Please cite as:
Xiao L-L, Liu T-L, Huang H-J and Liu R (2020) Temporal-spatial allocation of bottleneck capacity for managing morning commute with carpool. Transportation Research Part B. In press.

Temporal-spatial allocation of bottleneck capacity for managing morning commute with carpool

Ling-Ling Xiao[a], Tian-Liang Liu[b,1], Hai-Jun Huang[b], Ronghui Liu[c]
[a]*School of Economics and Management, Beijing Jiaotong University, Beijing 100044, PR China*
[b]*MoE Key Laboratory of Complex System Analysis and Management Decision, School of Economics and Management, Beihang University, Beijing 100191, China*
[c]*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK*

Abstract

Carpooling as one of demand management measures is effective in reducing highway congestion. Recent studies have shown that an appropriate spatial allocation of bottleneck capacity between carpool lane and general-purpose lane can lead to peak-spread of the morning commuters' departure time and reduce the system's total trip cost. What is not clear however is whether temporal allocation of bottleneck capacity can also be effective and if so, what the impact would be, and furthermore what the combined effects of temporal-spatial allocation of bottleneck capacity would be. This paper investigates the impacts of a temporal allocation of bottleneck capacity, when carpool lane is available only within a reserved time window, and a joint temporal-spatial capacity allocation, on morning commute patterns. User equilibrium commute patterns are derived for both the temporal-only and the joint temporal-spatial capacity allocation schemes, along a highway corridor with two driving modes: solo driving and carpooling. The extra costs associated with carpooling are considered alongside of travel time and schedule delay costs. We identify three different cases representing the relative barriers and attractions of carpooling to commuters, and we show that the optimal capacity allocations are sensitive to the accurate estimation of the commuters' extra carpool cost. To assist in evaluating the difference between a non-optimal and the optimal temporal-spatial allocation schemes, we derive analytically the upper bounds on the efficiency loss and present numerical illustrations on how the upper bounds vary with the different operational and behavioral variables.

Keywords: morning commute; bottleneck model; carpool; spatial-temporal capacity allocation

[1]Corresponding author. Tel: +86-10-8231-7839; Email: liutianliang@buaa.edu.cn

1. Introduction

Morning commute problem was first introduced in Vickrey (1969) where a population of car commuters has to pass a single bottleneck to get to work at a desired time. A queue forms when the capacity constraint does not allow all users to pass the bottleneck at their desired time, which forces users to make the trade-off between low queuing time and low schedule penalties. Since queuing time represents a deadweight loss, completely free competition is a rather inefficient way to allocate the scarce bottleneck capacity. The classic remedy to this inefficiency is time-varying congestion pricing (Arnott et al., 1990; Hall, 2018). However, due to its many limitations, e.g., bringing new inequity between heterogeneous users, congestion pricing has not been implemented widely, and researchers and practitioners have turned to other alternatives especially those that do not require any direct payment from commuters. One such alternative is carpooling that conceptualizes as an arrangement where two or more people share the use of a privately owned car for a trip, and the passengers share the driver's expenses (Li et al., 2020). Since the concept of carpooling was first introduced in the 1960s, there has been significant body of studies on the commuter behavior towards carpooling, as well as on the efficient allocation of road spaces to carpool traffic (Brown, 2020). This paper contributes to the discussion on the latter issue, and more specifically on the optimal design of road capacity allocation between carpool and general traffic that respond to both temporal and spatial distribution of bottleneck congestion.

As a more efficient and sustainable mode of travel than solo-driving, the take-up of using carpooling however has been relatively low (Ferguson, 1997; Huang et al., 2000; Chan and Shaheen, 2012). The reasons for this can be classified into physical or psychological barriers, attitudes and perceptions, e.g. incompatible work schedules, need of independence and privacy, lack of convenience (Teal, 1987; Koppelman et al., 1993; Baldassare et al., 1998; Buliung et al., 2009). Instead, monetary cost savings resulting from equal sharing of mileage, fuel costs and parking charges are attractions influencing the intention to start carpooling (Yang and Huang, 1999). The barriers and attractions of using carpooling work together to make up the (aggregated) extra carpool cost which, depending on their relative influence on the commuters, further complicates the commute patterns. By considering the combined time-based and distance-based inconvenience cost and the out-of-pocket cost for people sharing a car, Liu and Li (2017) derived all possible morning commute patterns with commuters' three roles, i.e., solo driver, carpooling driver, and carpooling rider, on the basis of the Vickrey's bottleneck model. They found that the arrival order between solo drivers and carpooling participants significantly depends on the relative magnitude of inconvenience cost and out-of-pocket cost. Furthermore, they derived a time-varying system optimum (SO) toll and a flat carpool price to support a stable equilibrium with minimum system cost.

In addition to combining carpool program with congestion pricing as a way to stimulate carpooling, providing exclusive carpool lanes for high occupancy vehicles (HOV) is another popular stimulus program to directly improve carpoolers' travel time savings or reliability, and to reduce congestion, air pollution and parking supply (Menendez and Daganzo, 2007). Despite their attractiveness, HOV or carpool lanes have also drawn criticism over time. Dahlgren (1998) reported that changing a mixed flow lane to a HOV lane is more effective only in reducing congestion of HOV lane, whilst significant delay remains on the other mixed flow lanes. Guiliano et al. (1990) questioned whether HOV lanes create new demand for private car travel or they simply attract people from solo-driving to carpooling. Partly to address these concerns, there have been theoretical studies to examine the effects of introducing HOV lanes and integrating them with other demand management measures, such as congestion pricing and parking availability. For example, at a static user equilibrium (UE) setting, Yang and Huang (1999) showed that the optimal toll scheme with the operation of carpool lanes is significantly different from that without its operation. Qian and Zhang (2011) analyzed the morning commute patterns involving three modes, driving-alone, carpool, and light rail transit, and discussed the effects of changes in transit fare, road toll and fuel cost on the system performance. Xiao et al. (2016) derived the multimodal morning commute patterns with the consideration of constant extra carpool cost and parking availability, and investigated the optimal spatial allocation of bottleneck capacity between a general-purpose (GP) lane and a carpool lane. Xiao et al. (2019) further proposed two tradable parking permit schemes for managing the morning commute with multiple modes, transit, driving alone and carpool, considering the parking space constraint at destination. The four key studies above have all assumed that carpool lanes are present along the entire highway connecting the origin to the destination and are in operation during the entire (morning) commuting period. This leads to significant under-utilization of the highway capacity (Xiao et al., 2016).

To address this under-utilization issue of carpool lanes, two alternative improvement schemes have been proposed and put into practice (Varaiya, 2007). One is converting carpool lanes to high-occupancy toll (HOT) lanes which charge solo drivers to use the carpool lanes. This scheme has attracted much attention of scholars in economics and transportation research. For example, Konishi and Mun (2010) explored the welfare effects of converting carpool lanes to HOT lanes when commuters have different carpool organization costs. Liu et al. (2009) and Lou et al. (2011) investigated the optimal dynamic pricing strategies for HOT lanes. Considering a carpool ratio that rises with travel delay, Ma and Zhang (2017) analyzed the morning commute problem with dynamic carpool. Zhong et al. (2020) extended the work of Ma and Zhang (2017) to model commuters' mode choice between solo-driving and carpooling under the situations without and with a HOV lane, and further discussed the impact of converting the HOV lane into a HOT lane on the commute pattern[2].

The other approach is to operate carpool lanes on a time-limited basis, i.e. a lane is operated as a carpool lane only during part of the commute period (e.g., 7:30-9:30 am and 5:30-7:30 pm in Shenzhen, China), while at other times, the lane is operated as a GP lane (Stamos et al., 2012). However, to the best of our knowledge, there has been very little analysis on time-limited carpool lanes, with the exception of the work by Fosgerau (2011) who proposed a fast and slow lane scheme, where a fast lane is reserved for a prioritized group of commuters during part of the morning commute period. The scheme is equivalent to a coarse toll and produces a Pareto-improvement. However, in the work of Fosgerau (2011), the group of prioritized commuters is assumed to be fixed exogenously and thus morning commute patterns do not vary with different extra carpool cost. This is a very strong assumption as it ignores the extra carpool cost on determining commuters' mode choice decisions as mentioned before.

There are many more unanswered questions on the design and performance of time-limited carpooling schemes. How much advantage does a time-limited one gain over a full-time one? What should be the optimal temporal allocation of bottleneck capacity to carpoolers and solo drivers corresponding to different cases of extra carpool cost components? When considering all barriers and attractions of using carpooling, the commuters' extra carpool cost may consist of not only constant components but also (travel) time-varying ones, and each component may also be positive or negative. Could a joint temporal-spatial capacity allocation further improve the morning commute, and if so, how to determine the optimal temporal-spatial capacity allocation? Due to various practical reasons, it is difficult for the government to accurately estimate the commuters' extra carpool costs and implement the

[2] The major differences of Zhong et al. (2020) to this paper are summarized here. Firstly, it is assumed in Zhong et al. (2020) that the HOV lane with a fixed capacity is in operation during the entire commute period, and solo drivers may commute on the HOV lane if paying a fixed toll. In contrast, a joint temporal-spatial capacity allocation to carpool lane and GP lane is investigated in this paper, and thus solo drivers can always commute outside the time interval reserved for carpool purpose without any payment. Secondly, Zhong et al. (2020) focused on analyzing one single case that carpoolers pass the bottleneck at the center of the rush hour period and solo drivers commute at the two tails, assuming that the inconvenience cost due to carpool increases with the number of riders. In this paper, three typical cases representing the relative barriers and attractions of carpooling to commuters are identified, and the upper bounds on the inefficiency arising from non-optimal capacity allocations are derived analytically for different cases. The new insight and significance of the differences of this paper from Zhong et al. (2020) are that improving the efficiency of the traffic system with HOV lanes needs the government's delicacy management, otherwise the benefit from operating HOV lanes may be relatively low.

optimal temporal-spatial capacity allocation. Then, how about the upper bound on the inefficiency of implemented temporal-spatial capacity allocation schemes compared with the optimal one when only limited information about extra carpool cost components, such as range of change, is known a priori?

This paper fills the gaps in the literature on time-limited carpooling, by addressing some of the above questions. More specifically, this paper models an one-to-one highway corridor problem with two commute modes, solo driving and carpooling, where the highway capacity is shared by a carpool lane and a GP lane. A single Vickrey's bottleneck dynamic setting is employed in order to make the model tractable. All commuters are assumed to have identical work schedules and freely choose their commute modes for travel. The barriers and attractions of using carpooling are represented by an extra carpool cost (in addition to the common travel time and schedule delay costs), and are modeled as consisting of a constant component (to represent the distance-based or free-flow time dependent parts for this single OD pair case) and a component proportional to the queuing delay. The two cost components can each be either positive or negative, to reflect the barriers and attractions of using carpooling. The carpool lane is designated to operate within a reserved time window at the center of the rush hour period; at other times, the lane is available to all commuters. We examine the relative impact of the barriers and attractions (and the associated carpool cost components) on commuters' travel pattern, and on the optimal design of temporal-only, and joint temporal-spatial, carpooling schemes.

The main contribution and the key findings of this paper are summarized as follows. Firstly, when the whole highway is operated for GP purpose, we identify three typical cases representing the relative barriers and attractions of using carpooling to commuters, i.e., Case 1(a), where carpoolers pass the bottleneck at the center of the rush hour period and solo drivers commute at the two tails, Case 2(a), where solo drivers pass the bottleneck at the center of the rush hour period and carpoolers commute at the two tails, and Case 3(a) where all commuters choose solo driving.

Secondly, we present analytically the UE solutions of a temporal-only capacity allocation scheme (a time-limited carpooling scheme) for each of the above three cases. We show that, the impact of the reserved time window on the commute patterns for both Case 1(a) and Case 3(a) is similar to that of the single-step coarse toll scheme (Arnott et al., 1990; Laih, 1994; Laih, 2004; Lindsey et al., 2012; Xiao et al., 2011, 2012; Nie and Yin, 2013; Liu et al., 2015; Li et al., 2017; Xu et al., 2019), and that two important variables are the constant component of extra carpool cost and the queuing time cost for the solo drivers just passing the bottleneck at the starting/ending time of the time window: if the latter is larger, it leads to excess queuing delay; whilst if the former is larger, the scheme will lead to excess capacity waste.

Thirdly, we present a joint temporal-spatial capacity allocation scheme, where the system's total trip cost is minimal when the full bottleneck capacity is spatially divided into a carpool lane and a GP lane and the reserved time window for carpool purpose is optimally set according to an accurate estimation of the commuters' extra carpool cost.

Finally, for non-optimal temporal-spatial allocation schemes, we further derive analytically the upper bounds on the inefficiency, and numerical examples are presented to illustrate how the upper bounds vary with different operational and behavioral parameters, such as the spatial allocation ratio of highway capacity, queuing time cost for the solo drivers just passing the bottleneck at the starting/ending time of the time window, number of commuters in a carpooling vehicle and rush hour period on the highway.

The rest of this paper is organized as follows. The Vickrey's highway bottleneck model is extended by incorporating carpooling mode in Section 2. Sections 3 and 4, respectively, introduce a temporal-only bottleneck capacity allocation, where the highway is designated for carpool use at the center of the rush hour period and a joint temporal-spatial capacity allocation, and investigate their impacts on morning commute patterns. Section 5 bounds the inefficiency of non-optimal temporal-spatial allocation schemes compared with the optimal one. Numerical examples are presented in Section 6. Section 7 concludes the paper.

2. Bottleneck model with carpooling

In this section, we consider the Vickrey's highway bottleneck model with carpooling. (The main variables and notations are summarized in Appendix A.1). Let s be the capacity of the bottleneck. A fixed number of homogeneous commuters, N, travel on the highway and hope to arrive at destination at the work starting time or the desired arrival time, t^*. Solo driving and carpooling are two alternative modes for the commuters' travel. Let N_s and N_c be the number of vehicles traveling on the highway by solo driving and carpooling, respectively. Then, the total number of vehicles on the highway N_f is:

$N_f = N_s + N_c$, and the travel demand N is given, $N = N_s + mN_c$. Here the constant m denotes the average number of commuters in a carpooling vehicle and $m \geq 2$[3].

The travel time on the highway for commuters departing from home at time t consists of the free-flow travel time, T^f, and the queuing time at the bottleneck, $T^v(t)$. Without loss of generality, it is assumed throughout this paper that $T^f = 0$. Because the total demand is fixed, the commuters' mode-choice decisions depend only on the difference between the costs of carpooling and the costs of solo driving. Thus, according to the ADL model (Arnott et al., 1990), the trip costs for solo drivers and carpoolers departing from home at time t can be expressed as, respectively[4]

$$c_s(t) = \alpha T^v(t) + \beta(\text{time early}) + \gamma(\text{time late}), \quad (1)$$

$$c_c(t) = \alpha T^v(t) + \beta(\text{time early}) + \gamma(\text{time late}) + \Delta(t), \quad (2)$$

where subscripts c and s denote carpooling and solo driving, α, β and γ are positive scalars that generally satisfy $\beta < \alpha < \gamma$ (Small, 1982). Eq. (1) and Eq. (2) assume the same value of time for solo-driving and carpooling and keep other carpooling-related cost components in $\Delta(t)$. Denote $Q(t)$ as the length of queue at the bottleneck for vehicles departing from home at time t, and then we have $T^v(t) = Q(t)/s$.

In Eq. (2), $\Delta(t)$ is the extra carpool cost that the carpoolers have to encounter besides the common queuing time cost and the schedule delay cost. It aggregates the additional cost for individual occupants of carpools as compared with the solo drivers. The monetary cost saving (e.g., sharing of parking fees, fuel costs and highway tolls) and the barriers (e.g., lack of convenience, independence and privacy) are both possibly dependent on the shared travel distance and travel time for the carpoolers. For the convenience of analysis, we define $\Delta(t)$ as a linear function of the shared queuing time $T^v(t)$ by putting all other factors into its constant component, Δ_1, including those distance-based or free-flow time dependent ones. It follows:

$$\Delta(t) = \Delta_1 + \Delta_2 T^v(t). \quad (3)$$

[3] Carpooling occurs when two or more people ride in one car simultaneously. Teal (1987) reported that purely household carpools are overwhelmingly composed of two persons whilst external carpools average 2.63 members. For a passenger car with five seats, m may at most take the value of 5.

[4] Carpoolers may have different free-flow travel time from solo drivers due to the pick-up, drop-off and matching friction. The difference can be integrated into the constant component of the extra carpool cost, so that the model and analysis still apply.

Component coefficients Δ_1 and Δ_2 may be positive or negative, depending on the difference between the compensation (monetary cost saving or attractions) and the inconvenience cost (barriers)[5]. Although it might be negative in some cases (as will be discussed later), $\Delta(t)$ is referred to in this paper as "the extra carpool cost" for convenience. To better understand the aggregated extra carpool cost, several representative determinants for its component coefficients are listed in Table 1.

Table 1. Determinants for the component coefficients of the extra carpool cost.

Determinants		Δ_1	Δ_2
Inconvenience cost	Pick up and drop off	Positive	--
	Matching friction/efforts	Positive	--
	Independence and privacy	Positive	Positive
Compensation	Sharing of parking fees	Negative	--
	Sharing of fuel costs	--	Negative
	Sharing of highway tolls	Maybe negative	Maybe negative
	Social preference for carpooling	Maybe negative	Maybe negative

The equilibrium condition for commuters' joint choice of departure time and travel mode in the single bottleneck model is that no commuter can reduce his/her individual trip cost by changing decisions bilaterally. At equilibrium, all commuters must have the identical trip cost, and then we have $dc_i(t)/dt = 0$, $i = s, c$. Accordingly, the arrival rates at the bottleneck for solo-driving and carpooling vehicles arriving at destination before and after t^* are given by, respectively

$$r_s^1 = \frac{\alpha}{\alpha-\beta}s\,,\ r_s^2 = \frac{\alpha}{\alpha+\gamma}s\,,\ r_c^1 = \frac{\alpha+\Delta_2}{\alpha+\Delta_2-\beta}s\,,\ r_c^2 = \frac{\alpha+\Delta_2}{\alpha+\Delta_2+\gamma}s \tag{4}$$

To ensure that all arrival rates and the equilibrium trip cost can take positive values and there may exist an interior solution, it is assumed in this paper that $\Delta_2 > \beta - \alpha$ and $-\delta N/(sm) < \Delta_1 \le \delta N/s$, where $\delta = \beta\gamma/(\beta+\gamma)$.

[5] With the consideration of all attractions and barriers, both Δ_1 and Δ_2 may depend on the number of commuters in a carpooling vehicle, m. Were we to specify function forms of Δ_1 and Δ_2 with respect to m, e.g., $\Delta_1 = \Delta_{11} + \Delta_{12}(m-1)$ and $\Delta_2 = \Delta_{21} + \Delta_{22}(m-1)$, the introduction of more parameters would result in a more complex analysis. It is without much loss of generality to assume that both Δ_1 and Δ_2 are independent of m since the number of commuters in a carpooling vehicle is assumed to be constant in this paper.

Since the arrival order between solo-driving vehicles and carpooling vehicles is determined by the trade-off among the schedule delay, queuing delay, and extra carpool cost (if involved), it is natural at equilibrium that the group with the higher (lower) arrival rate before (after) t^* will pass the bottleneck at the time slots closer to the desired arrival time. From Eq. (4), the arrival order between solo-driving vehicles and carpooling vehicles can be classified by three different cases[6] (Please refer to Appendix A.2 for details), as specified in Table 2.

Case 1: when $0 < \Delta_1 \le \delta N/s$, (***a***) if $\beta - \alpha < \Delta_2 < \Delta_2^{u_1}$, the carpooling vehicles pass the bottleneck at the center of the rush hour period (as Figure 1(i)), or (***b***) if $\Delta_2^{u_1} \le \Delta_2$, all commuters choose solo driving.

Case 2: when $-\delta N/(sm) < \Delta_1 < 0$, (***a***) if $\Delta_2 > \Delta_2^{l_1}$, the solo-driving vehicles pass the bottleneck at the center of the rush hour period (as Figure 1(ii)), or (***b***) if $\beta - \alpha < \Delta_2 \le \Delta_2^{l_1}$, all commuters choose carpooling.

Case 3: when $\Delta_1 = 0$, (***a***) if $\Delta_2 > 0$, all commuters choose solo driving, or (***b***) if $\beta - \alpha < \Delta_2 < 0$, all commuters choose carpooling, or (***c***) if $\Delta_2 = 0$, the departures and arrivals of solo-driving vehicles and carpooling vehicles are mixed.

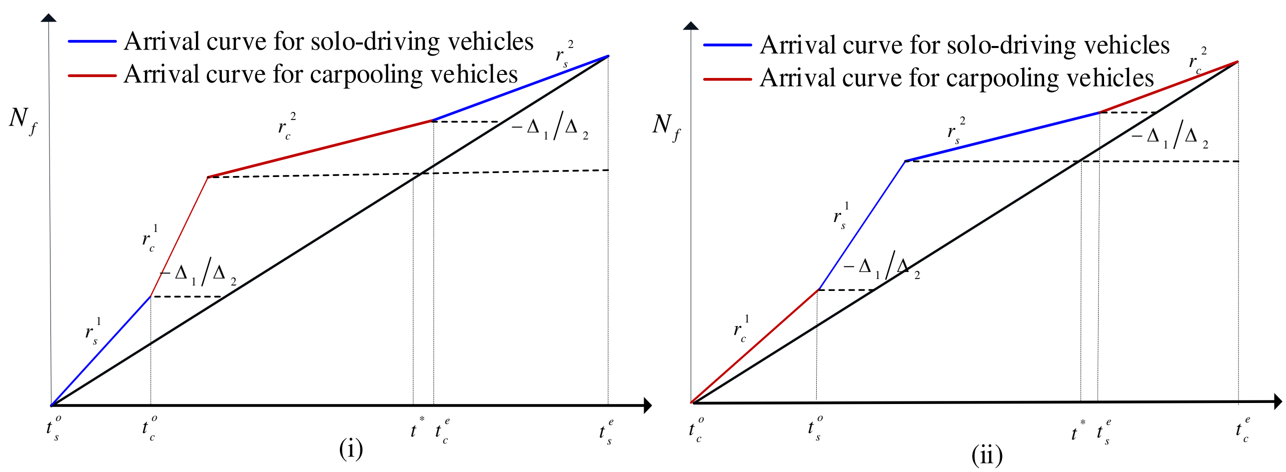

Fig. 1. Commute patterns for Case 1(a) and Case 2(a).

The equilibrium solutions for the above three cases, i.e., the equilibrium flow and the trip cost for each mode are derived in Appendix A.2. To separate different arrival time intervals, we denote t_i^o and t_i^e as the earliest and the latest arrival time at the bottleneck for role of commuters i, $i = s, c$. Figure 1

[6] The three cases here are similar to those discussed in Liu and Li (2017), where the detailed cost components due to ridesharing are defined. However, there is no clear empirical evidence on whether carpoolers tend to commute closer to the peak than solo drivers, or vice versa, although Brown (2020) found that people are more likely to share rides during peak periods.

shows the morning commute patterns for Case 1(a) and Case 2(a), respectively. The blue line is the arrival curve at the bottleneck for solo-driving vehicles, the red curve is that for carpooling vehicles and the black line is the departure curve from the bottleneck for all vehicles. At the critical time points, t_c^o and t_c^e in Figure 1(i) or t_s^o and t_s^e in Figure 1(ii), the queuing time for both roles is the same, and equals to $-\Delta_1/\Delta_2$ (the detailed derivation is presented in Appendix A.2). Using Eq. (3), at the critical time points with queuing time $-\Delta_1/\Delta_2$, the extra carpool cost equals to zero, i.e., $\Delta(t)=0$. Moreover, the commuters passing the bottleneck at the two tails endure a queuing time that is always smaller than $-\Delta_1/\Delta_2$. Thus, combining this with the conditions for each case, we can easily conclude that the commute pattern for carpooling corresponds to the situation with $\Delta(t)<0$, whilst that for solo driving corresponds to the situation with $\Delta(t)>0$.

In this section, the bottleneck-constrained highway is considered just as a GP lane with the fixed number of identical individuals for solo driving or carpooling. If $\Delta_1 \neq 0$, for both Case 1 and Case 2, all commuters are better off in the presence of carpooling. If $\Delta_1 = 0$, it is also clear for Case 3 that all commuters are not worse off in the presence of carpooling. In the following sections, we will introduce the temporal-only allocation as well as the joint temporal-spatial allocation of highway capacity for GP and carpool purposes to further improve the system efficiency for all cases.

Table 2. Cases of the arrival order between solo-driving vehicles and carpooling vehicles.

Arrival order		Δ_1		
		$0<\Delta_1 \le \delta N/s$	$-\delta N/(sm)<\Delta_1<0$	$\Delta_1=0$
Δ_2	$\Delta_2 \ge \Delta_2^{u_1}$	Case 1(b): Solo driving		
	$\beta-\alpha<\Delta_2<\Delta_2^{u_1}$	Case 1(a): Both modes and solo driving first for $\Delta_1 \in \left(0,(1-\dfrac{\beta}{\alpha})\dfrac{\delta N}{s}\right)$		
	$\Delta_2>\Delta_2^{l_1}$		Case 2(a): Both modes and carpooling first	
	$\beta-\alpha<\Delta_2 \le \Delta_2^{l_1}$		Case 2(b): Carpooling	
	$\Delta_2>0$			Case 3(a): Solo driving
	$\Delta_2=0$			Case 3(c): Mixed traffic
	$\beta-\alpha<\Delta_2<0$			Case 3(b): Carpooling

Note: $\Delta_2^{u_1}=-\Delta_1\alpha s/(\delta N)<0$ and $\Delta_2^{l_1}=-\alpha sm\Delta_1/\left(\Delta_1+\delta N/(sm)\right)>0$.

3. Temporal-only capacity allocation

Since not all commuters are worse off in the presence of carpooling in a completely free market mentioned above, it is natural to further stimulate carpooling by providing exclusive carpool lanes. However, if capacity is designated for carpool purpose at all times (during the peak period), the resultant equilibrium trip cost, i.e., $\Delta_1 + N/(ms)$, may be larger than that for Case 1(a) with no designation (See Appendix A.2). A possible remedy to such inefficiency is therefore to allocate bottleneck capacity for exclusive carpool use by time. That is to say, the entire morning commute period can be divided into a carpool period and a GP period. We define such a temporal-only bottleneck capacity allocation scheme as follows:

Definition 1. *The highway is designated for carpool use within a reserved time window,* $[t^+, t^-]$*, according to the queuing time cost for the solo drivers just passing the bottleneck at* t^+ *and* t^-*,* Δ_x*, whilst it is available for all commuters outside the time window during the morning commute period.*

The temporal-only bottleneck capacity allocation scheme allows a specified time window reserved for carpoolers according to Δ_x, which is the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^-. The time reservation is applied at t^+ before the work starting time and lifted at t^- after the work starting time (and its specific settings associated with a particular Δ_x for different cases will be discussed later). Accordingly, the commuters can be divided into three groups: the commuters passing the bottleneck before t^+ and after t^-, and the carpoolers passing the bottleneck inside $[t^+, t^-]$. Each selection of the parameters (t^+, t^-) associated with a particular Δ_x represents a particular temporal-only bottleneck capacity allocation scheme, and would also generate a corresponding morning commute pattern. However, it is unclear for each case of the arrival order between solo-driving vehicles and carpooling vehicles in Table 2 what the commute pattern will be like around the starting and ending points of the reserved time window, since the relationship between Δ_x and the commuters' extra carpool cost may be uncertain in practice.

3.1. Commute patterns for Case 1

As introduced in Section 2, if there is no time reservation for carpool purpose, the carpoolers pass the bottleneck at the center of the rush hour period and the solo drivers commute at the two tails for Case 1(a); all commuters choose solo driving for Case 1(b). But, once a temporal-only bottleneck capacity allocation scheme associated with Δ_x is implemented, the commute pattern for Case 1(b) would become similar to that for Case 1(a). A key question in analyzing the commute patterns under the temporal-only allocation scheme is how to deal with the discontinuity that takes place at the boundary of the reserved time window.

This discontinuity forces the commuters who arrive at the boundary of the reserved time window to have different travel delays, depending on whether or not they choose carpooling. The first carpooler who endures the extra carpool cost and has a lower travel delay must leave home later, in comparison with his/her immediate predecessor who escapes the time window by solo-driving. This implies that there must be a time period during which the arrival rate at the bottleneck is zero (e.g., the horizontal segment Δ_1/α for the scenario that $\Delta_x = \Delta_1$ as shown in Figure 2). In the spirit of the single step coarse toll, the discontinuity between the last person who commutes by carpooling and his/her immediate successor leads to a separated waiting (SW) behavioral assumption (Laih, 1994; Laih, 2004) about how commuters might respond to the temporal capacity allocation scheme. That is, the commuters who arrive at the bottleneck at the same time can use different waiting facilities, hence are allowed to have different travel delays.

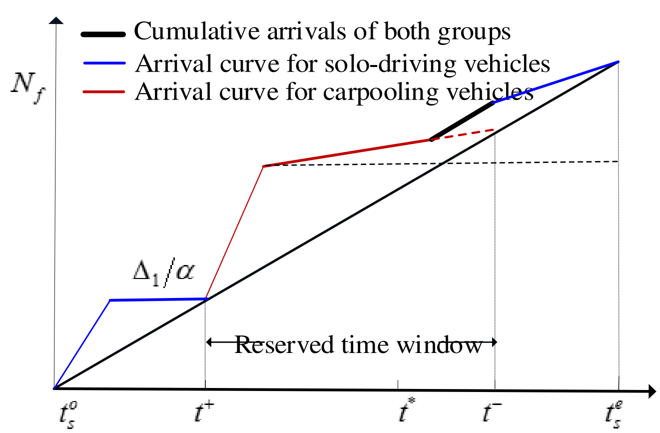

Fig. 2. Commute pattern under temporal-only capacity allocation for $\Delta_x = \Delta_1$

We first consider a special scenario, in which the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^- equals to the constant component of the extra carpool cost, i.e., $\Delta_x = \Delta_1$, such that the first carpoolers endure no queuing delay. Figure 2 depicts the commute pattern for this special scenario. Now, let us follow the SW assumption and explain how to deal with the discontinuity at the upper boundary of the reserved time window (t^+, t^-) associated with $\Delta_x = \Delta_1$. At equilibrium, the last carpoolers to arrive at the bottleneck before the carpooling reservation is lifted must have the same trip cost as the first solo driver to arrive at the bottleneck after the carpooling reservation is lifted. The latter must therefore incur the queuing time plus schedule delay costs that are Δ_1 higher than the former. This is impossible unless the solo drivers who choose to pass the bottleneck after the reserved time window can wait on a set of secondary lanes (on the separate lanes or shoulders, where the solo drivers can pull out of the traffic stream) without impeding the carpoolers who do pass the bottleneck in the reservation period.

In most of scenarios, the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^- may be not the constant component of the extra carpool cost, which renders the occurrence of a specific commute pattern depend on the relative values of Δ_x and Δ_1 as follows.

(i) Commute patterns with excess queuing delay for $\Delta_x > \Delta_1$

When the reserved period is over-compressed, i.e., the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^-, $\Delta_x = \alpha \overline{BD}$, is a little bit larger than Δ_1, after some arrivals of solo-driving vehicles there will be a period during which no one arrives at the bottleneck before t^+, but the excess queuing delay ($\overline{BD} - \overline{BC}$), due to the overestimation of the extra carpool cost, will be experienced by the first and the last carpoolers, as shown in Figure 3(i).

(ii) Commute patterns with excess capacity waste for $\Delta_x < \Delta_1$

When the reserved time window is longer (over-stretched), i.e., the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^-, $\Delta_x = \alpha \overline{BD}$, is less than Δ_1, then there is a time interval during which the bottleneck capacity is not fully utilized. This is referred to as "excess capacity waste". In other words, if the government underestimates the cost of carpooling, i.e., more people are supposed to carpool than they actually do, then the reserved time window is too long. Since the solo drivers cannot cross the bottleneck during the window, the bottleneck is un-utilized during part of the window. The reserved time window is set in terms of Δ_x, thus the capacity waste ($\overline{BC}-\overline{BD}$) occurs not only around the starting time but also at the ending point of the reserved time window, as shown in Figure 3(ii).

We now analyze the equilibrium solutions in Case 1. For any Δ_x, regardless of commute patterns with either excess capacity waste or excess queuing delay at t^+ and t^-, the rush hour period for arriving at the bottleneck can be determined by equalizing the trip costs of the first and last commuters for solo-driving or carpooling, and the starting time t^+ and the ending time t^- can be further determined by Δ_x (similar to the analyses in Lindsey et al. (2012) for the single-step coarse toll scheme and Nie and Yin (2013) for the travel permit case) as follows:

$$t_s^o = t^* - \frac{\delta N/m}{s\beta} + \frac{\Delta_x}{\beta m} - \frac{\Delta_x^+}{\beta},\ t_s^e = t^* + \frac{\delta N/m}{s\gamma} - \frac{\Delta_x}{m\gamma} + \frac{\Delta_x^+}{\gamma}, \tag{5a}$$

$$t_c^o = t_s^o + \Delta_x^+/\beta - \left(\Delta_x^+ - \Delta_1\right)/\alpha,\ t_c^e = t_s^e - \Delta_x^+/\gamma - \left(\Delta_x^+ - \Delta_1\right)/\alpha, \tag{5b}$$

$$t^+ = t_s^o + \Delta_x/\beta,\ t^- = t_s^e - \Delta_x/\gamma, \tag{6}$$

where $\Delta_x^+ = \max\{\Delta_1, \Delta_x\}$. It is easy to observe from Eq. (6) that Δ_x plays a very similar role to the single-step coarse toll in setting the reserved time window, and thus determines the number of solo drivers and the total number of vehicles on the carpool lane. The number of carpooling vehicles, the number of solo drivers and the equilibrium trip cost can be expressed as follows:

$$N_c = \left(t_c^e - t_c^o\right)s = \left(N - \Delta_x s/\delta\right)/m,\ N_s = \left(t_s^e - t^-\right)s + \left(t^+ - t_s^o\right)s = \Delta_x s/\delta \tag{7}$$

$$c^r = \delta N/(sm) + \left(m\Delta_x^+ - \Delta_x\right)/m, \tag{8}$$

which are independent of both α and Δ_2. Specifically for $\Delta_x = \Delta_1$, we easily get $c^r = \delta N_c/s + \Delta_1 = \delta N/(sm) + \Delta_1(m-1)/m$ from Eq. (8). At then, the first carpoolers passing the bottleneck within the reserved time window encounter the scenario with no excess queuing time and no capacity waste.

It should be noticed that, when $\Delta_x \geq -\alpha\Delta_1/\Delta_2$, where $-\alpha\Delta_1/\Delta_2 > \Delta_1$ due to the assumption that $\Delta_2 > \beta - \alpha$, the reserved time window is so short that the temporal-only capacity allocation does not impact the commuters' travel choice and corresponding commute pattern, and thus Eq. (8) would be changed to $c^r = \delta N/(ms) - (m-1)\alpha\Delta_1/(m\Delta_2)$, which is the equilibrium trip cost with no capacity allocation in Case 1 (See Appendix A.2). To make sense of the temporal-only capacity allocation, Δ_x

should take values from Δ_1 to $-\alpha\Delta_1/\Delta_2$ for the excess queuing delay scenario and from 0 to Δ_1 for the excess capacity waste scenario, respectively.

Proposition 1. For Case 1, the optimal temporal-only capacity allocation scheme for minimizing the equilibrium trip cost is that with $\Delta_x = \Delta_1$, i.e., there is no excess capacity waste and no excess queuing delay in the morning commute pattern.
Proof. See Appendix B.1.

Proposition 1 implies that for Case 1, the optimal temporal-only capacity allocation is very similar to the tolling/pricing control (c.f. Xiao et.al, 2011), since they both require determination of the reserved time window. If the time window is set inaccurately, the scenarios with excess capacity waste or excess queuing delay might occur in the morning commute.

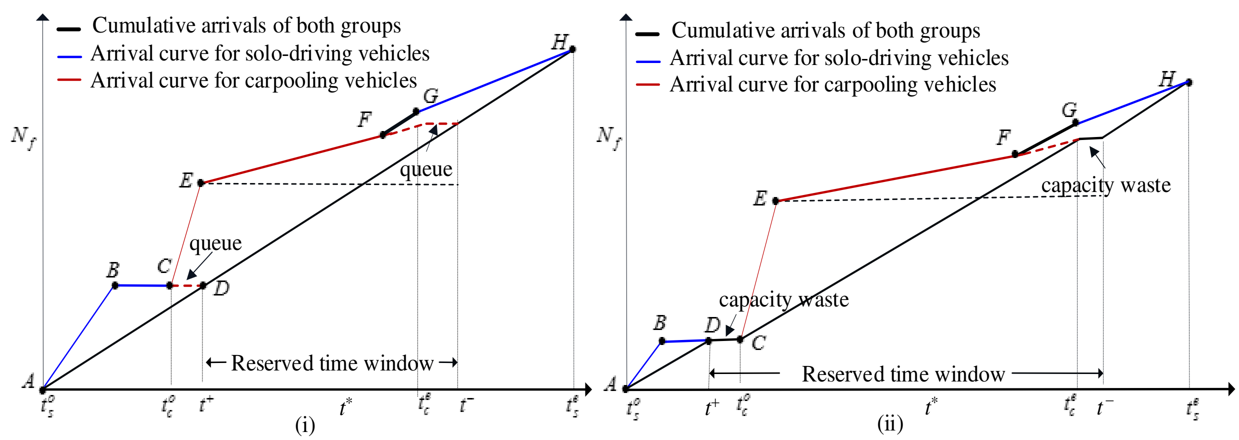

Fig. 3. Commute patterns under temporal-only capacity allocation for Case 1.

3.2. Commute patterns for Case 2

For Case 2(a), where $-\delta N/(sm) < \Delta_1 < 0$ and $\Delta_2^{l_1} < \Delta_2$, the carpoolers pass the bottleneck at the two tails and the solo drivers commute at the center of the rush hour period, if there is no time reservation for carpool purpose as introduced in Section 2. Under the temporal-only bottleneck capacity allocation scheme associated with Δ_x, all commuters can be divided into three groups, the commuters passing the highway bottleneck before t^+ and after t^- by solo driving or carpooling, and the carpoolers commuting inside $[t^+,t^-]$. Compared with the no time reservation in Section 2, the queue length at the bottleneck within the reserved time window $[t^+,t^-]$ will be reduced after the reservation is set, but at the time when the reservation starts, there exists only scenario with excess queuing delay (See Figure 4, where $\overline{BD}-\overline{BC}>0$). As done in Case 1, in this case we can derive the rush hour period, the starting time t^+ and the ending time t^-, which are presented in Appendix B.2, and the equilibrium trip cost

$$c^r = \Delta_1 + \delta N/(sm) + (\alpha\Delta_1/\Delta_2 + \Delta_x)(m-1)/m, \tag{9}$$

where $\Delta_x \in [-\alpha\Delta_1/\Delta_2, \bar{\Delta}_x]$ with $\bar{\Delta}_x = \delta N/s + (m-1)\alpha\Delta_1/\Delta_2 + m\Delta_1$. When $\Delta_x = \bar{\Delta}_x$, it is equivalent to the situation with no capacity allocation in Case 2(a) as analyzed in Section 2; and when $\Delta_x = -\alpha\Delta_1/\Delta_2$, all commuters choose carpooling with others, and thus the commute pattern would be changed to that with no capacity allocation in Case 2(b).

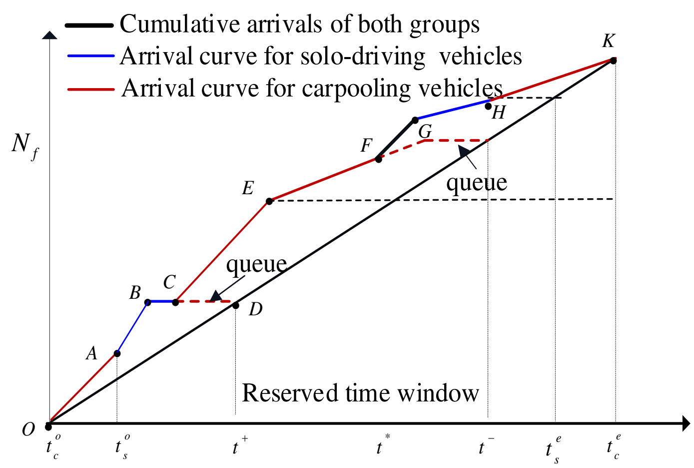

Fig. 4. Commute pattern under temporal-only capacity allocation for Case 2(a).

Proposition 2. For Case 2(a), the optimal temporal-only capacity allocation scheme for minimizing the equilibrium trip cost is that with $\Delta_x = -\alpha\Delta_1/\Delta_2$, where the time window reserved to the carpoolers is just long enough so that no one chooses solo driving.

Proof. Using Eq. (9) and $\Delta_2^{l_1} < \Delta_2$, we easily get $\mathrm{d}c^r/\mathrm{d}\Delta_x = (m-1)/m > 0$, and thus the minimum trip cost can be achieved at $\Delta_x = -\alpha\Delta_1/\Delta_2$. □

Proposition 2 suggests that with the optimal temporal-only capacity allocation for Case 2(a), all commuters would choose carpooling. The solo drivers who pass the bottleneck at t^+ or t^- would have the same queuing time as their immediate carpooling successors/predecessors (see Figure 4 and panel (ii) of Figure 1). For Case 2(b), where $-\delta N/(sm) < \Delta_1 < 0$ and $\beta - \alpha < \Delta_2 < \Delta_2^{l_1}$, all commuters freely choose carpooling with others under no time reservation scheme as discussed before. Thus, whether to set the reserved time window or not makes no difference to the commute pattern.

3.3. Commute patterns for Case 3

Since $\Delta_1 = 0$, it is easy to see that there is no need to implement the temporal capacity allocation for Case 3(b) and Case 3(c). However, for Case 3(a), all commuters would only choose solo driving if no temporal capacity allocation. Similar to the discussion as before, the carpool time reservation will be applied at t^+ before the work starting time and lifted at t^- after the work starting time. Under the temporal-only capacity allocation scheme associated with Δ_x, the commute pattern for Case 3(a) is similar to that for Case 1(a) with excess queuing delay (See Figure 3(i)), i.e., the carpoolers pass the bottleneck at the center of the reserved time window and the solo drivers travel before t^+ or later than t^-. Then, with the time reservation, for Case 3(a) where $\Delta_1 = 0$, the equilibrium trip cost can be expressed as follows:

$$c^r = \delta N/(sm) + (m-1)\Delta_x/m \text{, with } \Delta_x \in [0, \delta N/s]. \tag{10}$$

Proposition 3. For Case 3(a), the optimal temporal-only capacity allocation scheme for minimizing the equilibrium trip cost, is that $\Delta_x = 0$, i.e., the full bottleneck capacity is reserved to the carpoolers.

Proof. From Eq. (10), we easily get $\mathrm{d}c^r/\mathrm{d}\Delta_x = (m-1)/m > 0$ with $\Delta_x \in [0, \delta N/s]$. □

It can be observed from Propositions 1–3 that the optimal temporal-only capacity allocation scheme is to set $\Delta_x^o = \Delta_1$ for Case 1, $\Delta_x^o = -\alpha\Delta_1/\Delta_2$ for Case 2(a) and $\Delta_x^o = 0$ for Case 3(a), where superscript o denotes the optimal value of Δ_x. It suggests that, if the constant component of the extra carpool cost is positive so that Case 1 occurs, the optimal temporal-only allocation scheme would generate the commute pattern with no excess capacity waste and no excess queuing delay at both ends of the reserved time window, where the solo drivers still commute outside the time window. By giving commuting priority to the carpoolers during the rush hour period, the competition between solo drivers and carpoolers is reduced, and so is queuing delay endured by both carpoolers and trip cost. However, if the constant component of the extra carpool cost is non-positive such that Case 2 or Case 3 occurs, all commuters choose carpooling under the optimal temporal-only allocation scheme. Thus, the trip cost is reduced by squeezing all solo drivers out of market.

4. Joint temporal-spatial capacity allocation

4.1. Definition

In addition to re-assigning the peak demand by time on the carpool lane, re-assigning the peak demand by space on the highway is also important to the travel demand management. Following Definition 1 and the spatial capacity allocation scheme of bottleneck capacity discussed in Xiao et al. (2016), we here define a joint temporal-spatial capacity allocation scheme as follows.

Definition 2. *The highway is spatially allocated to a GP lane and a carpool lane, where the carpool lane is designated for carpool use within a reserved time window,* $[t^+, t^-]$, *according to the queuing time cost for the solo drivers who pas the bottleneck at* t^+ *and* t^-, Δ_x, *whilst it is available for all commuters outside the time window during the morning commute period.*

With this definition, we next use a two-lane equilibrium analysis to examine the effects of the temporal-spatial capacity allocation on the morning commute patterns for different cases.

4.2. Equilibrium analysis for different cases

Here, the highway bottleneck capacity is spatially allocated as a GP lane and a carpool lane, and the capacities on both GP and carpool lanes are assumed to be given and not endogenously affected by the composition of commuters. Superscripts g and h denote the GP lane and the carpool lane, respectively, and subscript i denote the specific role of commuters as defined before, solo driving or carpooling. Let N_i^g and N_i^h be the number of vehicles with role i $(i = s, c)$, traveling on the two parallel lanes of the highway, respectively, and $N^g = N_s^g + N_c^g$, $N^h = N_s^h + N_c^h$.

For simplicity, the capacity of the GP lane is further assumed to be $s^g = \theta s$, thus $s^h = (1-\theta)s$ is the capacity of the carpool lane. Here, the parameter $\theta \in [0,1]$ in fact determines the spatial capacity allocation of the bottleneck. According to the above assumptions, for a given θ, the carpooling time reservation would generate the commute patterns on the carpool lane as discussed in Section 3 for different cases. Define $c^g(N^g)$ and $c^h(N^h)$ as the respective equilibrium trip costs for commuters traveling on the GP lane and carpool lane. Clearly, we have $c^g(N^g) = c^h(N^h)$ at two-lane equilibrium.

Case 1: $0 < \Delta_1 \le \delta N/s$

According to the derivation of Eq. (7) and Appendix A.2, we easily get the numbers of solo driving vehicles on the GP lane and carpool lane in this case as follows:

$$N_s^g = -\frac{\Delta_1}{\Delta_2}\frac{\alpha\theta s}{\delta},\ N_s^h = \frac{\Delta_x}{\delta}(1-\theta)s, \tag{11}$$

where $\Delta_x \in [0, -\alpha\Delta_1/\Delta_2]$ determines the time reservation scheme.

Using $N = N_s^g + N_s^h + m(N_c^g + N_c^h)$ and the rush hour period for commuters passing the bottleneck is the same for both lanes, i.e., $N^g/s^g = N^h/s^h + (\Delta_x^+ - \Delta_x)/\delta$, with $\Delta_x^+ = \max\{\Delta_1, \Delta_x\}$. For the scenario that $\Delta_x < \Delta_1$, there exists the excess capacity waste during the time interval $(\Delta_1 - \Delta_x)/\delta$. Hence, we get the numbers of carpooling vehicles on the GP lane and carpool lane and the total vehicles on the highway as follows:

$$N_c^g = \theta\hat{N}_f - N_s^g,\ N_c^h = (1-\theta)\hat{N}_f - N_s^h \text{ and } N_f = \left(N + (m-1)\left(N_s^g + N_s^h\right)\right)/m, \tag{12}$$

where $\hat{N}_f = N_f + (\Delta_x^+ - \Delta_x)(1-\theta)s/\delta$.

To make $N_c^g > 0$ hold, it requires $\beta - \alpha < \Delta_2 < \Delta_2^{u_2} = -\alpha s\Delta_1/(\delta\hat{N}_f) = \dfrac{-\Delta_1\alpha(\theta + m(1-\theta))}{\delta N/s + (1-\theta)(m\Delta_x^+ - \Delta_x)}$ from Eqs. (11)–(12) and $\hat{N}_f \le N$. Then, the equilibrium trip cost for Case 1(a) can be expressed as follows:

$$c^r = \frac{\delta\hat{N}_f}{s} = \frac{\delta N}{ms} + \frac{(m-1)}{m}\left(\Delta_x(1-\theta) - \frac{\Delta_1}{\Delta_2}\alpha\theta\right) + (\Delta_x^+ - \Delta_x)(1-\theta), \tag{13}$$

where $\Delta_x^+ = \max\{\Delta_1, \Delta_x\}$. Otherwise, $N_c^g = 0$ with $\Delta_2 \ge \Delta_2^{u_2}$. According to Eq. (12), we get $N_f = \left(N + (m-1)N_s^h + (m-1)\theta(1-\theta)s(\Delta_x^+ - \Delta_x)/\delta\right)/\left(\theta + m(1-\theta)\right)$. The equilibrium trip cost for Case 1(b) can be expressed as follows:

$$c^r = \frac{\delta N}{s\left(\theta + m\left(1-\theta\right)\right)} + \frac{\left(1-\theta\right)}{\theta + m(1-\theta)}\left(\left(m-1\right)\Delta_x + m\left(\Delta_x^+ - \Delta_x\right)\right), \tag{14}$$

which is independent of Δ_2.

To sum up, in this section, under the temporal-spatial capacity allocation with given Δ_x and θ, the new condition for Case 1(a), where both modes are used on each lane as defined before, is such that $\beta - \alpha < \Delta_2 < \Delta_2^{u_2}$, whilst that for Case 1(b), where only solo drivers commute on the GP lane but both modes are used on the carpool lane, is such that $\Delta_2 \geq \Delta_2^{u_2}$.

Proposition 4. Under the temporal-spatial capacity allocation scheme with given Δ_x and θ, for the excess queuing scenario with $\Delta_x \in \left[\Delta_1, -\alpha\Delta_1/\Delta_2\right]$, the equilibrium trip cost increases with θ, i.e., $\partial c^r/\partial\theta \geq 0$. For the excess capacity waste scenario, if $\alpha/\beta \leq m$, the equilibrium trip cost always increases with θ for Case 1(a), i.e., $\partial c^r/\partial\theta > 0$ for $\Delta_x \in \left[0, \Delta_1\right]$; otherwise, the equilibrium trip cost may only do for $\Delta_x \in \left[\Delta_x^{\#}, \Delta_1\right]$, whilst decreasing with θ for $\Delta_x \in \left[0, \Delta_x^{\#}\right)$. Here, $\Delta_x^{\#} = \left(m - \left(m-1\right)\alpha/\left(-\Delta_2\right)\right)\Delta_1$ for Case 1(a) and $\Delta_x^{\#} = m\Delta_1 - \left(m-1\right)\delta N/s$ for Case 1(b).

Proof. See Appendix C.1.

Proposition 4 suggests that, the equilibrium trip cost for the excess capacity waste scenario does not always increase with the spatial allocation ratio θ. If the reserved time window for carpool use is non-optimally set, the spatially allocating the full bottleneck capacity to the carpool lane is not necessarily optimal and it depends on the condition $\alpha/\beta \leq m$. The implication of the condition can be understood from three aspects. Firstly, it is clear that when m is large, many people per vehicle benefit from the carpool lane. Secondly, the arrival rates and corresponding total travel delay for early-arrival commuters are both decreasing with α/β from Eq. (4). This suggests that the commuters also benefit from the carpool lane when α/β is small enough. Thirdly, from the proof of Proposition 4, the condition $\alpha/\beta \leq m$ together with the model assumption $\Delta_2 > \beta - \alpha$ can ensure $\Delta_2 > -\alpha\left(m-1\right)/m$ and thus $\partial c^r/\partial\theta > 0$ for Case 1(a), even when $\Delta_x = 0$ such that the maximal capacity waste would occur. Accordingly with this sufficient condition, at optimum, the full bottleneck capacity should be allocated to the carpool lane for Case 1(a). Otherwise, it would happen only if the reserved time window for carpool use is optimally set, i.e., $\Delta_x = \Delta_1$ from Proposition 1.

To sum up, it can be concluded from Proposition 4, together with Proposition 1, that the optimal temporal-spatial capacity allocation for minimizing the equilibrium trip cost for Case 1 can be achieved at $\theta = 0$ and $\Delta_x = \Delta_1$. That is, at optimum, the full bottleneck capacity is allocated to the HOV lane and there exists no excess capacity waste and no excess queuing delay in the commute pattern. Hence, for Case 1, the equilibrium trip cost with the optimal temporal-spatial capacity allocation is

$$c^* = \delta N/(ms) + (m-1)\Delta_1/m . \tag{15}$$

Case 2: $-\delta N/(ms) < \Delta_1 < 0$

For Case 2(a), the carpoolers pass the bottleneck at the two tails and the solo drivers commute at the center of the rush hour period on the GP lane. Under temporal-spatial bottleneck capacity allocation scheme, the numbers of solo driving vehicles, carpooling vehicles and the total vehicles on the highway can be derived as given in Appendix C.2. and thus the equilibrium trip cost can be expressed as follows:

$$c^r = \Delta_1 + \frac{\delta N_f}{s} = \Delta_1 + \frac{\delta N}{\left(\theta + m(1-\theta)\right)s} + \frac{(m-1)}{\theta + m(1-\theta)}\left(\left(\Delta_2 + \alpha\right)\frac{\Delta_1}{\Delta_2} + (1-\theta)\left(\Delta_x - \Delta_1\right)\right). \tag{16}$$

where $\Delta_x \in [-\alpha\Delta_1/\Delta_2, \bar{\Delta}_x]$, with $\bar{\Delta}_x = \delta N/s + (m-1)\alpha\Delta_1/\Delta_2 + m\Delta_1$. The new condition for Case 2(a) is that $\Delta_2 > \Delta_2^{l_2} = -\alpha\Delta_1\left(m + (m-1)(1-\theta)\right)/\left(\delta N/s + m\Delta_1 + (m-1)(1-\theta)\Delta_x\right)$, to make both modes be used on each lane.

Proposition 5. Under the temporal-spatial capacity allocation scheme for Case 2(a), where $-\delta N/(ms) < \Delta_1 < 0$ and $\Delta_2 > \Delta_2^{l_2}$, the equilibrium trip cost increases with θ, i.e., $\partial c^r/\partial\theta \geq 0$.

Proof. See Appendix C.3.

Proposition 5 implies that, the equilibrium trip cost for Case 2(a) increases with the spatial allocation ratio θ and thus it is optimal to spatially allocate the full bottleneck capacity to the carpool lane. If $\beta - \alpha < \Delta_2 \leq \Delta_2^{l_2}$, then $N_s^g = 0$ according to Appendix C.2. That is, all commuters choose carpooling on the GP lane, which is the commute pattern for Case 2(b). Thus, whether to set the reserved time window or not makes no difference to the commute pattern. Together with Propositions 2 and 5, it can be concluded for Case 2 that the optimal temporal-spatial capacity allocation for minimizing the equilibrium trip cost can be achieved at $\theta = 0$ and $\Delta_x = -\alpha\Delta_1/\Delta_2$. Hence, the equilibrium trip cost with the optimal temporal-spatial capacity allocation for Case 2 is $c^* = \delta N/(ms) + \Delta_1$.

Case 3: $\Delta_1 = 0$

For Case 3(a), where $\Delta_2 > 0$, the solo driving is a better choice than the carpooling without carpooling time reservation, then the number of carpoolers on the GP lane is zero, i.e., $N_c^g = 0$. Under the temporal-spatial capacity allocation, the commute pattern for Case 3(a) is reduced to that for Case 1(b) with the excess queuing delay (for details in Appendix C.4). Substituting $\Delta_1 = 0$ into Eq. (14), the equilibrium trip cost can be rewritten as follows:

$$c^r = \frac{\delta N}{\left(\theta + m(1-\theta)\right)s} + \frac{(m-1)(1-\theta)\Delta_x}{\theta + m(1-\theta)}, \tag{17}$$

where $\Delta_x \in [0, \delta N/s]$.

Proposition 6. Under the temporal-spatial capacity allocation scheme for Case 3(a), the equilibrium trip cost increases with θ, i.e., $\partial c^r/\partial\theta \geq 0$.

Proof. See Appendix C.5.

Similar to Proposition 5, Proposition 6 suggests that, the equilibrium trip cost for Case 3(a) also increases with the spatial allocation ratio θ and is optimal to spatially allocate the full bottleneck capacity to the carpool lane. For Case 3(b), where $\beta - \alpha < \Delta_2 < 0$, all commuters choose carpooling with others on the highway. Thus, whether to set the reserved time window or not makes no difference to the commute pattern. For Case 3(c), where $\Delta_2 = 0$, all commuters are indifferent to the arrival time. Although the equilibrium solutions are generally not unique for Case 3(c), it is always optimal to implement the temporal-spatial capacity allocation to discourage all solo drivers' travel. Hence, together with Proposition 6, it can be concluded for Case 3 that the optimal temporal-spatial capacity allocation for minimizing the equilibrium trip cost can be achieved at $\theta = 0$ and $\Delta_x = 0$, i.e., the full bottleneck capacity is allocated to the HOV lane and the entire peak period is reserved for carpool use. The optimal trip cost is $c^* = \delta N/(ms)$.

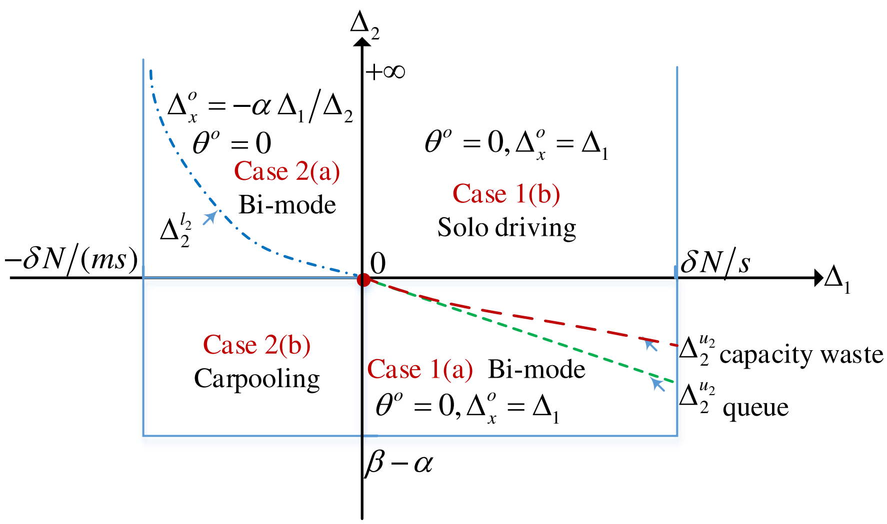

Fig. 5. Optimal temporal-spatial capacity allocation with (θ^o, Δ_x^o) for each case with $N = 6000\text{(person)}$, $s = 3000\text{(veh/h)}$, $\alpha = 6.4\text{(\$/h)}$, $\beta = 3.9\text{(\$/h)}$, $\gamma = 15.21\text{(\$/h)}$, $m = 2\text{(person/veh)}$, $\theta = 0.5$ and $\Delta_x = 1\text{(\$)}$.

Till now, we have derived all required conditions and critical values to distinguish the commute patterns for each case, which are summarized in Figure 5 together with the corresponding optimal temporal-spatial capacity allocation with (θ^o, Δ_x^o). In the next section, we will focus on bounding the inefficiency brought by any non-optimal temporal-spatial allocation for different cases.

5. Bounding the inefficiency of joint temporal-spatial allocation

Section 4 derives possible morning commute patterns and the optimal allocation, for different combinations of Δ_1 and Δ_2, under a joint temporal-spatial allocation scheme. In practice, it is difficult for the government to accurately estimate the commuters' extra carpool cost, and without which the implemented temporal-spatial capacity allocations would be non-optimal and inefficient. In this section, we derive the upper bounds on the inefficiency when only limited information about extra carpool cost components Δ_1 and Δ_2, such as their ranges for different cases, is known a priori.

Lemma 1 shows how the equilibrium trip cost c^r varies with Δ_1 and Δ_2 for each case discussed in Section 4 (Proofs can be found in Appendix D.1), which are useful for analyzing and bounding the inefficiency of a temporal-spatial allocation.

Lemma 1. For Case 1 and Case 2(a), the equilibrium trip cost increases with Δ_1 and Δ_2, i.e., $\partial c^r / \partial \Delta_2 \geq 0$ and $\partial c^r / \partial \Delta_1 \geq 0$, whilst for Case 3(a) it is independent of Δ_2, i.e., $\partial c^r / \partial \Delta_2 = 0$.

Compared with the optimal one corresponding to each case, the inefficiency of a temporal-spatial allocation scheme can be defined as

$$\rho = \frac{c^r}{c^*}, \tag{18}$$

where c^* is the equilibrium trip cost under the optimal temporal-spatial capacity allocation scheme. It clearly holds that $\rho \geq 1$. With the help of Lemma 1, we next analyze when the worst-case equilibrium trip costs would happen for different cases with the changes of Δ_1 and Δ_2, and then derive the corresponding upper bounds on the inefficiency arising from the non-optimal allocation, $\bar{\rho}$.

5.1 Specific bound for Case 1

Substituting Eqs. (13)–(15) into Eq. (18), we easily get

$$\rho = \frac{\delta N/(ms) + (m-1)\left(\Delta_x(1-\theta) - \dfrac{\Delta_1}{\Delta_2}\alpha\theta\right) \Big/ m + \left(\Delta_x^+ - \Delta_x\right)(1-\theta)}{\delta N/(ms) + (m-1)\Delta_1/m}, \tag{19}$$

for Case 1(a), where $0 < \Delta_1 \leq \delta N/s$ and $\beta - \alpha < \Delta_2 < \Delta_2^{u_2}$, and

$$\rho = \frac{\left(\delta N/s + (1-\theta)\left((m-1)\Delta_x + m\left(\Delta_x^+ - \Delta_x\right)\right)\right) \Big/ \left(\theta + m(1-\theta)\right)}{\delta N/(ms) + (m-1)\Delta_1/m}, \tag{20}$$

for Case 1(b), where $0 < \Delta_1 \leq \delta N/s$ and $\Delta_2^{u_2} \leq \Delta_2 < 0$.

Obviously, given a non-optimal temporal-spatial allocation scheme with Δ_x and θ, the inefficiency values in Eqs. (19) and (20) both depend on the specific values of Δ_1 and Δ_2. Since the required conditions of Δ_1 and Δ_2 for Case 1 are known till now, we can derive the upper bounds on the inefficiency arising from the non-optimal allocation as given in Proposition 7.

Proposition 7. Under a non-optimal temporal-spatial allocation scheme with given Δ_x and θ, the commute pattern can be either with excess queuing delay or with excess capacity waste for Case 1. The upper bound on the inefficiency for the excess queuing delay scenario,

$$\bar{\rho} = \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1\to 0^+} = \frac{\delta N/s + (m-1)(1-\theta)\Delta_x}{\big(\theta + m(1-\theta)\big)\delta N/(ms)}, \tag{21}$$

where $\Delta_x \in [0, \delta N/s]$. For the excess capacity waste scenario,

$$\bar{\rho} = \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s} = \frac{\big(1+(1-\theta)m\big)\delta N/s - (1-\theta)\Delta_x}{\big(\theta + m(1-\theta)\big)(\delta N/s)}, \tag{22}$$

where $\Delta_x \in [\max(0, \min(\hat{\Delta}_x, \delta N/s)), \delta N/s]$, or

$$\bar{\rho} = \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\Delta_x} = \frac{\delta N/s + (1-\theta)(m-1)\Delta_x}{\big(\theta + m(1-\theta)\big)\big(\delta N/(ms) + (m-1)\Delta_x/m\big)}, \tag{23}$$

where $\Delta_x \in [0, \max(0, \min(\hat{\Delta}_x, \delta N/s))]$ with $\hat{\Delta}_x = \dfrac{(m\theta - 1)\delta N/s}{(m-1)(1-\theta)}$ and $\Delta_2^{u_2} = \dfrac{-\Delta_1\alpha(\theta + m(1-\theta))}{\delta N/s + (1-\theta)(m\Delta_1 - \Delta_x)}$.

Proof. See Appendix D.2.

Eqs. (21)–(23) are all obtained for $\Delta_2 = \Delta_2^{u_2}$, through which it just happens that all commuters on the GP lane choose to drive alone for both Case 1(a) and Case 1(b). Further, it is required that $\Delta_1 \to 0^+$ for the excess queuing delay scenario where $\Delta_x \geq \Delta_1$, suggesting that the temporal capacity allocation should be optimal at $\Delta_x = 0$. However, the government estimates a nonzero extra carpool cost, which leads to the worst-case equilibrium trip cost, $c^r = \big(\delta N/s + (m-1)(1-\theta)\Delta_x\big)/\big(\theta + m(1-\theta)\big)$, as compared with the optimal one, $c^* = \delta N/(ms)$.

For the excess capacity waste scenario where $\Delta_x \leq \Delta_1$, the situation is more complicated. If spatial capacity ratio $\theta \leq 1/m$, then $\hat{\Delta}_x \leq 0$ holds. Thus, Eq. (22) would work for any $\Delta_x \in [0, \delta N/s]$, and the worst-case equilibrium trip cost is associated with $\Delta_1 = \delta N/s$, suggesting that there should be no time window reserved for carpool purpose at optimum. Unfortunately, the government actually set a carpool time reservation according to Δ_x although no commuter would like to choose carpooling. If spatial capacity ratio $\theta \geq m/(2m-1)$, thus $\hat{\Delta}_x \geq \delta N/s$ holds and Eq. (23) would work for any $\Delta_x \in [0, \delta N/s]$, and the worst-case equilibrium trip cost and the optimal one are $c^r = \left(\delta N/s + (m-1)(1-\theta)\Delta_x\right)/\left(\theta + m(1-\theta)\right)$ and $c^* = \delta N/(ms) + (m-1)\Delta_x/m$, respectively, which are both associated with $\Delta_1 = \Delta_x$. Otherwise, Eq. (22) happens to work for $\Delta_x \in [\hat{\Delta}_x, \delta N/s]$ whilst Eq. (23) does for $\Delta_x \in [0, \hat{\Delta}_x]$.

Remark 1. All upper bounds in Eqs. (21)–(23) can be calculated once the values of parameters θ, Δ_x, m and $\delta N/s$ are specifically given. They can be relaxed by the expressions only with respect to m, if the values of θ and Δ_x chosen by the government are a priori unknown except that the scenario is with excess queuing delay or with excess capacity waste. For example, the upper bound on the inefficiency for the excess queuing scenario, $\bar{\rho}$ in Eq. (21), can be further relaxed by Δ_x to reach the maximum, m, when $\Delta_x = \delta N/s$. It means that the worst-case equilibrium trip cost, here $c^r = \delta N/s$, happens when there is no time reservation for carpool use (i.e. all commuters have to choose solo driving). The reader can refer to Appendix D.2 for details. We will numerically discuss the respective impacts of these parameters on the bounds in Section 6.

5.2 Specific bounds for other cases

For Case 2(a), using Eq. (16) and $c^* = \delta N/(ms) + \Delta_1$, we get

$$\rho = \frac{\Delta_1 + \dfrac{\delta N}{\left(\theta + m(1-\theta)\right)s} + \dfrac{(m-1)}{\theta + m(1-\theta)}\left(\left(\Delta_2 + \alpha\right)\dfrac{\Delta_1}{\Delta_2} + (1-\theta)\left(\Delta_x - \Delta_1\right)\right)}{\dfrac{\delta N}{ms} + \Delta_1}, \text{ for } \Delta_2^{l_2} < \Delta_2. \quad (24)$$

Given a non-optimal temporal-spatial allocation scheme with Δ_x and θ, the inefficiency value in Eq. (24) depends on the specific values of Δ_1 and Δ_2. We can similarly derive the upper bound on the inefficiency arising from the non-optimal allocation according to the required conditions of Δ_1 and Δ_2 for Case 2(a), which is given in Proposition 8.

Proposition 8. Under a non-optimal temporal-spatial allocation scheme with given Δ_x and θ, there exists only commute pattern with excess queuing delay for Case 2(a), and thus the upper bound on the inefficiency is

$$\rho \le \bar{\rho} = \rho\Big|_{\Delta_2 \to +\infty, \Delta_1 \to -\delta N/(sm)} \to +\infty, \text{ for } \Delta_x \in [0, \delta N/s], \tag{25}$$

Proof. See Appendix D.3.

Proposition 8 shows the worst-case inefficiency arising from the non-optimal temporal-spatial allocation tends to be the infinity for Case 2(a). This happens to be a very extreme case, where all commuters are unwilling to choose carpooling due to $\Delta_2 \to +\infty$ and the optimal equilibrium trip cost approaches zero since $\Delta_1 \to -\delta N/(sm)$.

Recall that the commute pattern for Case 3(a) is similar to that for Case 1(a) with excess queuing delay. Thus, substituting $\Delta_1 = 0$ into Eq. (16), we can easily get the same formulation of the upper bound on the inefficiency arising from the non-optimal temporal-spatial allocation as Eq. (21).

6. Numerical examples

In this section, we adopt the shadow values of common travel time, early arrival time, and late arrival time from Arnott et al. (1990), i.e., $\alpha = 6.4$ (\$/h), $\beta = 3.9$ (\$/h), and $\gamma = 15.21$ (\$/h), and consider the situation with $N = 6000$ (person), $s = 3000$ (veh/h), and $m = 2$ (person/veh). This yields $\Delta_2 > \beta - \alpha = -2.5$ and $-3.10 \approx -\delta N/(sm) < \Delta_1 \le \delta N/s \approx 6.21$ according to the model assumptions, i.e. both Δ_1 and Δ_2 can be negative or positive. Here, we limit the values of Δ_1 from -0.5 (\$) to 2 (\$) and the values of Δ_2 from -2 (\$/h) to 2 (\$/h)[7]. With the input parameters except for those specially mentioned, we next numerically discuss the equilibrium trip cost, the inefficiency arising from a joint

[7] Values of parameters Δ_1 and Δ_2 can be aggregated by referring to the empirical studies, since they both depend on the difference between the compensation (monetary cost saving or attractions) and the inconvenience cost (barriers). For example, Li et al. (2007) reported that the typical two-person carpool requires an average of five minutes to form; the average commute time is about 18 minutes according to the 2001 National Household Transportation Survey (NHTS) (Jacobson and King, 2009); the out-of-pocket cost per mile is estimated to be about 0.37\$ in 2016 as reported by the American Automobile Association (AAA) (Liu and Li, 2017); Weinberger et al. (2010) reported that it would cost 8\$ to park at CBD (outside the ParkSmart pilot areas) for a half day in New York City; Lyft rideshare trips were about \$2 cheaper on average compared to regular ride-hail trips (Brown, 2020); Monchambert (2020) found that the value of travel time for a carpool driver is on average 13% higher than that when driving alone, and carpool passengers incur a discomfort cost of on average 4.5 euros per extra passenger in the same vehicle.

temporal-spatial capacity allocation and its upper bounds for different cases with the change of Δ_1 and Δ_2 or that of θ and Δ_x, respectively.

6.1 Equilibrium trip cost

To illustrate the equilibrium trip cost with the change of θ and Δ_x, we further set $\Delta_1 = 1$ (\$) and $\Delta_2 = -2$ (\$/h) in Figure 6(i) for Case 1, $\Delta_1 = -0.5$ (\$) and $\Delta_2 = 2$ (\$/h) in Figure 6(ii) for Case 2(a), and $\Delta_1 = 0$ (\$) and $\Delta_2 = 2$ (\$/h) in Figure 6(iii) for Case 3(a). Clearly, Figure 6 shows that for any given Δ_x, the equilibrium trip costs always increase with the spatial allocation ratio θ. This implies that the more capacity is allocated to the GP lane, the higher the equilibrium trip cost, for these three cases. For Case 1 in Figure 6(i), the optimal equilibrium trip cost can be achieved at $\Delta_x = \Delta_1 = 1$ (\$) and $\theta = 0$, which is consistent with Proposition 1 and Proposition 4. That is to say, for minimizing the equilibrium trip cost for a given combination of Δ_1 and Δ_2, the full bottleneck capacity should be spatially allocated to the HOV lane and the reserved time window for carpooling should be set to lead to the morning pattern with no excess queue and no excess capacity waste. As depicted in Figure 6(ii), the equilibrium trip cost for Case 3(a) happens to be minimized at $\theta = 0$ and $\Delta_x = -\alpha\Delta_1/\Delta_2 = 1.6$ (\$) for Case 2(a), which is consistent with Proposition 2 and Proposition 5. Figure 6(iii) shows that the optimal equilibrium trip cost for Case 3(a) is achieved at $\theta = 0$ and $\Delta_x = 0$ (\$), which is consistent with Propositions 3 and 6.

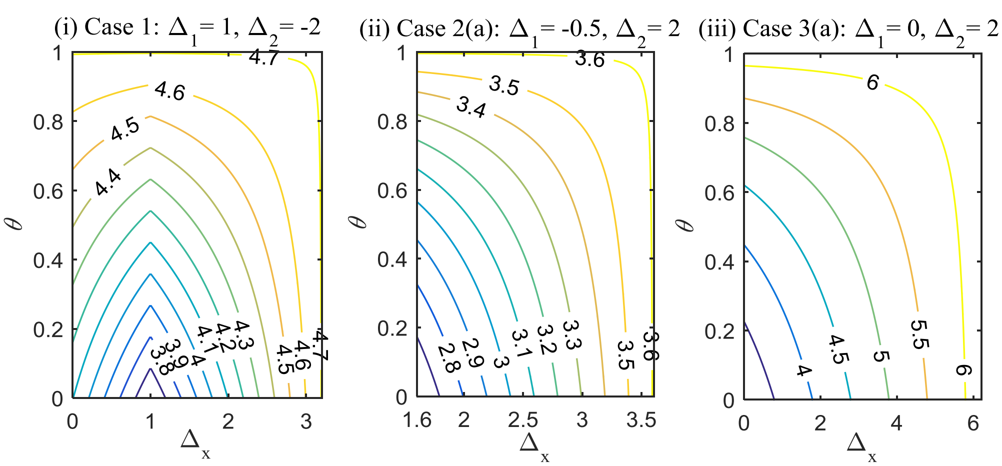

Fig. 6. Indifference curves of equilibrium trip cost with the change of θ and Δ_x.

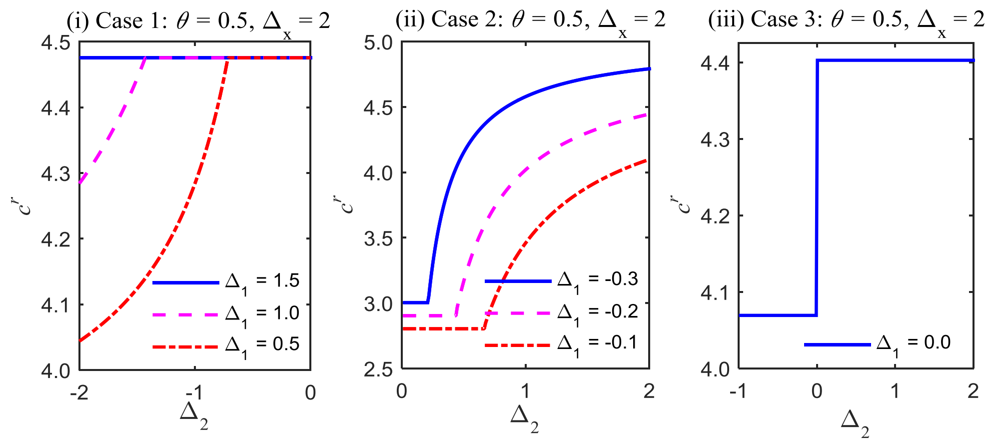

Fig. 7. Equilibrium trip cost varying with Δ_2 and Δ_1.

Given the temporal-spatial capacity allocation scheme with $\theta = 0.5$ and $\Delta_x = 2$, Figure 7 shows that the equilibrium trip cost is a step function but always increases with Δ_2 or Δ_1 for Case 1 and Case 2, respectively, whilst keeping piece-wise constant with the change of Δ_2. These are consistent with Lemma 1. In Figure 7(i), as Δ_2 increases, the commute pattern consecutively goes through two different situations for Case 1, i.e., Case 1(a) and Case 1(b) analyzed in Section 4. Specially, for Case 1(b), solo driving dominates carpooling and then no carpooler commute on the GP lane, so that the equilibrium trip cost is independent of Δ_1 and Δ_2. Similarly, for Case 2 in Figure 7(ii), when Δ_2 is smaller, all commuters choose carpooling and then the trip cost presents as different horizon lines which only depend on Δ_1. In Figure 7(iii), the trip cost for Case 3(a) with only carpooling and that for Case 3(b) with only solo driving are both unchanged with Δ_2.

6.2 Inefficiency of temporal-spatial capacity allocation

Propositions 1–6 will also apply to analyze how the inefficiency arising from a given temporal-spatial allocation varies with Δ_x and θ. Recalling $\rho = c^r / c^*$, the equilibrium trip cost under the optimal temporal-spatial capacity allocation scheme c^* keeps constant with respect to Δ_x and θ, and thus the change of ρ is the same as that of the equilibrium trip cost under a given temporal-spatial allocation scheme c^r. Here, we do not numerically discuss this to save space. Next, we only discuss the change of ρ with Δ_1 and Δ_2, given $\Delta_x = 2$ and $\theta = 0.5$ for Case 1, Case 2(a) and Case 3(a).

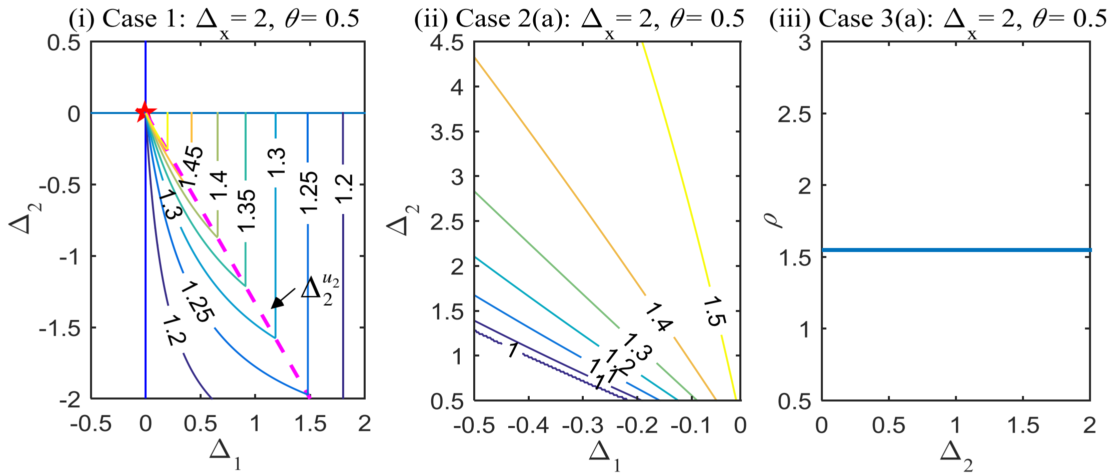

Fig. 8. Inefficiency ρ varying with Δ_1 and Δ_2.

As shown in Figure 8, the inefficiency of the temporal-spatial allocation with $\Delta_x = 2$ and $\theta = 0.5$ always increase with Δ_1 and Δ_2 for Case 1(a) and Case 2(a) whilst decreasing with Δ_1 for Case 1(b) and being constant for Case 3(a). The upper bound on ρ can be achieved at $\Delta_1 = 0$ and $\Delta_2 = 0$ for Case 1 (the red point), and line $\Delta_2 = \Delta_2^{u_2}$ divides the feasible region of Δ_1 and Δ_2 for the scenario with excess queue delay and that with excess capacity waste, which is consistent with the analysis in Section 4.

6.3 Upper bounds on the inefficiency

Here, we numerically discuss the respective impacts of parameters Δ_x, θ, m and N/s on the upper bounds on the inefficiency arising from the temporal-spatial allocation scheme only for Case 1, as shown in Figures 9–13.

Figure 9 shows the indifference curves of the upper bounds on the inefficiency with the change of θ and Δ_x for Case 1. For the scenario with excess queue delay, it is clear that the upper bound $\bar{\rho}$ in Eq. (21) increases with θ and Δ_x, i.e., the more capacity is allocated to the GP lane is, the larger the worst-case inefficiency is. This is consistent with Proposition 4, in which the equilibrium trip cost is minimized at $\theta = 0$. But for the scenario with excess capacity waste, the inefficiency bound $\bar{\rho}$ according to either Eq. (22) or Eq. (23), depending on the relationship between Δ_x and $\hat{\Delta}_x$, first decreases and then increases with θ, whilst it consistently decreases with Δ_x. It means that to allocate some capacity to the GP lane may be helpful for reducing the worst-case inefficiency for the scenario with excess capacity waste when $\Delta_x \le \hat{\Delta}_x$ as shown in Appendix D.2.

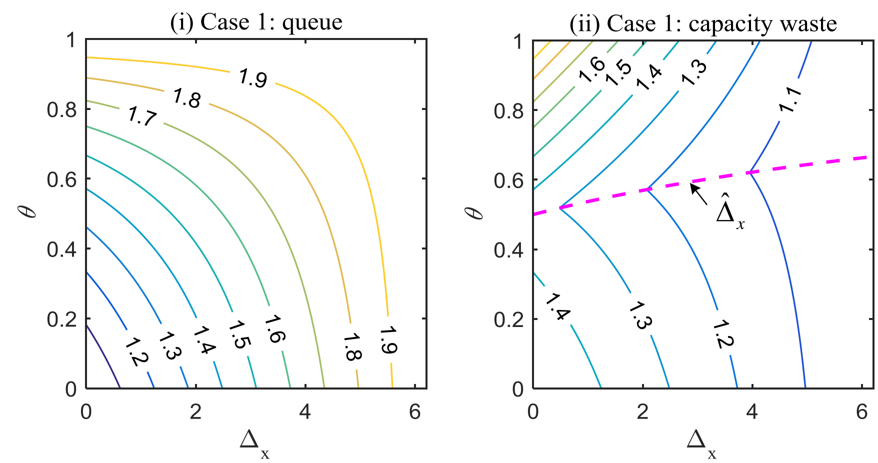

Fig. 9. Upper bounds on the inefficiency varying with θ and Δ_x.

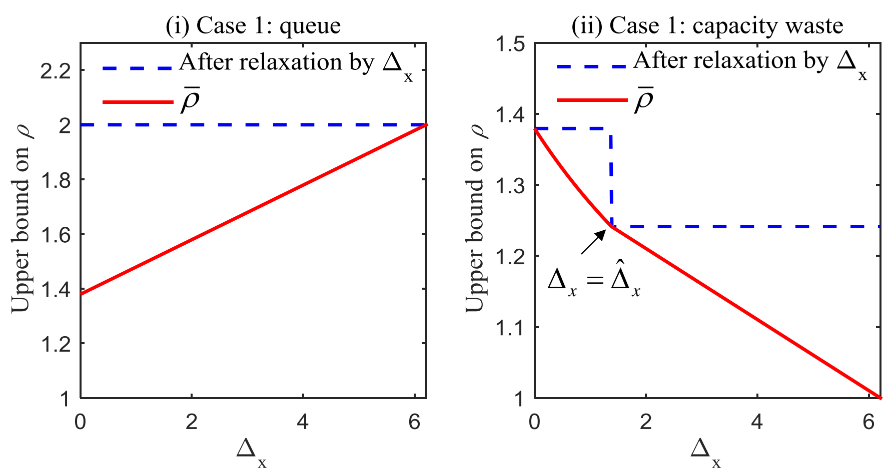

Fig. 10. Upper bounds on the inefficiency with the change of Δ_x.

Figure 10 depicts the changes of the upper bounds on the inefficiency $\bar{\rho}$ with Δ_x for Case 1 with $\theta = 0.55$. For the scenario with excess queue delay, it is obvious that $\bar{\rho}$ in Eq. (21) is increasing with Δ_x, and reaches its maximum, $m = 2$, at $\Delta_x = \delta N/s \approx 6.21$ (\$). This suggests the worst-case commute pattern for Case 1 with excess queue delay, where there is no time reservation for carpool purpose and all commuters drive alone on the highway. But for the scenario with excess capacity waste, $\bar{\rho}$ in either

Eq. (22) or Eq. (23) is decreasing with Δ_x, and reaches its maximum, $m/\left(\theta+m(1-\theta)\right)\approx 1.38$ for $\Delta_x\in[0,\hat{\Delta}_x]$ and $m^2(1-\theta)/\left((m-1)\left(\theta+m(1-\theta)\right)\right)\approx 1.24$ for $\Delta_x\in[\hat{\Delta}_x,\delta N/s]$ with $\hat{\Delta}_x\approx 1.38$ (\$). The derivations of all upper bounds relaxed by Δ_x can be found in Appendix D.2.

Figure 11 depicts the changes of the upper bounds on the inefficiency with θ before and after the relaxation by θ, besides those calculated in Propositions 7 for Case 1 when $\Delta_x=2$. it can be noted that, for the scenario with excess queue delay, $\bar{\rho}$ in Eq. (21) is increasing with θ, and reaches its maximum, $m=2$, at $\theta=1$. This corresponds to the worst case where all commuters solo drive alone on the highway in equilibrium, as even though the full bottleneck capacity is (spatially) allocated to the carpool lane but there is no time reservation for carpool purpose. The scenario with excess capacity waste is more complex, where $\bar{\rho}$ in Eq. (22) or Eq. (23) first decreases and then increases with θ, and reaches its minimum at $\theta=0.57$. Furthermore, the upper bound before the relaxation by θ, i.e., that relaxed only by Δ_x but depends on the value of θ, and the upper bound further relaxed by θ (thus independent of θ) are both segmented at $\theta=\bar{\theta}=1/m=0.5$ and $\theta=0.57$; see their derivations in Appendix D.2. These results imply that for the scenario with excess capacity waste, a spatial allocation of bottleneck capacity between the carpool lane and the GP lane at ratio $\theta=0.57$ can reduce the worst-case inefficiency.

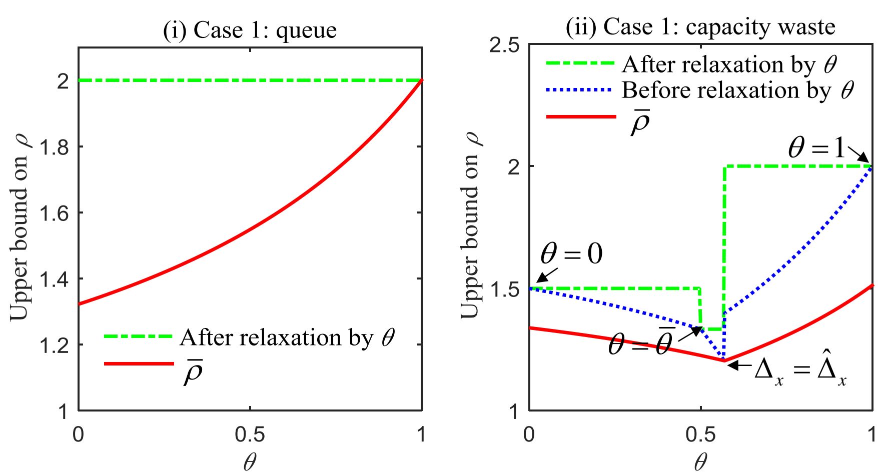

Fig. 11. Upper bounds on the inefficiency with the change of θ.

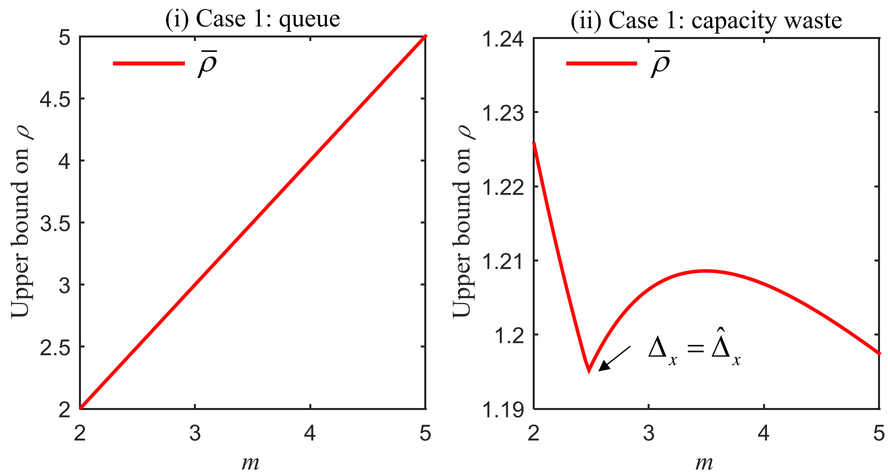

Fig. 12. Upper bounds on the inefficiency with the change of m .

Given the temporal-spatial allocation scheme with $\Delta_x = 2$ and $\theta = 0.5$, Figure 12 depicts the changes of the upper bounds on the inefficiency with respect to m . It shows clearly that $\bar{\rho}$ increases with m for the scenario with excess queue delay, implying that increasing the number of commuters in a carpooling vehicle would always increase the worst-case inefficiency arising from a non-optimal temporal-spatial allocation. This is however not necessarily true for the scenario with excess capacity waste, where there exists the right number of commuters in a carpooling vehicle, $m = 2.48$ here, that reduces the inefficiency. When $2 \le m \le 2.48$ which corresponds to $\Delta_x \ge \hat{\Delta}_x \ge 0$, there should be no time window reserved for carpool purpose (see also derivations in Appendix D.2), thus increasing m will reduce the inefficiency upper bound in Eq. (22). When $m > 2.48$ which corresponds to $\Delta_x < \hat{\Delta}_x$, the optimal equilibrium trip cost is associated with that with $\Delta_1 = \Delta_x$, which will render the inefficiency upper bound in Eq. (23) first increase and then decrease with m .

Figure 13 depicts the changes of the inefficiency upper bounds with N/s for given parameter values $\Delta_x = 2$ and $\theta = 0.55$. It is found that, $\bar{\rho}$ in Eq. (21) decreases with N/s for the scenario with excess queue delay, suggesting that, to shorten the rush hour period on the highway can increase the worst-case inefficiency arising from a non-optimal temporal-spatial allocation. For the scenario with excess capacity waste, $\bar{\rho}$ in Eq. (22) and Eq. (23) which correspond to $N/s \le 2.90$ and $N/s > 2.90$ respectively, increase with N/s .

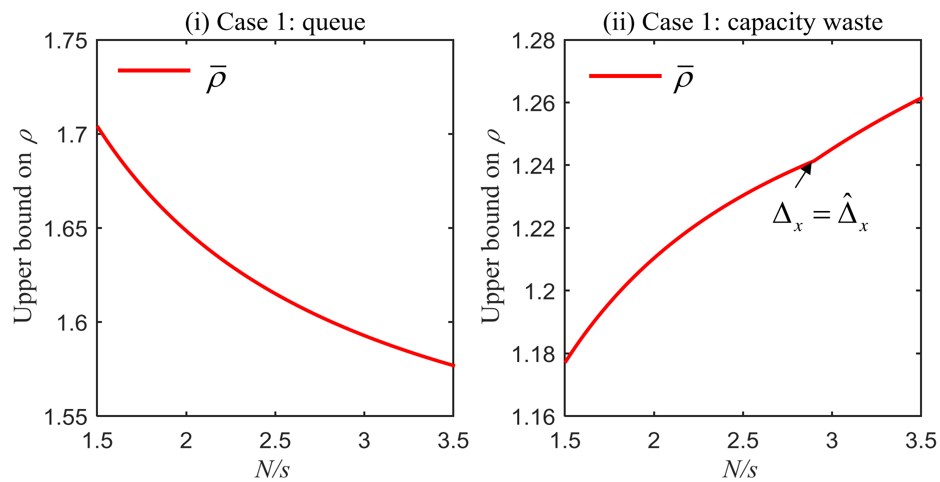

Fig. 13. Upper bounds on the inefficiency with the change of N/s .

7. Conclusions

This paper proposes temporal-spatial bottleneck capacity allocation schemes to manage the morning commute with carpool. A congested highway with one single bottleneck is assumed to serve commuters' travel, and its bottleneck capacity is proportionally shared by a GP lane and a carpool lane. Homogeneous commuters are assumed to make the joint choice of departure time and travel mode (solo-driving or carpooling) to minimize their trip costs, consisting of travel delay cost, schedule delay cost, and a time-dependent extra carpool cost where applied.

With a temporal-only allocation of bottleneck capacity, where the carpool lane is available only within a reserved time window, we show that the resultant morning commute patterns with carpool are significantly different from those generated by the standard bottleneck model without carpool whilst the impact of the reserved carpool time window on the commute patterns is similar to that of the single-step toll scheme studied in the literature. An optimal temporal-allocation scheme requires prior knowledge on the extra carpool cost Δ (or the combination of Δ_1 and Δ_2). If, in practice, the government makes an overestimation or underestimation of the extra carpool cost, excess queue delay or excess capacity waste will occur at the tails of the reserved time window.

We analyse the equilibrium commute patterns under a joint temporal-spatial carpool allocation scheme. It is proved that when the unit cost of schedule delay early is not too small as compared with that of travel time, the equilibrium trip cost decreases gradually with increasing allocation of capacity to the carpool lane. Furthermore, we show that it is possible to design an optimized joint temporal-spatial capacity allocation scheme that minimizes the system's total trip cost. For the optimized joint scheme, the full bottleneck capacity should be spatially designated to the carpool lane and the reserved time window for carpooling should be set in the middle of the peak period such that the first carpoolers just pass the bottleneck without experiencing excess queuing delay.

The optimal joint temporal-spatial capacity allocation schemes are also dependent on the accurate government estimation of the commuters' extra carpool cost. Inaccurate estimation leads to sub-optimal schemes, and efficiency loss. We formulate tight analytical expressions to quantify the upper bounds on the inefficiency arising from a non-optimal temporal-spatial allocation, and demonstrate through numerical illustrations how the worst-case inefficiency can be affected by a range of operational and behavioral variables, including the ratio of capacity allocation between GP and carpool lanes, the proportion of carpool traffic, and the length of peak period.

The model and key findings in this paper have not only theoretical, but also practical and policy implications. Firstly, as far as we are aware, this is the first comprehensive study on the complex morning commute problem with temporally and/or spatially designated carpool facilities, where all possible commute patterns under UE conditions are derived. These analytical derivations provide a solid foundation and practical tool for policy design such as congestion tolling, travel incentives and subsidies to induce commuters' travel, and modeling extensions such as incorporating the parking at destination, the use of autonomous unmanned vehicles (AUVs) (e.g., Liu, 2018; Tian et al., 2019), the competition and shifts between private vehicle and public transit. Secondly, in practice, it is difficult for the government to accurately estimate the commuters' extra carpool costs and then implement the optimal temporal-spatial capacity allocation. This inaccurate estimation may have very bad consequences. As shown in Eq. (25) and numerically discussed in Section 6.3, when the free-flow travel time is negligible, the upper bound on the inefficiency due to the non-optimal temporal-spatial capacity allocation may tend to the infinity. When only limited information about extra carpool cost components, such as range of change, is known a priori, the derivations of the upper bounds on the inefficiency provides a possible tool for the government agencies to evaluate the worst-case system performance due to the inaccurate estimation of the commuters' extra carpool cost and the impacts of some other operational and behavioral parameters.

Along the line of thought of this paper, our future studies will consider four possible extensions. Firstly, it is meaningful to convert the carpool lane to the HOT lane by allowing solo drivers commute during the reserved time window if they pay a toll (Zhong et al., 2020). Secondly, the number of commuters in a carpool vehicle is assumed to be constant during the rush hour period for simplicity. It is worthy to extend to incorporate carpoolers' matching process and explore its impact on commute patterns (Ma and Zhang, 2017; Wang et al., 2019). Thirdly, if the government does make a serious mistake in estimating the commuters' extra carpool cost, a trial and error approach (Yang et al., 2004) can be adopted to repeatedly adjust the reserved time window until excess queuing is eliminated without inducing capacity waste. Fourthly, many people either cannot, or strongly prefer not to, carpool. Carpooling is more attractive for commuters with lower incomes, longer trips to work, and somewhat restricted access to auto. Furthermore, the sign and magnitude of extra carpool cost may vary considerably across individuals. Accordingly, it is of interest to integrate user heterogeneity into the temporal-spatial capacity allocation problem for managing morning commute with carpool (Chen et al., 2015; Wu and Huang, 2015; Liu and Nie, 2017; Yu et al., 2019).

Acknowledgments

The research described in this paper was jointly supported by the National Key Research and Development Program of China (2018YFB1600900), the National Natural Science Foundation of China (71501012, 71771007, 71890971/71890972/71890970, 71971020), the Beijing Social Sciences Foundation (16GLC054), the UK Department for Transport (Project "Future Streets") and the joint project of National Natural Science Foundation of China and Joint Programming Initiative Urban Europe (NSFC–JPI UE) ('U-PASS', 71961137005). The authors would like to thank the Associate Editor, Robin Lindsey, and the anonymous referees for their helpful comments and suggestions, which improved the contents and composition of this paper substantially.

Appendix A: Bottleneck model with carpooling

A.1. Main variables and notation

α	Unit cost of in-vehicle travel time
β	Unit cost of arriving at destination early
γ	Unit cost of arriving at destination late
c_i	Trip cost for role i, $i = s, c$
$Q(t)$	Length of the queue at time t
T^f	Free-flow travel time
$T^v(t)$	Queuing time at the bottleneck

$\Delta(t)$	Extra carpool cost
Δ_1	Constant component of extra carpool cost
Δ_2	Time-varying component of extra carpool cost
N	Number of commuters
N_f	Number of vehicles
N_g	Number of vehicles on the GP lane
N_h	Number of vehicles on the HOV lane
N_c	Number of carpooling vehicles on the HOV lane
N_s	Number of solo-driving vehicles on the HOV lane
m	Average number of commuters in a carpooling vehicle
θ	Spatial capacity allocation ratio to the GP lane
r_i	Arrival rate at the bottleneck for vehicles with role i, $i = s, c$
s	Capacity of the bottleneck
t	Time
t^*	Desired arrival time at destination
t_i^o	Earliest arrival time at the bottleneck for each role of commuters i, $i = s, c$
t_i^e	Latest arrival time at the bottleneck for each role of commuters i, $i = s, c$
t^+	Starting time of temporal reservation
t^-	Ending time of temporal reservation
c^r	Trip cost with any temporal-spatial capacity allocation
c^*	Trip cost with optimal temporal-spatial capacity allocation
ρ	Inefficiency arising from an temporal-spatial capacity allocation
$\bar{\rho}$	Upper bound on the inefficiency by Δ_1 and Δ_2

A.2. Derivations for the bottleneck model with carpooling

Case 1: $0 < \Delta_1 < \delta N/s$

When $c_s = \delta N/s < \Delta_1$, all commuters would choose only solo driving to the destination. Thus, the necessary condition to ensure that both modes may be used is $0 < \Delta_1 < \delta N/s$ when $\Delta_1 > 0$. We first consider Case 1(a). As shown in Figure 1(i), the carpoolers pass the bottleneck at the center of the rush hour period and the solo drivers commute at the two tails. The queuing time (i.e., dividing queuing length by capacity) for both roles at t_c^o or t_c^e is the same, i.e., $Q_s = Q_c$. Using $\alpha Q_s = (\alpha + \Delta_2) Q_c + \Delta_1$, we can get the queuing time at the two critical points, $-\Delta_1/\Delta_2$.

Recall that $N = mN_c + N_s$ and $N_f = N_c + N_s$. According to the arrival order between the two roles, the earliest and the latest arrival time at the bottleneck for each role can be determined by equalizing the trip costs of the first and last commuters as follows:

$$t_s^o = t^* - \frac{\delta N_f}{s\beta},\ t_s^e = t^* + \frac{\delta N_f}{s\gamma},\ t_c^o = t_s^o - \frac{\Delta_1}{\Delta_2}\frac{\alpha - \beta}{\beta},\ t_c^e = t_s^e + \frac{\Delta_1}{\Delta_2}\frac{\alpha + \gamma}{\gamma}.$$

The numbers of solo driving vehicles and carpooling vehicles are, respectively

$$N_s = (t_c^o - t_s^o) r_s^1 + (t_s^e - t_c^e) r_s^2 = -\alpha \Delta_1 s / (\delta \Delta_2),\ N_c = (t_c^e - t_c^o) s = (N + \Delta_1 \alpha s/(\delta \Delta_2))/m.$$

To make $N_c > 0$ hold, it requires $\beta - \alpha < \Delta_2 < \Delta_2^{u_1}$, with $\Delta_2^{u_1} = -\Delta_1 \alpha s/(\delta N) < 0$. As a result, only if $0 < \Delta_1 < (1 - \beta/\alpha)\delta N/s$ and $\beta - \alpha < \Delta_2 < \Delta_2^{u_1}$, the commute pattern is consistent with Figure 1(i) for Case 1(a).

Clearly according to the ADL model (Arnott et al., 1990), the equilibrium trip cost is

$$\hat{c} = \delta N_f / s = \delta N/(ms) - (m-1)\alpha \Delta_1/(m\Delta_2).$$

Since $N_f = N_c + N_s < N$, $\hat{c}$ in the presence of carpooling is lower than the equilibrium trip cost in the absence of carpooling ($\delta N/s$). Hence, all commuters are better off due to carpooling.

Next, we consider the scenario that $0 < \Delta_1 < \delta N/s$ and $\Delta_2 \geq \Delta_2^{u_1}$ for Case 1(b). The time-varying component of the extra carpool cost (Δ_2) is so large that all commuters choose solo driving, and thus the equilibrium trip cost is $\hat{c} = \delta N/s$.

Case 2: $0 > \Delta_1 > -\delta N/(sm)$

When $c_c = \delta N/(ms) + \Delta_1 > 0$, all commuters would only choose carpooling to the destination. Thus, the necessary condition to ensure that both modes may be used is $0 > \Delta_1 > -\delta N/(sm)$ when $\Delta_1 < 0$. For Case 2(a), the commute pattern is that the carpoolers pass the bottleneck at the two tails and the solo drivers commute at the center of the rush hour period, as shown in Figure 1(ii). The earliest arrival time and the latest arrival time at the bottleneck for each role in this case are:

$$t_c^o = t^* - \frac{\delta N_f}{s\beta},\ t_c^e = t^* + \frac{\delta N_f}{s\gamma},\ t_s^o = t_c^o - \frac{\Delta_1}{\beta}\left(1 + \frac{\alpha - \beta}{\Delta_2}\right),\ t_s^e = t_c^e + \frac{\Delta_1}{\gamma}\left(1 + \frac{\alpha + \gamma}{\Delta_2}\right).$$

Using Eq. (4), the numbers of vehicles for solo driving and carpooling are, respectively

$$N_s=\left(t_s^e-t_s^o\right)s=N+m\left(\alpha+\Delta_2\right)\Delta_1 s\big/\left(\delta\Delta_2\right),\ N_c=\left(t_s^o-t_c^o\right)r_c^1+\left(t_c^e-t_s^e\right)r_c^2=-\left(\alpha+\Delta_2\right)\Delta_1 s\big/\left(\delta\Delta_2\right).$$

The corresponding equilibrium trip cost is

$$\hat{c}=\Delta_1+\delta N_f\big/s=\Delta_1+\delta N\big/s+\left(m-1\right)\Delta_1\left(\alpha+\Delta_2\right)\big/\Delta_2 .$$

Since $\Delta_1<0$ and $N_f<N$, all commuters are also better off due to carpooling in this case.

Due to $N_s>0$ and $-\delta N/(sm)<\Delta_1<0$ in Case 2(a), we easily get $\Delta_2>\Delta_2^{l_1}$ with $\Delta_2^{l_1}=-\alpha sm\Delta_1\big/\left(\Delta_1+\delta N/(sm)\right)>0$. Thus for Case 2(b), when $\beta-\alpha<\Delta_2\le\Delta_2^{l_1}$, all commuters choose carpooling with others and the corresponding equilibrium trip cost is $\hat{c}=\Delta_1+\delta N\big/\left(ms\right)$.

Case 3: $\Delta_1=0$

When $\Delta_1=0$, all commuters are indifferent to the arrival time with no queue and the equilibrium solutions depend on the value of Δ_2.

(a) If $\Delta_2>0$, solo driving is a better choice than carpooling at any time t, and thus $N_s=N$, $N_c=0$, and the corresponding equilibrium trip cost is $\hat{c}=\delta N/s$.

(b) If $\beta-\alpha<\Delta_2<0$, carpooling with others is a better choice than solo driving at any time t, and thus $N_c=N/m$, $N_s=0$, and the corresponding equilibrium trip cost is $\hat{c}=\delta N\big/\left(ms\right)$.

(c) If $\Delta_2=0$, all commuters are indifferent to any arrival time, and thus the order of departure is indeterminate.

Appendix B: Temporal-only capacity allocation

B.1. Proof of Proposition 1

Under the temporal-only bottleneck capacity allocation scheme (t^+,t^-), the reserved time window for carpooling is set according to the queuing time cost for the solo drivers just passing the bottleneck at t^+ and t^-, Δ_x, which can be smaller or larger than Δ_1. When $\Delta_x<\Delta_1$, then $\Delta_x^+=\Delta_1$. Using Eq. (8), we get $\mathrm{d}c^r\big/\mathrm{d}\Delta_x=-1/m<0$. If $\Delta_x>\Delta_1$, then $\Delta_x^+=\Delta_x$, we get $\mathrm{d}c^r\big/\mathrm{d}\Delta_x=1-1/m>0$. This completes the proof. □

B.2. Derivations for Case 2(a) under temporal-only capacity allocation

Under temporal-only bottleneck capacity allocation scheme, commuters can be divided into three groups, the commuters passing the highway bottleneck before t^+ and after t^- by driving alone or carpooling, and the carpoolers commuting inside $[t^+, t^-]$. Denote N_c^r and N_c^u as the numbers of carpooling vehicles passing the bottleneck inside and outside of the reserved time window, respectively. Then we get the numbers of carpooling vehicles and total vehicles, $N_c = N_c^r + N_c^u$ and $N_f = N_c + N_s$. Since the commuters arriving at the bottleneck at the times corresponding to points B and C of Figure 4 experience the same schedule early delay, the extra carpool cost should be equal to the difference between the two queuing delay costs. For a given Δ_x, the rush hour period for arriving at the bottleneck and the reserved time window can be determined by equalizing the trip costs of the first and last commuters for solo-driving or carpooling as follows:

$$t_c^o = t^* - \frac{\delta N/m}{s\beta} - \frac{(m-1)}{m}\frac{\alpha\Delta_1/\Delta_2 + \Delta_x}{\beta},\ t_c^e = t^* + \frac{\delta N/m}{s\gamma} + \frac{(m-1)}{m}\frac{\alpha\Delta_1/\Delta_2 + \Delta_x}{\gamma},$$

$$t_s^o = t_c^o - \frac{\Delta_1}{\Delta_2}\frac{\alpha + \Delta_2 - \beta}{\beta},\ t_s^e = t_c^e + \frac{\Delta_1}{\Delta_2}\frac{\alpha + \Delta_2 + \gamma}{\gamma},$$

$$t^+ = t_c^o + \frac{(\Delta_x - \Delta_1)}{\beta},\ t^- = t_c^e - \frac{(\Delta_x - \Delta_1)}{\gamma}.$$

And the number of solo drivers and the number of carpooling vehicles passing the bottleneck within the reserved time window or not can be expressed as follows:

$$N_s = (t_s^e - t^-)s + (t^+ - t_s^o)s = (\alpha\Delta_1/\Delta_2 + \Delta_x)s/\delta,$$

$$N_c^r = (t^- - t^+)s = N_f - \frac{(\Delta_x - \Delta_1)}{\delta}s,\ N_c^u = (t_c^e - t_s^e)s + (t_s^o - t_c^o)s = -\frac{\Delta_1}{\Delta_2}\frac{(\alpha + \Delta_2)s}{\delta}.$$

Appendix C: Temporal-spatial capacity allocation

C.1. Proof of Proposition 4

For Case 1(a), we easily get from Eq. (13)

$$\frac{\partial c^r}{\partial\theta} = \frac{(m-1)}{m}\left(-\Delta_x - \alpha\frac{\Delta_1}{\Delta_2}\right) - (\Delta_x^+ - \Delta_x).$$

For the excess queue delay, $\Delta_x^+ = \Delta_x$. If $\Delta_x = -\alpha\Delta_1/\Delta_2$, we get $\partial c^r/\partial\theta = 0$, and then for $\Delta_x \in [\Delta_1, -\alpha\Delta_1/\Delta_2]$, that $\partial c^r/\partial\theta \geq 0$ always holds.

For the excess capacity waste with $\Delta_x \in [0, \Delta_1]$, $\Delta_x^+ = \Delta_1$, then $\partial c^r / \partial\theta = \Delta_x / m + \{(m-1)\alpha / m(-\Delta_2) - 1\}\Delta_1$, which is increasing with Δ_x. Since $\beta - \alpha < \Delta_2 < \Delta_2^{u_2} < 0$, we get $\partial c^r / \partial\theta > 0$ if $\Delta_x = \Delta_1$. But for $\Delta_x = 0$, we get $\partial c^r / \partial\theta = \left((m-1)\alpha / (-\Delta_2 m) - 1\right)\Delta_1$. It suggests that if $\Delta_2 \geq -\alpha(m-1)/m$, then $\partial c^r / \partial\theta \geq 0$ for $\forall \Delta_x \in [0, \Delta_1]$; otherwise, there exists $\Delta_x^{\#} = \left(m - (m-1)\alpha / (-\Delta_2)\right)\Delta_1 < \Delta_1$ to make $\partial c^r / \partial\theta \big|_{\Delta_x^{\#}} = 0$ hold. Therefore, if $\Delta_2 > -\alpha(m-1)/m$, we get $\partial c^r / \partial\theta > 0$ for $\forall \Delta_x \in [0, \Delta_1]$, whilst $\partial c^r / \partial\theta < 0$, at $\Delta_x \in \left[0, \Delta_x^{\#}\right)$ and $\partial c^r / \partial\theta \geq 0$ at $\Delta_x \in \left[\Delta_x^{\#}, \Delta_1\right]$. Since $\Delta_2 > \beta - \alpha$, it is obvious that $\Delta_2 > -\alpha(m-1)/m$ when $m \geq \alpha/\beta$.

For Case 1(b), we get from Eq. (14)

$$\frac{dc^r}{d\theta} = \frac{(m-1)\delta N/s}{\left(\theta + m(1-\theta)\right)^2} - \frac{(m-1)\Delta_x + m\left(\Delta_x^+ - \Delta_x\right)}{\left(\theta + m(1-\theta)\right)^2}.$$

For the excess queue delay, since $\Delta_x^+ = \Delta_x$, we have $dc^r / d\theta = (m-1)(\delta N/s - \Delta_x) / \left(\theta + m(1-\theta)\right)^2$. Substituting $\Delta_x \leq -\alpha\Delta_1/\Delta_2$ into the equation, we get $\partial c^r / \partial\theta \geq 0$. Hence, for $\Delta_x \in [\Delta_1, -\alpha\Delta_1/\Delta_2]$, we have $\partial c^r / \partial\theta \geq 0$.

For the capacity waste with $\Delta_x \in [0, \Delta_1]$, $\Delta_x^+ = \Delta_1$, $0 < \Delta_1 \leq \delta N/s$ and $dc^r / d\theta = \left((m-1)\delta N/s - (m\Delta_1 - \Delta_x)\right) / \left(\theta + m(1-\theta)\right)^2$, which is increasing with Δ_x. If $\Delta_x = \Delta_1$, we get $\partial c^r / \partial\theta > 0$. But for $\Delta_x = 0$, we get $dc^r / d\theta = \left((m-1)\delta N/s - m\Delta_1\right) / \left(\theta + m(1-\theta)\right)^2$. It suggests that there exists $\Delta_x^{\#} = m\Delta_1 - (m-1)\delta N/s < \Delta_1$ to make $\partial c^r / \partial\theta \big|_{\Delta_x^{\#}} = 0$ hold. Therefore, if $\alpha/\beta \leq m$, we get $\partial c^r / \partial\theta \geq 0$ for $\forall \Delta_x \in [0, \Delta_1]$, whilst $\partial c^r / \partial\theta < 0$, at $\Delta_x \in \left[0, \Delta_x^{\#}\right)$ and $\partial c^r / \partial\theta \geq 0$ at $\Delta_x \in \left[\Delta_x^{\#}, \Delta_1\right]$.

This completes the proof. □

C.2. Derivations for Case 2(a) under temporal-spatial capacity allocation

For Case 2(a), according to the derivations in Appendix A.2 and Appendix B.2, we easily get the numbers of carpooling vehicles on the GP lane and solo-driving vehicles on the HOV lane as follows:

$$N_c^g = -\frac{\Delta_1}{\Delta_2}\frac{(\alpha + \Delta_2)}{\delta}\theta s,\ N_s^h = \left(\Delta_x + \alpha\frac{\Delta_1}{\Delta_2}\right)\frac{(1-\theta)s}{\delta},$$

where $-\alpha\Delta_1/\Delta_2 \le \Delta_x \le \overline{\Delta}_x$, with $\overline{\Delta}_x = \delta N/s + (m-1)\alpha\Delta_1/\Delta_2 + m\Delta_1$.

Since the peak period is the same for both lanes, i.e., $N^g/s^g = N^h/s^h$, and using $N = N^g + mN^h$, $N^g = N_s^g + N_s^h$ and $N^h = N_c^g + N_c^h$, then we get the numbers of vehicles for solo driving on the GP lane and carpooling on the HOV lane and the total vehicles on the highway as follows:

$$N_s^g = \theta N_f - N_c^g,\ N_c^h = (1-\theta)N_f - N_s^h \text{ and } N_f = \frac{N + (m-1)\left(N_s^h - N_c^g\right)}{\theta + m(1-\theta)}.$$

To make $N_s^g > 0$, it requires $\Delta_2 > \Delta_2^{l_2} = -\alpha\Delta_1\left(m + (m-1)(1-\theta)\right)/\left(\delta N/s + m\Delta_1 + (m-1)(1-\theta)\Delta_x\right)$, with $-\delta N/(ms) < \Delta_1 < 0$. As a result, when $-\delta N/(ms) < \Delta_1 < 0$, only if $\Delta_2 > \Delta_2^{l_2}$, the commute pattern is that for Case 2(a) under temporal-spatial capacity allocation, otherwise if $\beta - \alpha < \Delta_2 \le \Delta_2^{l_2}$, all commuters choose carpooling for Case 2(b), and thus whether to set the reserved time window or not makes no difference to the commute.

C.3. Proof of Proposition 5

For Case 2(a) with $\Delta_2^{l_2} < \Delta_2$, we get from Eq. (16)

$$\frac{\partial c^r}{\partial \theta} = \frac{(m-1)}{\left(\theta + m(1-\theta)\right)^2}\left(\frac{\delta N}{s} - \left(\Delta_x - \Delta_1\right)\right) + \frac{(m-1)^2}{\left(\theta + m(1-\theta)\right)^2}\left(\Delta_2 + \alpha\right)\frac{\Delta_1}{\Delta_2}.$$

Substituting $\Delta_x = \overline{\Delta}_x = \delta N/s + (m-1)\alpha\Delta_1/\Delta_2 + m\Delta_1$ into the above equation, we get $\partial c^r/\partial\theta = 0$. Since $\partial c^r/\partial\theta$ is decreasing with Δ_x for , we get $\partial c^r/\partial\theta \ge 0$ for $\Delta_x \in [-\alpha\Delta_1/\Delta_2, \overline{\Delta}_x]$. This completes the proof. □

C.4. Derivations for Case 3(a) under temporal-spatial capacity allocation

For Case 3(a), where $\Delta_2 > 0$, driving alone is a better choice than carpooling without carpooling reservation, then the number of carpoolers on the GP lane is zero, i.e., $N_c^g = 0$. Similar to that in Case 1(a), the number of solo drivers on the HOV lane can be formulated as:

$$N_s^h = \frac{\Delta_x}{\delta}(1-\theta)s.$$

Since the peak period is the same for both lanes, i.e., $N^g/s^g = N^h/s^h$, and using $N = N^g + mN^h$, $N^g = N_s^g + N_s^h$ and $N^h = N_c^g + N_c^h$, we get the numbers of vehicles for solo driving on the GP lane and carpooling on the HOV lane and the total vehicles on highway as follows:

$$N_s^g = \theta N_f,\ N_c^h = (1-\theta)N_f - N_s^h \text{ and } N_f = \frac{N + (m-1)N_s^h}{\theta + m(1-\theta)}.$$

C.5. Proof of Proposition 6

From Eq. (17), we get

$$\frac{\partial c^r}{\partial \theta} = \frac{(m-1)}{\left(\theta + m(1-\theta)\right)^2}\left(\frac{\delta N}{s} - \Delta_x\right).$$

Substituting $\Delta_x = \delta N/s$ into the above equation, we get $\partial c^r/\partial\theta = 0$. Obviously, $\partial c^r/\partial\theta$ is decreasing with Δ_x, thus $\partial c^r/\partial\theta \geq 0$ for $\Delta_x \in [0, \delta N/s]$. This completes the proof. □

Appendix D: Bounding the inefficiency

D.1. Proof of Lemma 1

For Case 1(a), where $\beta - \alpha < \Delta_2 < \Delta_2^{u_2}$, from Eq. (13), we get

$$\frac{\partial c^r}{\partial \Delta_2} = \frac{\Delta_1}{(\Delta_2)^2}\frac{(m-1)}{m}\alpha\theta > 0 \text{ and } \frac{\partial c^r}{\partial \Delta_1} = \frac{(m-1)}{m}\left((1-\theta) - \frac{\alpha\theta}{\Delta_2}\right) > 0.$$

Using Eq. (14), we get $\partial c^r/\partial\Delta_2 = 0$ and $\partial c^r/\partial\Delta_1 = 0$ for Case 1(b).

For Case 2(a), where $-\delta N/(ms) - (m-1)(1-\theta)\Delta_x/m < \Delta_1 < 0$ and $\Delta_2 > \Delta_2^{l_2}$, from Eq. (16), we get

$$\frac{\partial c^r}{\partial \Delta_2} = \frac{(m-1)}{\theta + m(1-\theta)}\left(-\frac{\alpha\Delta_1}{(\Delta_2)^2}\right) > 0 \text{ and } \frac{\partial c^r}{\partial \Delta_1} = 1 + \frac{(m-1)}{\theta + m(1-\theta)}\left(\theta + \frac{\alpha}{\Delta_2}\right) > 0.$$

For Case 3(a), where $\Delta_1 = 0$ and $\Delta_2 > 0$, from Eq. (17), we get $\partial c^r/\partial\Delta_2 = 0$.

This completes the proof. □

D.2. Proof of Proposition 7

From the proof of Lemma 1, we have $\partial c^r / \partial \Delta_2 > 0$ for Case 1(a), where $\beta - \alpha < \Delta_2 < \Delta_2^{u_2} = \frac{-\Delta_1 \alpha(\theta + m(1-\theta))}{\delta N/s + (1-\theta)(m\Delta_x^+ - \Delta_x)}$, and $\partial c^r / \partial \Delta_2 = 0$ for Case 1(b), where $\Delta_2 \geq \Delta_2^{u_2}$. Thus, from Eq. (19), we get for $0 \leq \Delta_x \leq -\alpha\Delta_1/\Delta_2$

$$\rho \leq \rho\Big|_{\Delta_2=\Delta_2^{u2}} = \frac{\left(\delta N/s + (1-\theta)\left((m-1)\Delta_x + m\left(\Delta_x^+ - \Delta_x\right)\right)\right)/\left(\theta + m(1-\theta)\right)}{\delta N/(ms) + (m-1)\Delta_1/m},$$

which $\rho\Big|_{\Delta_2=\Delta_2^{u2}}$ is the same for Case 1(a) and Case 1(b).

For the excess queuing scenario with $\Delta_1 \leq \Delta_x \leq -\alpha\Delta_1/\Delta_2\Big|_{\Delta_2=\Delta_2^{u2}}$, substituting $\Delta_x^+ = \Delta_x$ into $\rho\Big|_{\Delta_2=\Delta_2^{u2}}$ and using $\Delta_1 > 0$, we easily get for $\Delta_1 \in [0, \Delta_x]$ and $\Delta_x \in [0, \delta N/s]$

$$\rho\Big|_{\Delta_2=\Delta_2^{u2}} \leq \rho\Big|_{\Delta_2=\Delta_2^{u2}, \Delta_1 \to 0^+} = \frac{\left(\delta N/s + (1-\theta)(m-1)\Delta_x\right)/\left(\theta + m(1-\theta)\right)}{\delta N/(ms)}$$

$$\leq \rho\Big|_{\Delta_2=\Delta_2^{u2}, \Delta_1 \to 0^+, \Delta_x = \delta N/s} = \frac{\delta N/s}{\delta N/(ms)} = m.$$

For the excess capacity waste scenario with $0 \leq \Delta_x \leq \Delta_1$, using $\Delta_x^+ = \Delta_1$, we easily get

$$\rho\Big|_{\Delta_2=\Delta_2^{u2}} = \frac{\left(\delta N/s + (1-\theta)\left(m\Delta_1 - \Delta_x\right)\right)/\left(\theta + m(1-\theta)\right)}{\delta N/(ms) + (m-1)\Delta_1/m}$$

$$= \frac{m}{(m-1)\left(\theta + m(1-\theta)\right)}\left(m(1-\theta) + \frac{(m\theta - 1)\delta N/s - (1-\theta)(m-1)\Delta_x}{\delta N/s + (m-1)\Delta_1}\right).$$

Letting $\hat{\Delta}_x = \frac{(m\theta-1)\delta N/s}{(m-1)(1-\theta)}$, $\hat{\Delta}_x \geq 0$ if $1 \geq \theta \geq \bar{\theta} = 1/m$; $\hat{\Delta}_x < 0$ otherwise. Next, we will discuss two different situations.

(1) When $\Delta_x \geq \hat{\Delta}_x$, $\rho\Big|_{\Delta_2=\Delta_2^{u2}}$ is increasing with Δ_1. Hence, further if $1 \geq \theta \geq \bar{\theta}$, we have for $\Delta_1 \in [\Delta_x, \delta N/s]$ and $\Delta_x \in [\hat{\Delta}_x, \delta N/s]$,

$$\rho\Big|_{\Delta_2=\Delta_2^{u2}} \leq \rho\Big|_{\Delta_2=\Delta_2^{u2}, \Delta_1=\delta N/s} = \frac{\left(\left(1 + (1-\theta)m\right)\delta N/s - (1-\theta)\Delta_x\right)/\left(\theta + m(1-\theta)\right)}{\delta N/s}$$

$$\le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s,\Delta_x=\hat{\Delta}_x} = \frac{m^2(1-\theta)}{(m-1)\big(\theta+m(1-\theta)\big)} \le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s,\Delta_x=\hat{\Delta}_x,\theta=\overline{\theta}} = \frac{m^2}{m^2-(m-1)}.$$

If $0\le\theta\le\overline{\theta}$, we have for $\Delta_1\in[\Delta_x,\delta N/s]$ and $\Delta_x\in[0,\delta N/s]$

$$\rho\le\rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s} = \frac{\Big(\big(1+(1-\theta)m\big)\delta N/s-(1-\theta)\Delta_x\Big)\Big/\big(\theta+m(1-\theta)\big)}{\delta N/s}$$

$$\le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s,\Delta_x=0} = \frac{1+m(1-\theta)}{\theta+m(1-\theta)} \le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\delta N/s,\Delta_x=0,\theta=0} = (1+m)/m.$$

(2) When $\Delta_x\le\hat{\Delta}_x$, $\rho\Big|_{\Delta_2=\Delta_2^{u2}}$ is decreasing with Δ_1, with $\theta\ge\overline{\theta}$. Hence, we have for $\Delta_1\in[\Delta_x,\delta N/s]$ and $\Delta_x\in[0,\min(\hat{\Delta}_x,\delta N/s)]$

$$\rho\Big|_{\Delta_2=\Delta_2^{u2}} \le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\Delta_x} = \frac{\big(\delta N/s+(1-\theta)(m-1)\Delta_x\big)\Big/\big(\theta+m(1-\theta)\big)}{\delta N/(ms)+(m-1)\Delta_x/m}$$

$$\le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\Delta_x,\Delta_x=0} = m\Big/\big(\theta+m(1-\theta)\big) \le \rho\Big|_{\Delta_2=\Delta_2^{u2},\Delta_1=\Delta_x,\Delta_x=0,\theta=1} = m.$$

This completes the proof. □

D.3. Proof of Proposition 8

Using Eq. (24) and $\Delta_2>\Delta_2^{l_2}=\dfrac{-\alpha\Delta_1\big(m+(m-1)(1-\theta)\big)}{\delta N/s+m\Delta_1+(m-1)(1-\theta)\Delta_x}$, we can get

$$\rho=\frac{\Delta_1+\dfrac{\delta N}{\big(\theta+m(1-\theta)\big)s}+\dfrac{(m-1)}{\theta+m(1-\theta)}\left((\Delta_2+\alpha)\dfrac{\Delta_1}{\Delta_2}+(1-\theta)(\Delta_x-\Delta_1)\right)}{\dfrac{\delta N}{ms}+\Delta_1}.$$

Since $\partial c^r/\partial\Delta_2>0$ from Lemma 1, we get

$$\rho\le\rho\Big|_{\Delta_2\to+\infty}=\frac{\dfrac{\delta N}{s}+m\Delta_1+(m-1)(1-\theta)\Delta_x}{\big(\theta+m(1-\theta)\big)\left(\dfrac{\delta N}{ms}+\Delta_1\right)}=\frac{m}{\theta+m(1-\theta)}+\frac{(m-1)(1-\theta)\Delta_x}{\big(\theta+m(1-\theta)\big)\left(\dfrac{\delta N}{ms}+\Delta_1\right)}.$$

Since $\Delta_x\ge0$, $\rho\Big|_{\Delta_2\to+\infty}$ is decreasing with Δ_1. Hence, for $\Delta_1\in(-\delta N/(sm),0)$ and $\Delta_x\in[0,\delta N/s]$, we have

$$\rho\Big|_{\Delta_2\to+\infty} \le \rho\Big|_{\Delta_2\to+\infty,\Delta_1\to-\delta N/(sm)} \to +\infty .$$

This completes the proof. □

References

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of Urban Economics 27(1), 111–130.

Baldassare, M., Ryan, S., Katz, C., 1998. Suburban attitudes toward policies aimed at reducing solo driving. Transportation 25, 99–117.

Brown, A.E, 2020. Who and where rideshares? Rideshare travel and use in Los Angeles. Transportation Research Part A 136, 120–134.

Buliung, R.N, Soltys, K., Habel, C., 2009. Driving factors behind successful carpool formation and use. Transportation Research Record 2118, 31–38.

Chan, N.D., Shaheen, S.A, 2012. Ridesharing in North America: past, present, and future. Transport Reviews 32, 93–112.

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

Dahlgren, J.W., 1998. High occupancy vehicle lanes: Not always more effective than mixed flow lanes. Transportation Research Part A 32(2), 99–114.

Ferguson, E., 1997. The rise and fall of the American carpool: 1970–1990. Transportation 24(7), 349–376.

Fosgerau, M., 2011. How a fast lane may replace a congestion toll. Transportation Research Part B 45(6), 845–851.

Guiliano, G., Levine, D.W., Teal, R.F., 1990. Impact of high occupancy vehicle lanes on carpooling behavior. Transportation 17(2), 159–177.

Hall, J.D., 2018. Pareto improvements from Lexus lanes: The effects of pricing a portion of the lanes on congested highways. Journal of Public Economics 158, 113–125.

Huang, H.J., Yang, H., Bell, M.G.H., 2000. The models and economics of carpools. Annals of Regional Science 34, 55–68.

Jacobson, S.H., King, D.M., 2009. Fuel saving and ridesharing in the US: motivations, limitations, and opportunities. Transportation Research Part D 14(1), 14–21.

Konishi, H., Mun, S.I., 2010. Carpooling and congestion pricing: HOV and HOT lanes. Regional Science and Urban Economics 40(4), 173–186.

Koppelman, F., Bhat, C., Schofer, J., 1993. Market research evaluation of actions to reduce suburban traffic congestion: commuter travel behavior and response to demand reduction actions. Transportation Research Part A 27, 383–393.

Laih, C. H., 1994. Queuing at a bottleneck with single- and multi-step tolls. Transportation Research Part A 28(3), 197–208.

Laih, C.H., 2004. Effects of the optimal step toll scheme on equilibrium commuter behavior. Applied Economics 36(1), 59–81.

Li, J., Embry, P., Mattingly, S.P., Sadabadi, K.F., Rasmidatta, I., Burris, M.W., 2007. Who chooses to carpool and why? Examination of Texas carpoolers. Transportation Research Record 2021, 110–117.

Li, Z.-C., Huang, H.-J., Yang, H., 2020. Fifty years of the bottleneck model: a bibliometric review and future research directions. Transportation Research Part B, 139, 311–342.

Li, Z.-C., Lam, W.H.K., Wong, S.C., 2017. Step tolling in an activity-based bottleneck model. Transportation Research Part B 101, 306–334.
Lindsey, R., van den Berg, V.A., Verhoef, E.T., 2012. Step tolling with bottleneck queuing congestion. Journal of Urban Economics 72(1), 46–59.
Liu, T.-L, Huang, H.-J, Tian, L.-J., 2009. Microscopic simulation of multi-lane traffic under dynamic tolling and information feedback. Journal of Central South University of Technology 5, 167–172.
Liu, W., 2018. An equilibrium analysis of commuter parking in the era of autonomous vehicles. Transportation Research Part C 92, 191–207.
Liu, W., Yang, H., Yin, Y., 2015. Efficiency of a highway use reservation system for morning commute. Transportation Research Part C 56, 293–308.
Liu, Y., Li, Y., 2017. Pricing scheme design of ridesharing program in morning commute problem. Transportation Research Part C 79, 156–177.
Liu, Y., Nie, Y.M., 2017. A credit-based congestion management scheme in general two-mode networks with multiclass users. Networks and Spatial Economics 17(3), 681–711.
Lou, Y., Yin, Y., Laval, J.A., 2011. Optimal dynamic pricing strategies for high-occupancy/toll lanes. Transportation Research Part C 19(1), 64–74.
Ma, R., Zhang H.M., 2017. The morning commute problem with ridesharing and dynamic parking charges. Transportation Research Part B 106, 345–374.
Menendez, M., Daganzo, C.F., 2007. Effects of HOV lanes on freeway bottlenecks. Transportation Research Part B 41(8), 809–822.
Monchambert, G., 2020. Why do (or don't) people carpool for long distance trips? a discrete choice experiment in France. Transportation Research Part A 132, 911–931.
Nie, Y.M., Yin, Y., 2013. Managing rush hour travel choice with tradable credit scheme. Transportation Research Part B 50, 1–19.
OECD/ITF, 2014. Valuing Convenience in Public Transport. ITF Round Tables, No. 156. OECD Publishing, France. https://doi.org/10.1787/9789282107683-en
Qian, Z., Zhang, H.M. 2011. Modeling multi-modal morning commute in a one-to-one corridor network. Transportation Research Part B 19, 254–269.
Small, K.A., 1982. The scheduling of consumer activities: work trips. American Economic Review 72(3), 467–479.
Stamos, I., Kitis, G., Basbas, S., Tzevelekis, I., 2012. Evaluation of a high occupancy vehicle lane in central business district Thessaloniki. Procedia-Social and Behavioral Sciences 48, 1088–1096.
Teal, R.F., 1987. Carpooling: who, how and why. Transportation Research Part A 21, 203–214.
Tian, L.-J., Sheu, J.B., Huang, H.-J., 2019. The morning commute problem with endogenous shared autonomous vehicle penetration and parking space constraint. Transportation Research Part B 123, 258–278.
Varaiya, P. Effectiveness of California's high occupancy vehicle (HOV) system, California PATH Research Report, UCB-ITS-PRR-2007-5, 2007.
Vickrey, W.S., 1969. Congestion theory and transport investment. American Economic Review 59(2), 251–261.
Wang, J.-P., Ban, X., Huang, H.-J., 2019. Dynamic ridesharing with variable-ratio charging compensation scheme for morning commute. Transportation Research Part B 122, 390–415.
Weinberger, R., Kaehny, J., Rufo, M., 2010. U.S. parking policies: An overview of management strategies. Institute for Transportation and Development Policy, New York.
Wu, W.-X., Huang, H.-J., 2015. An ordinary differential equation formulation of the bottleneck model with user heterogeneity. Transportation Research Part B 81, 34–58.

Xiao, F., Qian, Z., Zhang H.M., 2011. The morning commute problem with coarse toll and nonidentical commuters. Networks Spatial Economics 11, 343–369.

Xiao, F., Shen, W., Zhang, H.M., 2012. The morning commute under flat toll and tactical waiting. Transportation Research Part B 46(10), 1346–1359.

Xiao, L.-L., Liu, T.-L., Huang, H.-J., 2016. On the morning commute with carpooling behavior under parking space constraint. Transportation Research Part B 91(1), 383–407.

Xiao, L.-L., Liu, T.-L., Huang, H.-J., 2019. Tradable permit schemes for managing morning commute with carpool under parking space constraint. Transportation, https://doi.org/10.1007/s11116-019-09982-w.

Xu, D., Guo, X., Zhang, G. 2019. Constrained optimization for bottleneck coarse tolling. Transportation Research Part B 128, 1–22.

Yang, H., Huang, H.-J., 1999. Carpooling and congestion pricing in a multilane highway with high-occupancy-vehicle lanes. Transportation Research Part A 33, 139–155.

Yang, H., Meng, Q., Lee, D.H., 2004. Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. Transportation Research Part B 38(6), 477–493.

Yu, X., van den Berg, V.A., Verhoef, E.T., 2019. Carpooling with heterogeneous users in the bottleneck model. Transportation Research Part B 127, 178–200.

Zhong, L., Zhang, K., Nie, Y.M., Xu, J., 2020. Dynamic carpool in morning commute: Role of high-occupancy-vehicle (HOV) and high-occupancy-toll (HOT) lanes. Transportation Research Part B 135, 98–119.