-(41), (B.24) and (B.25), if $f_{grs}^{CETE}(\ell)>0$, then we have $f_{rs}^{CETE}(\ell)=\overline{f}_{grs}^{CETE}(\ell)$, $p_{rs}^{CET}(\ell)=\overline{p}_{grs}^{CET}(\ell)$, and $C_{grs}^{CETE}(\ell)=\overline{C}_{grs}^{CETE}(\ell)$. Therefore, we have

$$C_{grs}^{CETE}(\ell)\begin{cases}=\overline{C}_{grs}^{CETE}(\ell), & \text{if } f_{grs}^{CETE}(\ell)>0,\\ \geq \overline{C}_{grs}^{CETE}(\ell), & \text{if } f_{grs}^{CETE}(\ell)=0,\end{cases}\forall g\in G, r\in R, s\in S, \ell\leq 0\,. \tag{B.28}$$

Based on Eq. (41), if $f_{grs}^{CETE}(\ell)>0$, we have $\overline{f}_{grs}^{CETE}(\ell)>0$. Combining conditions (B.27) and (B.28), we have

$$C_{grs}^{CETE}(\ell)\begin{cases}=\pi_{g}^{CETE}, & \text{if } f_{grs}^{CETE}(\ell)>0,\\ \geq \pi_{g}^{CETE}, & \text{if } f_{grs}^{CETE}(\ell)=0,\end{cases}\forall g\in G, r\in R, s\in S, \ell\leq 0\,, \tag{B.29}$$

which implies that the arrival flow pattern defined by Eq. (41) satisfies CETE condition (35). This completes the proof. □

Appendix C: The derivation related to the NTE

C.1. Retrieving traffic demand from equilibrium generalized trip costs

We consider a given vector of equilibrium generalized trip costs $\boldsymbol{\pi}$. Let $\vec{g}_{rs}^{0}$ be the first group with a positive retrieved number of commuters using route *r* and experiencing the *s*-th type schedule delay. For any given group *g* with a positive retrieved number of commuters using route *r* and experiencing the *s*-th type schedule delay, we let $\vec{g}_{grs}$ be the successor group of group *g* with a positive retrieved number of commuters using route *r* and experiencing the *s*-th type schedule delay. If the retrieved number of commuters of group *g* using route *r* and experiencing the *s*-th type schedule delay equals zero (i.e., $N_{grs}=0$) or group *g* is the last group with a positive retrieved number of commuters using road $r\in R$ and experiencing the *s*-th type schedule delay, then $\vec{g}_{grs}=0$; otherwise, $\vec{g}_{grs}>0$. Let $\vec{\mathbf{g}}^{0}=[g_{rs}^{0}, r\in R, s\in S]$ and $\vec{\mathbf{g}}=[g_{grs}, g\in G, r\in R, s\in S]$.

Let $\overline{\ell}_{gg'rs}$ be the relative arrival time such that the commuters of both groups g and g' who use road $r\in R$, experience the *s*-th type schedule delay, and have the same relative arrival time and the same equilibrium travel time. By definition, we have

$$\frac{\pi_{g}-c_{r}^{0}+\beta_{gs}\overline{\ell}_{gg'rs}}{\alpha_{g}}=\overline{\tau}_{grs}(\overline{\ell}_{gg'rs})=\overline{\tau}_{g'rs}(\overline{\ell}_{gg'rs})=\frac{\pi_{g'}-c_{r}^{0}+\beta_{g's}\overline{\ell}_{gg'rs}}{\alpha_{g'}}, \forall g\in G, g'\in G\setminus\{g\}, r\in R, s\in S\,. \tag{C.1}$$

Rearranging Eq. (C.1), we have

$$\overline{\ell}_{gg'rs}=-\frac{\alpha_{g'}\pi_{g}-\alpha_{g}\pi_{g'}-\alpha_{g'}c_{r}^{0}+\alpha_{g}c_{r}^{0}}{\alpha_{g'}\beta_{gs}-\alpha_{g}\beta_{g's}}, \forall g\in G, g'\in G\setminus\{g\}, r\in R, s\in S\,. \tag{C.2}$$

For any given groups $g\in G$ and $g'\in G\setminus\{g\}$, if the arrival order ID of group *g* is smaller than that of group g', we have $\overline{\tau}_{grs}(\ell)>\overline{\tau}_{g'rs}(\ell)$ for all $\ell\in(-\infty,\overline{\ell}_{gg'rs})$ and $\overline{\tau}_{grs}(\ell)<\overline{\tau}_{g'rs}(\ell)$ for all $\ell\in(\overline{\ell}_{gg'rs},0]$. In particular, if $g'=\vec{g}_{grs}\neq 0$ (i.e., group g' is the successor group of group *g*), $\overline{\ell}_{gg'rs}$ is the watershed line of the relative arrival time that separates the commuters of groups g and g'.

Using Eq. (18), we define the following function, which is used to retrieve traffic demand from equilibrium generalized trip costs in Procedure C.1:

$$\overline{F}_{grs}(\ell)=\int_{\overline{\ell}_{grs}}^{\ell}\overline{f}_{grs}(l)dl=\frac{\alpha_g L_r\zeta_r Q_r\kappa}{\beta_{gs}(\kappa+1)}\left(\frac{\pi_g-c_r^0+\beta_{gs}\ell}{\alpha_g L_r}-\frac{\tau_r^0}{L_r}\right)^{\frac{\kappa+1}{\kappa}},\forall g\in G,r\in R,s\in S\,, \tag{C.3}$$

where $\overline{\ell}_{grs}$ is defined by Eq. (15).

Procedure C.1 can be used to retrieve traffic demand from equilibrium generalized trip costs. In Procedure C.1, Lines 4-12 are used to determine the groups with a positive number of commuters who use road *r* and experience the *s*-th type of schedule delay, and also the corresponding successor group of those groups. Lines 13-19 are used to update the retrieved traffic demand of each group.

Procedure C.1 Retrieving traffic demand from equilibrium generalized trip costs at the NTE

Input: A vector of equilibrium generalized trip costs $\boldsymbol{\pi}$.
1: Set $\tilde{\mathbf{D}}(\boldsymbol{\pi})=\mathbf{0}$, $\vec{\mathbf{g}}^0=\mathbf{0}$, and $\vec{\mathbf{g}}=\mathbf{0}$.
2: **for each** $r\in R$ **do**
3: **for each** $s\in S$ **do**
4: $i=\arg\min_{i'\in I}\overline{\ell}_{g_s(i')rs}$ and $\vec{g}_{rs}^0=g_s(i)$.
5: **while** $i<n$ **do**
6: $j=\arg\min_{j'>i}\overline{\ell}_{g_s(i)g_s(j'),rs}$.
7: **if** $\overline{\ell}_{g_s(i)g_s(j)rs}\ge 0$ then
8: Set $i=n$.
9: **else**
10: Set $\vec{g}_{g_s(i)rs}=g_s(j)$ and $i=j$.
11: **end if**
12: **end while**
13: Set $g=g_{rs}^0$ and $g'=\vec{g}_{grs}$.
14: **while** $g'>0$ **do**
15: Set $\tilde{D}_g(\boldsymbol{\pi})=\tilde{D}_g(\boldsymbol{\pi})+\overline{F}_{grs}(\overline{\ell}_{gg'rs})$, and $\tilde{D}_{g'}(\boldsymbol{\pi})=\tilde{D}_{g'}(\boldsymbol{\pi})-\overline{F}_{g'rs}(\overline{\ell}_{gg'rs})$.
16: Set $g=g'$ and $g'=\vec{g}_{grs}$.
17: **end while**
18: Set $\tilde{D}_g(\boldsymbol{\pi})=\tilde{D}_g(\boldsymbol{\pi})+\overline{F}_{grs}(0)$.
19: **end for**
20: **end for**
Output: $\tilde{\mathbf{D}}(\boldsymbol{\pi})$, $\vec{\mathbf{g}}^0$, and $\vec{\mathbf{g}}$.

C.2. The gradient of the objective function of the unconstrained optimization problem

Based on Eq. (C.2), we have $\overline{\ell}_{gg'rs}=\overline{\ell}_{g'grs}$. Taking derivatives on both sides of Eq. (C.2), we have

$$\frac{\partial \overline{\ell}_{g'grs}}{\partial \pi_g} = \frac{\alpha_{g'}}{\alpha_g \beta_{g's} - \alpha_{g'} \beta_{gs}}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S \quad \text{and} \tag{C.4}$$

$$\frac{\partial \overline{\ell}_{g'grs}}{\partial \pi_{g'}} = \frac{\alpha_g}{\alpha_{g'} \beta_{gs} - \alpha_g \beta_{g's}}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S . \tag{C.5}$$

Substituting Eq. (C.2) into Eq. (C.3), taking derivatives of the resultant expression, and simplifying the resulting derivatives using Eqs. (C.4) and (C.5), we have

$$\begin{aligned} \frac{\partial \overline{F}_{grs}(\overline{\ell}_{gg'rs})}{\partial \pi_g} &= \frac{\partial}{\partial \pi_g}\left[\frac{\alpha_g L_r \zeta_r Q_r \kappa}{\beta_{gs}(\kappa+1)} \left(\frac{\pi_g - c_r^0 + \beta_{gs}\overline{\ell}_{g'grs}}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{\kappa+1}{\kappa}} \right] \\ &= \left(1 + \frac{\beta_{gs} \partial \overline{\ell}_{g'grs}}{\partial \pi_g} \right) \cdot \frac{\zeta_r Q_r}{\beta_{gs}} \left(\frac{\pi_g - c_r^0 + \beta_{gs}\overline{\ell}_{g'grs}}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{1}{\kappa}} = \frac{\alpha_g \beta_{g's} \overline{f}_{grs}(\overline{\ell}_{gg'rs})}{\beta_{gs}(\alpha_g \beta_{g's} - \alpha_{g'}\beta_{gs})}, \end{aligned} \tag{C.6}$$

$$\begin{aligned} \frac{\partial \overline{F}_{grs}(\overline{\ell}_{gg'rs})}{\partial \pi_{g'}} &= \frac{\partial}{\partial \pi_{g'}}\left[\frac{\alpha_g L_r \zeta_r Q_r \kappa}{\beta_{gs}(\kappa+1)} \left(\frac{\pi_g - c_r^0 + \beta_{gs}\overline{\ell}_{g'grs}}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{\kappa+1}{\kappa}} \right] \\ &= \frac{\partial \overline{\ell}_{g'grs}}{\partial \pi_{g'}} \cdot \zeta_r Q_r \left(\frac{\pi_g - c_r^0 + \beta_{gs}\overline{\ell}_{g'grs}}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{1}{\kappa}} = \frac{\alpha_g \overline{f}_{grs}(\overline{\ell}_{gg'rs})}{\alpha_{g'}\beta_{gs} - \alpha_g \beta_{g's}}, \end{aligned} \quad \text{and} \tag{C.7}$$

$$\frac{\partial \overline{F}_{grs}(0)}{\partial \pi_g} = \frac{\partial}{\partial \pi_g}\left[\frac{\alpha_g L_r \zeta_r Q_r \kappa}{\beta_{gs}(\kappa+1)} \left(\frac{\pi_g - c_r^0}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{\kappa+1}{\kappa}} \right] = \frac{1}{\beta_{gs}} \cdot \zeta_r Q_r \left(\frac{\pi_g - c_r^0}{\alpha_g L_r} - \frac{\tau_r^0}{L_r} \right)^{\frac{1}{\kappa}} = \frac{\overline{f}_{grs}(0)}{\beta_{gs}}. \tag{C.8}$$

Using derivatives (C.6)-(C.8) and Procedure C.2, we can obtain the Jacobian of the retrieved traffic demand function.

Procedure C.2 The Jacobian of the retrieved traffic demand function

---

Inputs: A vector of equilibrium generalized trip costs $\boldsymbol{\pi}$, a vector of the first arrival group $\vec{\mathbf{g}}^0$, and a vector of successor groups $\vec{\mathbf{g}}$.

Set $\nabla \tilde{\mathbf{D}}(\boldsymbol{\pi}) = \mathbf{0}$.

for each $r \in R$ **do**

  for each $s \in S$ **do**

    Set $g = g_{rs}^0$ and $g' = \vec{g}_{grs}$.

    while $g' > 0$ **do**

      Set $\frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_g} = \frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_g} + \frac{\partial \overline{F}_{grs}(\overline{\ell}_{gg'rs})}{\partial \pi_g}$, $\frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_{g'}} = \frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_{g'}} + \frac{\partial \overline{F}_{grs}(\overline{\ell}_{gg'rs})}{\partial \pi_{g'}}$,
      $\frac{\partial \tilde{D}_{g'}(\boldsymbol{\pi})}{\partial \pi_g} = \frac{\partial \tilde{D}_{g'}(\boldsymbol{\pi})}{\partial \pi_g} - \frac{\partial \overline{F}_{g'rs}(\overline{\ell}_{gg'rs})}{\partial \pi_g}$, and $\frac{\partial \tilde{D}_{g'}(\boldsymbol{\pi})}{\partial \pi_{g'}} = \frac{\partial \tilde{D}_{g'}(\boldsymbol{\pi})}{\partial \pi_{g'}} - \frac{\partial \overline{F}_{g'rs}(\overline{\ell}_{gg'rs})}{\partial \pi_{g'}}$.

      Set $g = g'$ and $g' = \vec{g}_{grs}$.

end while

Set $\frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_g} = \frac{\partial \tilde{D}_g(\boldsymbol{\pi})}{\partial \pi_g} + \frac{\partial \overline{F}_{grs}(0)}{\partial \pi_g}$.

end for

end for

Output: $\nabla \tilde{\mathbf{D}}(\boldsymbol{\pi})$.

Using the Jacobian $\nabla \tilde{\mathbf{D}}(\boldsymbol{\pi})$, we can obtain the gradient of the objective function of problem (24) as follows:

$$\frac{\partial \eta(\boldsymbol{\pi})}{\partial \pi_g} = 2\sum_{g' \in G} [\tilde{D}_{g'}(\boldsymbol{\pi}) - D_{g'}] \frac{\partial \tilde{D}_{g'}(\boldsymbol{\pi})}{\partial \pi_g}. \tag{C.9}$$

Appendix D: The derivation related to the CETE

D.1. Retrieving traffic demand from equilibrium generalized trip costs

At the NTE, the relative cost of schedule delay to travel time of each group is independent of commuter's travel time and the arrival of each group at the destination in the whole studied period follows a predetermined order. However, at the CETE, the relative cost of schedule delay to travel time of each group, i.e., $\beta_{gs} / \alpha_{gr,k}$, is dependent on commuter's travel time, and the arrival of each group at the destination in the whole studied period does not follow a predetermined order. The arrival of each group at the destination follows the ascending order of $\beta_{gs} / \alpha_{gr,k}$ just when $\ell \in [\vec{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]$. At the CETE, let $g_{rs,k}(i)$ be the group ID of the i-th arrival group from road r experiencing the s-th type schedule delay and having relative arrival time $\ell \in [\vec{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]$. Similar to the case of $\ell \in [\vec{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]$, based on Eq. (30), it is found that the arrival of each group at the destination follows the ascending order of β_{gs} when $\ell \in (\overleftarrow{\ell}_{rs,k}, \vec{\ell}_{rs,k+1})$. Let $\tilde{g}_s(i)$ be the group ID of the i-th arrival group experiencing the s-th type schedule delay and having relative arrival time $\ell \in (\overleftarrow{\ell}_{rs,k}, \vec{\ell}_{rs,k+1})$ for all roads $r \in R$ and linear interpolants $k \in K$. By definition, $[g_{rs,k}(i), i \in I, r \in R, s \in S, k \in K]$ and $[\tilde{g}_s(i), i \in I, s \in S]$ provide mappings from arrival order IDs to group IDs.

We consider a given vector of equilibrium generalized trip costs $\boldsymbol{\pi}^{CET}$. Let $\overline{g}^0_{rs,k}$ be the first group with a positive retrieved number of commuters who use route r, experience the s-th type schedule delay, and have travel time in the range $(\tau_{r,k}, \tau_{r,k+1})$. For any given group g with a positive retrieved number of commuters using route r, experiencing the s-th type schedule delay, and having travel time in the range $(\tau_{r,k}, \tau_{r,k+1})$, we let $\overline{g}_{grs,k}$ be the successor group of group g with a positive retrieved number of commuters who use route r, experience the s-th type schedule delay, and have travel time in the range $(\tau_{r,k}, \tau_{r,k+1})$. Let $\overline{\mathbf{g}}^0 = [\overline{g}^0_{rs,k}, r \in R, s \in S, k \in K]$ and $\overline{\mathbf{g}} = [\overline{g}_{grs,k}, g \in G, r \in R, s \in S, k \in K]$. If the retrieved number of commuters of group g using route r, experiencing the s-th type schedule delay, and having the travel time in the range $[\tau_{r,k}, \tau_{r,k+1}]$ equals zero or group g is the last group, then $\overline{g}_{grs,k} = 0$; otherwise, $\overline{g}_{grs,k} > 0$. Let

$\tilde{g}^{0}_{rs,k}$ be the first group with a positive retrieved number of commuters who use route *r*, experience the *s*-th type schedule delay, and have the travel time $\tau_{r,k+1}$. For any given group *g* with a positive retrieved number of commuters using route *r*, experiencing the *s*-th type schedule delay, and having the travel time $\tau_{r,k+1}$, let $\tilde{g}_{grs,k}$ be the successor group of group *g* with a positive retrieved number of commuters who use route *r*, experience the *s*-th type schedule delay, and have the travel time $\tau_{r,k+1}$. Let $\tilde{\mathbf{g}}^{0} = [\tilde{g}^{0}_{rs,k}, r \in R, s \in S, k \in K]$ and $\tilde{\mathbf{g}} = [\tilde{g}_{grs,k}, g \in G, r \in R, s \in S, k \in K]$. If the retrieved number of commuters of group *g* using route *r*, experiencing the *s*-th type schedule delay, and having the travel time $\tau_{r,k+1}$ equals zero or group *g* is the last group, then $\tilde{g}_{grs,k} = 0$; otherwise, $\tilde{g}_{grs,k} > 0$. Let $\overline{\ell}_{gg'rs,k}$ and $\tilde{\ell}_{gg'rs,k}$ be the relative arrival times such that the following conditions are satisfied, respectively:

$$\zeta_r Q_r \left(\frac{\pi_g^{CETE} - c^0_{r,k} + \beta_{gs} \overline{\ell}_{gg'rs,k} - \alpha_{gr,k} \tau_r^0}{\alpha_{gr,k} L_r} \right)^{\frac{1}{\kappa}} = \zeta_r Q_r \left(\frac{\pi_{g'}^{CETE} - c^0_{r,k} + \beta_{g's} \overline{\ell}_{gg'rs,k} - \alpha_{g'r,k} \tau_r^0}{\alpha_{g'r,k} L_r} \right)^{\frac{1}{\kappa}} \quad \text{and} \tag{D.1}$$

$$\pi_g^{CETE} - c_r^0 - \alpha_g \tau_{r,k+1} + \beta_{gs} \tilde{\ell}_{gg'rs,k} = \pi_{g'}^{CETE} - c_r^0 - \alpha_{g'} \tau_{r,k+1} + \beta_{g's} \tilde{\ell}_{gg'rs,k} . \tag{D.2}$$

Based on Eqs. (37) and (D.1), if $\overline{\ell}_{gg'rs,k} \in [\overrightarrow{\ell}_{grs,k}, \overleftarrow{\ell}_{grs,k}] \cap [\overrightarrow{\ell}_{g'rs,k}, \overleftarrow{\ell}_{g'rs,k}]$, we have $\overline{f}^{CETE}_{grs,k}(\overline{\ell}_{gg'rs,k}) = \overline{f}^{CETE}_{g'rs,k}(\overline{\ell}_{gg'rs,k})$. Based on Eqs. (38) and (D.2), if $\overline{\ell}_{gg'rs,k} \in (\overleftarrow{\ell}_{grs,k}, \overrightarrow{\ell}_{grs,k+1}) \cap (\overleftarrow{\ell}_{g'rs,k}, \overrightarrow{\ell}_{g'rs,k+1})$, we have $\overline{p}^{CET}_{grs,k}(\tilde{\ell}_{gg'rs,k}) = \overline{p}^{CET}_{g'rs,k}(\tilde{\ell}_{gg'rs,k})$.

Rearranging Eqs. (D.1) and (D.2), we have

$$\overline{\ell}_{gg'rs,k} = -\frac{\alpha_{g'r,k}\pi_g^{CETE} - \alpha_{gr,k}\pi_{g'}^{CETE} - \alpha_{g'r,k} c^0_{r,k} + \alpha_{gr,k} c^0_{r,k}}{\alpha_{g'r,k}\beta_{gs} - \alpha_{gr,k}\beta_{g's}}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S, k \in K \quad \text{and} \tag{D.3}$$

$$\tilde{\ell}_{gg'rs,k} = -\frac{\pi_g^{CETE} - \pi_{g'}^{CETE} - (\alpha_g - \alpha_{g'})\tau_{r,k+1}}{\beta_{gs} - \beta_{g's}}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S, k \in K . \tag{D.4}$$

For any given interpolant $k \in K$ and any given groups $g \in G$ and $g' \in G \setminus \{g\}$, if the arrival order ID of group *g* is smaller than that of group g' for $\ell \in [\overrightarrow{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]$, we have $\overline{f}^{CETE}_{grs,k}(\ell) > \overline{f}^{CETE}_{g'rs,k}(\ell)$ for all $\ell \in [\overrightarrow{\ell}_{rs,k}, \overline{\ell}_{gg'rs,k})$ and $\overline{f}^{CETE}_{grs,k}(\ell) < \overline{f}^{CETE}_{g'rs,k}(\ell)$ for all $\ell \in (\overline{\ell}_{gg'rs,k}, \overleftarrow{\ell}_{rs,k}]$. If the arrival order ID of group *g* is smaller than that of group g' for $\ell \in (\overleftarrow{\ell}_{rs,k}, \overrightarrow{\ell}_{rs,k+1})$, we have $\tilde{p}^{CET}_{grs,k}(\ell) > \tilde{p}^{CET}_{g'rs,k}(\ell)$ for all $\ell \in [\overleftarrow{\ell}_{rs,k}, \tilde{\ell}_{gg'rs,k})$ and $\tilde{p}^{CET}_{grs,k}(\ell) < \tilde{p}^{CET}_{g'rs,k}(\ell)$ for all $\ell \in (\tilde{\ell}_{gg'rs,k}, \overrightarrow{\ell}_{rs,k+1}]$. In particular, for a given group $g \in G$, if $g' = \overline{g}_{grs,k} > 0$, $\overline{\ell}_{gg'rs,k}$ is the watershed line that separates the commuters of group g and its successor group g' for $\ell \in [\overrightarrow{\ell}_{rs,k}, \overleftarrow{\ell}_{rs,k}]$. If $g' = \tilde{g}_{grs,k} > 0$, $\tilde{\ell}_{gg'rs,k}$ is the watershed line that separates the commuters of group g and its successor group g' for $\ell \in (\overleftarrow{\ell}_{rs,k}, \overrightarrow{\ell}_{rs,k+1})$.

We define the following function, which is used to retrieve the CETE traffic demand from equilibrium generalized trip costs:

$$\overline{F}_{grs,k}^{CETE}(\ell)=\int_{\overline{\ell}_{grs,k}}^{\ell}\overline{f}_{grs,k}^{CETE}(l)dl=\frac{\alpha_{gr,k}L_r\zeta_rQ_r\kappa}{\beta_{gs}(\kappa+1)}\left(\frac{\pi_g^{CETE}-c_{r,k}^0+\beta_{gs}\ell}{\alpha_{gr,k}L_r}-\frac{\tau_r^0}{L_r}\right)^{\frac{\kappa+1}{\kappa}}, \tag{D.5}$$

where

$$\overline{\ell}_{grs,k}=-\frac{\pi_g^{CETE}-c_{r,k}^0-\alpha_{gr,k}\tau_r^0}{\beta_{gs}}.$$

Procedure D.1 can be used to retrieve traffic demand from equilibrium generalized trip costs. In Procedure D.1, Lines 5-14 are used to determine the arrival sequence of the commuters who use road *r*, experience the *s*-th type of schedule delay, and have relative arrival time $\ell\in[\overrightarrow{\ell}_{rs,k},\overleftarrow{\ell}_{rs,k}]$. Lines 15-20 are used to update the retrieved traffic demand of each group having relative arrival time $\ell\in[\overrightarrow{\ell}_{rs,k},\overleftarrow{\ell}_{rs,k}]$. Lines 21-30 are used to determine the arrival sequence of the commuters who use road *r*, experience the *s*-th type of schedule delay, and have relative arrival time $\ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1})$. Lines 31-36 are used to update the retrieved traffic demand of each group having relative arrival time $\ell\in(\overleftarrow{\ell}_{rs,k},\overrightarrow{\ell}_{rs,k+1})$.

Procedure D.1 Retrieving CETE traffic demand from equilibrium generalized trip costs

---

Input: A vector of generalized trip costs $\boldsymbol{\pi}^{CETE}$

1: Set $\tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE})=\mathbf{0}$, $\overline{\mathbf{g}}^0=\mathbf{0}$, $\overline{\mathbf{g}}=\mathbf{0}$, $\tilde{\mathbf{g}}^0=\mathbf{0}$, and $\tilde{\mathbf{g}}=\mathbf{0}$.
2: **for** each $r\in R$ **do**
3:     **for** each $s\in S$ **do**
4:         Set $k=1$
5:         **while** $\min_{i'\in I}\overrightarrow{\ell}_{g_{rs,k}(i')rs,k}<0$ **do**
6:             $i=\arg\min_{i'\in I}\overrightarrow{\ell}_{g_{rs,k}(i')rs,k}$ and $\overline{g}_{rs,k}^0=g_{rs,k}(i)$
7:             **while** $i<n$ **do**
8:                 $j=\arg\min_{j'>i}\overline{\ell}_{g_{rs,k}(i)g_{rs,k}(j')rs,k}$
9:                 **if** $\overline{\ell}_{g_{rs,k}(i)g_{rs,k}(j)rs,k}\geq\min\{\overleftarrow{\ell}_{g_{rs,k}(i)rs,k},0\}$ then
10:                     Set $i=n$.
11:                 **else**
12:                     Set $\overline{g}_{g_{rs,k}(i)rs,k}=g_{rs,k}(j)$ and $i=j$.
13:                 **end if**
14:             **end while**
15:             Set $g=\overline{g}_{rs,k}^0$, $\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})=\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})-\overline{F}_{grs,k}^{CETE}(\overrightarrow{\ell}_{grs,k})$, and $g'=\overline{g}_{grs,k}$.
16:             **while** $g'>0$ **do**
17:                 Set $\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})=\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})+\overline{F}_{grs,k}^{CETE}(\overline{\ell}_{gg'rs,k})$ and
                    $\tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})=\tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})-\overline{F}_{g'rs,k}^{CETE}(\overline{\ell}_{gg'rs,k})$.
18:                 Set $g=g'$ and $g'=\overline{g}_{grs,k}$.
19:             **end while**
20:             Set $\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})=\tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})+\overline{F}_{grs,k}^{CETE}(\min\{\overleftarrow{\ell}_{grs,k},0\})$.

if $\min_{i' \in I} \overleftarrow{\ell}_{\tilde{g}_s(i')rs,k} < 0$ **do**

$i = \arg\min_{i' \in I} \overleftarrow{\ell}_{\tilde{g}_s(i')rs,k}$ and $\tilde{g}^0_{rs,k} = \tilde{g}_s(i)$

while $i < n$ **do**

$j = \arg\min_{j' > i} \tilde{\ell}_{\tilde{g}_s(i)\tilde{g}_s(j')rs,k}$

if $\tilde{\ell}_{\tilde{g}_s(i)\tilde{g}_s(j)rs,k} \geq \min\{\vec{\ell}_{\tilde{g}_s(i)rs,k+1}, 0\}$ then

Set $i = n$.

else

Set $\tilde{g}_{\tilde{g}_s(i)rs,k} = \tilde{g}_s(j)$ and $i = j$.

end if

end while

Set $g = \tilde{g}^0_{rs,k}$, $\tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) = \tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) - \vec{\ell}_{grs,k}\tilde{f}^{CETE}_{r,k}$, and $g' = \tilde{g}_{grs,k}$.

while $g' > 0$ **do**

Set $\tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) = \tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) + \tilde{\ell}_{gg'rs,k}\tilde{f}^{CETE}_{r,k}$ and
$\tilde{D}^{CETE}_{g'}(\boldsymbol{\pi}^{CETE}) = \tilde{D}^{CETE}_{g'}(\boldsymbol{\pi}^{CETE}) - \tilde{\ell}_{gg'rs,k}\tilde{f}^{CETE}_{r,k}$.

Set $g = g'$ and $g' = \tilde{g}_{grs,k}$.

end while

Set $\tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) = \tilde{D}^{CETE}_g(\boldsymbol{\pi}^{CETE}) + \min\{\vec{\ell}_{grs,k+1}, 0\} \cdot \tilde{f}^{CETE}_{r,k}$.

Set $k = k + 1$.

end if

end while

end for

end for

Output: $\tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE})$, $\overline{\mathbf{g}}^0$, $\overline{\mathbf{g}}$, $\tilde{\mathbf{g}}^0$, and $\tilde{\mathbf{g}}$.

D.2. The gradient of the objective function of the unconstrained optimization problem

Based on Eq. (D.3), we have $\overline{\ell}_{gg'rs,k} = \overline{\ell}_{g'grs,k}$. Substituting Eq. (D.3) into Eq. (D.5), and taking derivatives of the resultant expression, we have

$$\frac{\partial \overline{F}^{CETE}_{grs,k}(\overline{\ell}_{gg'rs,k})}{\partial \pi^{CETE}_g} = \frac{\partial \overline{F}^{CETE}_{grs,k}(\overline{\ell}_{g'grs,k})}{\partial \pi^{CETE}_g} = \frac{\alpha_{g,k}\beta_{g's}\overline{f}^{CETE}_{grs,k}(\overline{\ell}_{gg'rs,k})}{\beta_{gs}(\alpha_{g,k}\beta_{g's} - \alpha_{g',k}\beta_{gs})}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S, k \in K, \tag{D.6}$$

$$\frac{\partial \overline{F}^{CETE}_{grs,k}(\overline{\ell}_{gg'rs,k})}{\partial \pi^{CETE}_{g'}} = \frac{\partial \overline{F}^{CETE}_{grs,k}(\overline{\ell}_{g'grs,k})}{\partial \pi^{CETE}_{g'}} = \frac{\alpha_{g,k}\overline{f}^{CETE}_{grs,k}(\overline{\ell}_{gg'rs,k})}{\alpha_{g',k}\beta_{gs} - \alpha_{g,k}\beta_{g's}}, \forall g \in G, g' \in G \setminus \{g\}, r \in R, s \in S, k \in K, \tag{D.7}$$

$$\frac{\partial \overline{F}^{CETE}_{grs,k}(\vec{\ell}_{grs,k})}{\partial \pi^{CETE}_g} = \frac{\partial \overline{F}^{CETE}_{grs,k}(\overleftarrow{\ell}_{grs,k})}{\partial \pi^{CETE}_g} = 0, \forall g \in G, r \in R, s \in S, k \in K, \text{ and} \tag{D.8}$$

$$\frac{\partial \overline{F}^{CETE}_{grs,k}(0)}{\partial \pi^{CETE}_g} = \frac{\overline{f}^{CETE}_{grs,k}(0)}{\beta_{gs}}, \forall g \in G, r \in R, s \in S, k \in K. \tag{D.9}$$

Based on Eq. (36), we have

$$\frac{\partial \vec{\ell}_{grs,k}}{\partial \pi_g^{CETE}} = \frac{\partial \overleftarrow{\ell}_{grs,k}}{\partial \pi_g^{CETE}} = -\frac{1}{\beta_{gs}}, \forall g \in G, r \in R, s \in S, k \in K. \tag{D.10}$$

Using the derivatives (D.6)-(D.10) and Procedure D.2, we can obtain the Jacobian of the retrieved traffic demand function.

Procedure D.2 The Jacobian of retrieved traffic demand function at the CETE

Inputs: A vector of equilibrium generalized trip costs $\boldsymbol{\pi}^{CET}$, vectors of the first arrival group $\overline{\mathbf{g}}^0$ and $\tilde{\mathbf{g}}^0$, and vectors of successor groups $\overline{\mathbf{g}}$ and $\tilde{\mathbf{g}}$.

Set $\nabla \tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE}) = \mathbf{0}$.

for each $r \in R$ **do**

  for each $s \in S$ **do**

    Set $k = 1$.

    while $\overline{g}_{rs,k}^0 > 0$ **do**

      Set $g = \overline{g}_{rs,k}^0$ and $g' = \overline{g}_{grs,k}$.

      while $g' > 0$ **do**

        Set $\frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} = \frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} + \frac{\partial \overline{F}_{grs,k}^{CETE}(\overline{\ell}_{gg'rs,k})}{\partial \pi_g^{CETE}}$,

$$\frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} = \frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} + \frac{\partial \overline{F}_{grs,k}^{CETE}(\overline{\ell}_{gg'rs,k})}{\partial \pi_{g'}^{CETE}},$$

$$\frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} = \frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} - \frac{\partial \overline{F}_{g'rs,k}^{CETE}(\overline{\ell}_{gg'rs,k})}{\partial \pi_{g'}^{CETE}}, \text{ and}$$

$$\frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} = \frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} - \frac{\partial \overline{F}_{g'rs,k}^{CETE}(\overline{\ell}_{gg'rs,k})}{\partial \pi_g^{CETE}}.$$

        Set $g = g'$ and $g' = \overline{g}_{grs,k}$.

      end while

      if $\overleftarrow{\ell}_{grs,k} > 0$ then

        Set $\frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} = \frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} + \frac{\partial \overline{F}_{grs,k}^{CETE}(0)}{\partial \pi_g^{CETE}}$.

      end if

      if $\tilde{g}_{rs,k}^0 > 0$ **do**

        Set $g = \tilde{g}_{rs,k}^0$, $\tilde{D}_g^{CETE} = \tilde{D}_g^{CETE} - \vec{\ell}_{g_{s,k}(i')rs,k} \tilde{f}_{r,k}^{CETE}$, and $g' = \overline{g}_{grs,k}$.

        while $g' > 0$ **do**

          Set $\frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} = \frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_g^{CETE}} + \frac{\partial \tilde{\ell}_{gg'rs,k}}{\partial \pi_g^{CETE}} \tilde{f}_{r,k}^{CETE}$,

$$\frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} = \frac{\partial \tilde{D}_g^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} + \frac{\partial \tilde{\ell}_{gg'rs,k}}{\partial \pi_{g'}^{CETE}} \tilde{f}_{r,k}^{CETE},$$

$$\frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} = \frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g'}^{CETE}} - \frac{\partial \tilde{\ell}_{gg'rs,k}}{\partial \pi_{g'}^{CETE}} \tilde{f}_{r,k}^{CETE}, \text{ and}$$

$$\frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}} = \frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}} - \frac{\partial \tilde{\ell}_{gg'rs,k}}{\partial \pi_{g}^{CETE}} \tilde{f}_{r,k}^{CETE}.$$

Set $g = g'$ and $g' = \overline{g}_{grs,k}$.

end while

if $\vec{\ell}_{grs,k+1} < 0$ then

Set $\frac{\partial \tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}} = \frac{\partial \tilde{D}_{g}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}} + \frac{\partial \vec{\ell}_{grs,k+1}}{\partial \pi_{g}^{CETE}} \tilde{f}_{r,k}^{CETE}$.

end if

Set $k = k+1$.

end if

end while

end for

end for

Output: $\nabla \tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE})$.

Using the Jacobian $\nabla \tilde{\mathbf{D}}^{CETE}(\boldsymbol{\pi}^{CETE})$, we can obtain the gradient of the objective function of problem (45) as follows:

$$\frac{\partial \eta^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}} = 2\sum_{g' \in G} [\tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE}) - D_{g'}] \frac{\partial \tilde{D}_{g'}^{CETE}(\boldsymbol{\pi}^{CETE})}{\partial \pi_{g}^{CETE}}. \tag{D.11}$$

Appendix E: Congestion toll equilibrium with general user heterogeneity and route choice

Similar to the CTE with homogeneous users (Chu, 1995), the marginal system travel cost of commuters under the CTE with general user heterogeneity and route choice can be formulated as follows:

$$MC_{grs}^{CT}(\ell) = C_{grs}(\ell) + \alpha_g \kappa \left[\tau_{rs}(\ell) - \tau_r^0 \right], \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{E.1}$$

Based on Eq. (E.1), we have

$$MC_{grs}^{CT}(\ell) - C_{grs}(\ell) = \alpha_g \kappa \left[\tau_{r}(\ell) - \tau_r^0 \right], \forall g \in G, r \in R, s \in S, \ell \le 0. \tag{E.2}$$

Eq. (E.2) implies that the difference between the marginal system travel cost $MC_{grs}^{CT}(\ell)$ and the marginal private cost $C_{grs}(\ell)$ depends on commuters' VOT. To have an anonymous toll scheme, we design the following CT scheme:

$$p_{rs}^{CT}(\ell) = \alpha_r^{CT} \kappa \left[\tau_{rs}(\ell) - \tau_r^0 \right], \forall g \in G, r \in R, s \in S, \ell \le 0, \tag{E.3}$$

where α_r^{CT} is a parameter associated with the CT scheme for road *r* and the vector $\boldsymbol{\alpha}^{CT} = [\alpha_r^{CT}, \forall r \in R]$ is required to be optimized for the scheme. For the CTE problem in this section, $\boldsymbol{\alpha}^{CT}$ is assumed to be given. However, an optimal vector of $\boldsymbol{\alpha}^{CT}$ can be obtained by solving a bi-level optimization model similar to the model presented in Section 4.1.

Commuters' generalized trip cost under the CT scheme (E.3) can be formulated as follows:

$$c_{grs}^{CTE}(\ell) = c_{grs}(\ell) + p_{rs}^{CT}(\ell) = (c_r^0 - \alpha_r^{CT}\kappa\tau_r^0) + (\alpha_g + \alpha_r^{CT}\kappa)\tau_{rs}(\ell) + \beta_{gs}\ell, \forall g \in G, r \in R, s \in S, \ell \le 0, \quad \text{(E.4)}$$

where the superscript "CTE" is added to the notations associated with the NTE to represent the CTE counterparts.

If the vector $\boldsymbol{\alpha}^{CT}$ is given, the generalized trip cost under the CTE (i.e., in Eq. (E.4)) is similar to that under the NTE with c_r^0 replaced with $c_r^0 - \alpha_r^{CT}\kappa\tau_r^0$ and α_g replaced with $\alpha_g + \alpha_r^{CT}\kappa$. Therefore, we can solve the CTE problem by solving an NTE problem.

References

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. J. Urb. Econ. 27 (1), 111-130.

Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: A traffic bottleneck with elastic demand. Am. Econ. Rev. 161-179.

Arnott, R., de Palma, A., Lindsey, R., 1994. The welfare effects of congestion tolls with heterogeneous commuters. J. Transp. Econ. Policy 28 (2), 139-161.

Aziz, H.A., Ukkusuri, S.V., Zhan, X., 2017. Determining the impact of personal mobility carbon allowance schemes in transportation networks. Netw. Spat. Econ. 17(2), 505-545.

California Department of Transportation, 1999. Performance measurement system. Tech. Rep., Sacramento, California.

Carey, M., Watling, D., 2012. Dynamic traffic assignment approximating the kinematic wave model: System optimum, marginal costs, externalities and tolls. Transp. Res. Part B 46(5), 634-648.

Chen, H., Liu, Y., Nie, Y., 2015. Solving the step-tolled bottleneck model with general user heterogeneity. Transp. Res. Part B 81, 210-229.

Chu, X., 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. J. Urb. Econ. 37, 324-348.

Coria, J., Zhang, X.B., 2017. Optimal environmental road pricing and daily commuting patterns. Transp. Res. Part B 105, 297-314.

Dial, R.B., 2006. A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. Transp. Res. Part B 40, 917-936.

Friesz, T.L., Bernstein, D., Smith, T.E., Tobin, R.L., Wie, B., 1993. A variational inequality formulation of the dynamic networks user equilibrium problem. Oper. Res. 41 (1), 179-191.

Florian, M., Mahut, M., Tremblay, N., 2008. Application of a simulation-based dynamic traffic assignment model. Eur. J. Oper. Res. 189(3), 1381-1392.

Ghali, M.O., Smith, M.J., 1993. Traffic assignment, traffic control and road pricing. In: Daganzo, C.F. (Ed.), Transportation and Traffic Theory. Amsterdam, Netherlands: Elsevier, pp. 147-169.

Guo, R.Y., Yang, H., Huang, H.J., 2018. Are we really solving the dynamic traffic equilibrium problem with a

departure time choice? Transp. Sci. 52(3), 603-620.
Harks, T., Klimm, M., Schneider, M., 2018. Bottleneck routing with elastic demands. Oper. Res. Lett. 46(1), 93-98.
Hall, J.D., 2019. Can tolling help everyone? Estimating the aggregate and distributional consequences of congestion pricing. Working Paper.
Hendrickson, C., Kocur, G., 1981. Schedule delay and departure time decisions in a deterministic model. Transp. Sci. 15 (1), 62-77.
Henderson, J.V., 1974. Road congestion: a reconsideration of pricing theory. J. Urb. Econ. 1, 346-355.
Henderson, J.V., 1977. Economic Theory and the Cities. Academic Press, New York chapter 8.
Henderson, J.V., 1981. The economics of staggered work hours. J. Urb. Econ. 9(3), 349-364.
Laih, C.H., 1994. Queueing at a bottleneck with single-and multi-step tolls. Transp. Res. Part A 28 (3), 197-208.
Li, Y., Ukkusuri, S.V., Fan, J., 2018. Managing congestion and emissions in transportation networks with dynamic carbon credit charge scheme. Comput. Oper. Res. 99, 90-108.
Lindsey, R., 2004. Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. Transp. Sci. 38 (3), 293-314.
Lindsey, C.R. Van den Berg, V.A.C. Verhoef, E.T., 2012. Step tolling with bottleneck queuing congestion. J. Urb. Econ. 72, 46-59.
Liu, Y., Li, Y., Hu, L., 2018. Departure time and route choices in bottleneck equilibrium under risk and ambiguity. Transp. Res. Part B 117, 774-793.
Liu, Y., Nie, Y., Hall, J., 2015. A semi-analytical approach for solving the bottleneck model with general user heterogeneity. Transp. Res. Part B 71, 56-70.
Long, J., Szeto, W.Y., Gao, Z., Huang, H.J., Shi, Q., 2016. The nonlinear equation system approach to solving dynamic user optimal simultaneous route and departure time choice problems. Transp. Res. Part B 83, 179-206.
Long, J., Szeto, W.Y., Shi, Q., Gao, Z., Huang, H.J., 2015. A nonlinear equation system approach to the dynamic stochastic user equilibrium simultaneous route and departure time choice problem. Transportmetrica A 11(5), 388-419.
Ma, R., Ban, X.J., Szeto, W.Y., 2017. Emission modeling and pricing on single-destination dynamic traffic networks. Transp. Res. Part B 100, 255-283.
Mayeres, I., Ochelen, S., Proost, S., 1996. The marginal external costs of urban transport. Transp. Res. Part D 1 (2), 111-130.
Nagurney, A., 1999. Network Economics: A Variational Inequality Approach, the second and revised edition. Kluwer Academic Publishers, Dordrecht.
Pang, J.S., 1985. Asymmetric variational inequality problems over product sets: Applications and iterative methods. Math. Program. 31 (2), 206-219.

Penic, M.A., Upchurch, J., 1992. TRANSYT-7F: Enhancement for fuel consumption, pollution emissions, and user costs. Transp. Res. Rec. 1360, 104-111.

Ren, H.L., Xue, Y., Long, J.C., Gao, Z.Y., 2016. A single-step-toll equilibrium for the bottleneck model with dropped capacity. Transportmetrica B 4(2), 92-110.

Shafiei, S., Gu, Z., Saberi, M., 2018. Calibration and validation of a simulation-based dynamic traffic assignment model for a large-scale congested network. Simul. Model. Pract. Th. 86, 169-186.

Sharma, S., Mishra, S., 2013. Intelligent transportation systems-enabled optimal emission pricing models for reducing carbon footprints in a bimodal network. J. Intell. Transp. Syst. 17(1), 54-64.

Shi, Z.J., Shen, J., 2005. New inexact line search method for unconstrained optimization, J. Optim. Theory Appl. 127 (2), 425-446.

Small, K.A., 1982. The scheduling of consumer activities: Work trips. Am. Econ. Rev. 72 (3) 467-479.

Sumalee, A., Xu, W., 2011. First-best marginal cost toll for a traffic network with stochastic demand. Transp. Res. Part B 45(1), 41-59.

Tian, L.J., Yang, H., Huang, H.J., 2013. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. Transp. Res. Part E 54, 1-13.

van den Berg, V.A., Verhoef, E.T., 2016. Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity. Transp. Res. Part B 94, 43-60.

Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59 (2), 251-260.

Wu, W.X., Huang, H.J., 2015. An ordinary differential equation formulation of the bottleneck model with user heterogeneity. Transp. Res. Part B 81, 34-58.

Xiao, L.L., Huang H.J., Liu, R.H., 2015. Congestion behavior and tolls in a bottleneck model with stochastic capacity. Transp. Sci. 49 (1), 46-65.

Xiao, L.L., Liu, T.L., Huang, H.J., 2016. On the morning commute problem with carpooling behavior under parking space constraint. Transp. Res. Part B 91, 383-407.

Yang, H., Huang, H.J., 2005. Mathematical and Economic Theory of Road Pricing. Elsevier.

Yildirim, M.B., Hearn, D.W., 2005. A first best toll pricing framework for variable demand traffic assignment problems. Transp. Res. Part B 39(8), 659-678.

Yin, Y., Lawphongpanich, S., 2006. Internalizing emission externality on road networks. Transp. Res. Part D 11(4), 292-301.

Zhang, X.N., Huang, H.J., Zhang, H., 2008. Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city. Transp. Res. Part B 42, 38-56.

Zhang, X.N., Yang, H. Huang, H.J., 2011. Improving travel efficiency by parking permits distribution and trading. Transport. Res. Part B 45 1018-1034.

Zhu, T., Long, J., Liu, H., 2019. Optimal official work start times in activity-based bottleneck models with staggered work hours. Transportmetrica B 7(1), 657-683.