Congestion Tolling for Mixed Urban Freight and Passenger Traffic

Chaoda Xie[1] and Xifu Wang[1]

[1]*School of Traffic and Transportation, Beijing Jiaotong University, 100044 Beijing, China*

Abstract. This paper investigates the welfare effects of optimal tolling on urban traffic congestion, in a bottleneck model, with mixed freight and passenger users. The users' marginal utility of time is considered to be varying with time. Under both no-toll equilibrium and socially optimal tolling, the users are found to sort their arrival time according to the increasing rates of marginal utility at the destination. The optimal toll that maximizes social welfare does not change each user's indirect utility relative to the no-toll equilibrium, but completely removes the queue, which also removes the barrier of freight carriers to accept congestion pricing by relating their marginal utilities directly to the toll. When the toll is equally rebated, the proposed social optimal tolling is a Pareto improvement relative to the no-toll equilibrium. Those more productive users also suffer more in both no-toll equilibrium and optimal tolling, which indicates that a differentiated redistribution of toll revenues could be an incentive to improve productivity.

1 Introduction

This paper focuses on the dynamics of traffic congestion induced by two distinct transportation demands, i.e. passengers and freights. Freight vehicles are evidently found to be competing with passenger cars for the capacity of urban roads. Empirical evidence [1] shows that the that freight traffic is a significant proportion of the total traffic volume at the major interstate crossings of the Port Authority of New York and New Jersey (PANYNJ), and the flow of freight and passenger traffic are overlapped. GPS data of Tokyo Metropolis Area [2] indicates that trucks are concentrating in the city center during rush hours. In a survey of Italian cities [3], the peak of freight traffic often overlaps with the peak of passenger traffic, during 8AM-11AM and 3PM-5PM, due to regulatory concerns. A report of Beijing, China, indicates that over 80% of freight traffic volume in terms of the number of trips occurs during daytime in the central area of the city, including 11.7% occurs during peak hours of passenger traffic [4].

Despite the existence of mixed freight and passenger traffic, the study of the dynamics of urban traffic congestion begins by treating the users as commuters only. As the pioneer work, [5] first depicts the scheduling preference of commuters as their travel cost when passing through a single bottleneck in the city center. The model assumes that commuters valued the travel time per se at a rate of α. They are supposed to arrive at an exogenous preferred arrival time (PAT), and are subjected to a scheduling cost at a rate of β per time unit for early arrival or γ per time unit for late arrival, which is well-knows as the $\alpha-\beta-\gamma$ specification. In light of this simple and tractable scheduling preference specification, the congestion dynamics with respect to the heterogeneity of passenger traffic has been extensively discussed. Limited number of user classes and continuously distributed user classes are considered by [6] and [7] respectively, but the ratio of β/γ is fixed for the purpose of analytic convenience. [8] discusses equilibrium properties by assuming a more general scheduling delay cost function, but the PAT remains exogenous. The specification of PAT with an indifference band has also been considered in a number of studies to allow for flexibility of arrival time (e.g. [9], [10]).

However, it is not generally true that the preferred arrival times can be exogenously treated and are unrelated with scheduling cost, since the PAT itself is ambiguous without the mechanism underlying the preference decision. The seminal work of [11] first connects the scheduling preference with the utilities generated by the activities conducted before and after the travel. This specification relates the PAT with the marginal utility of being at the base and the destination, which facilitates the analyze of PAT decision behavior as well as more general time-varying marginal utilities. The time-varying marginal utility has been empirically identified among commuters, as evidenced by [12–14]. By elaborating the model of [11], [15] and [16] notes that if the travel duration is fixed and is irrelevant to time, then the traveler will optimally choose to travel between the times when his marginal utility of being at the base equals his marginal utility of being at the destination. [17] provides a useful review on the connection between the time-varying marginal utility specification and the common α-β-γ preference. [18] further associates the agglomeration

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

of co-workers with their marginal utilities, as well as commuters' scheduling preferences. However, none of these literatures deal with the heterogeneity of marginal utilities, and the type of user are also limited to commuter.

Freight carriers differs from the private commuters in at least two ways. First, they tend to use vehicles with lager capacity, collecting several goods for multiple receivers each trip to lower the marginal cost of delivery. Second, the fact that commuters travel time is usually not paid by their employers. Therefore, higher value of travel time per vehicle of freight users than that of private commuters has been empirically found [19]. On the other hand, a freight delivery trip to the city center often ends at an area where a number of receivers spatially close to each other. Due to the limited capacity of urban facilities and the heterogeneity of receivers' preferences, it is usually not possible for the receivers who locate at the city center arrive at their workplace simultaneously. As a result, the numbers of available receivers distribute unevenly over a specific period of time, which will cause the carrier's productivity to vary over time at the destination. This is exactly the same mechanism as the time-varying productivity of commuters with the number of working dependents [20]. It is therefore possible to treat the commuting and freight users by their heterogeneous marginal utilities.

Given these findings, it is now clear that the dynamics of urban road congestion involves both freight and commuting user who are heterogeneous users in terms of their marginal utilities, and that their marginal utilities is likely to vary over time. This naturally raise two questions: First, how the users with heterogeneous time-varying marginal utilities behave when queuing at a bottleneck, compared with homogeneous users? Second, if their behavior affects welfare distribution of the first-best pricing strategy?

To deal with these two questions, this paper introduces the user heterogeneity to the time-varying marginal utility model [11]. The scheduling preference of freight and commuting users is defined by a marginal utility function with a contentiously distributed increasing rate. This methodology is built in light of [16] that shows a way of analyzing the properties of congestion when the users' travel distances to the urban bottleneck are continuously distributed. By applying the model framework of [11], they find that users sort their arrival time at the bottleneck according to the travel distance to the bottleneck, and that the distant users tend to gain more in optimal tolling scheme. Our paper differs from [16] in two major respects. First, it concerns user heterogeneity in marginal utility of time, instead of travel distance. Second, the achieved utility function of user is parametrized, which allows for a continuously distributed rate of marginal utility. We thereby find users sort their arrival time at the bottleneck by the increasing rate of marginal utilities at the destination under both no-toll equilibrium and socially optimal tolling. The social optimal toll removes the queue completely and does not change social welfare when no toll revenue is rebated. The finding implies that equal redistribution to all users can be a Pareto improvement.

In the context of mixed freight and passenger traffic, extending the model of [16] by introducing the heterogeneity of marginal utility also relates with the attitude of freight carrier towards congestion charging. Since the carrier calculates the cost of delivery per trip instead of the number of receivers they serve, the congestion pricing policy with a time-dependent cordon toll, which is the most conventional way in practice, cannot enter the marginal utility of freight carrier. The toll, therefore, becomes a fixed cost and does not vary with the number of receivers per trip. When the charge rate of delivery has to be set to the marginal cost under a complete competitive market, the receivers cannot get any price signal from the cordon toll to choose less congested delivery times [21]. To cope with this fact, [22] proposes a time-distance pricing scheme to show an incentive of off-peak delivery to the carriers, which, however, not applicable when receivers are densely concentrated in the city center. In the setting of the present paper, the above problems are solved by setting the optimal toll to be varying with the amount of available receivers. Indeed, this paper does not intent to discuss the market equilibrium between the freight carriers and the receivers. Throughout the discussion, the charge rate to the receivers is assumed to be fixed. We also limit the freight traffic to light freight vehicles physical externalities (e.g. size and acceleration performance) are similar to private passenger vehicles.

In Section 2, the scheduling model of freight and passenger users is described in a unified formulation. Section 3 shows the conditions when the no-toll equilibrium uniquely exist. Then the socially optimal tolling that exactly removes the equilibrium queue and the corresponding welfare effects are discussed in Section 4. A numerical example is simulated in Section 5, where the distribution of productivity increasing rate is bimodal. The setting of distribution is to reflect the largely differentiated empirical value of travel time and schedule delay time estimated in freight and passenger users. Section 6 concludes the key findings and identifies the possible extension to be introduced. The proofs of propositions are all deferred to the Appendix B.

2 A unified model

Consider that N road users consisting of freight carriers and private commuters travel along a single urban corridor from the suburb to the city center. The freight carrier loads a batch of packages to be delivered to multiple receivers. Commuter each drives a car and travels alone. Before arriving at the destination, they travel without congestion first and then passing through a bottleneck with a constant capacity of φ vehicles per time unit. All users follow the FIFO(first-in-first-out) discipline. The freight carriers and commuters achieve their utility by spending their time with relating parties at their respective bases and the destination. Both freight carriers and commuters incur no further penalties other than the loss utility of time, when their arrivals at the

destination delays. The agglomeration of their respective relating parties varies with time of day, therefore, at either ends of the travel, users' marginal utilities of time (MUT) are time-varying. The exogenous agglomeration magnitudes are assumed to be decreasing at their respective base and increasing at the destination. The receivers locate very closely in the city center, then the travel time between different receivers is ignored. The commuters also do not need additional travel time when interacting with their co-workers.

Both the carriers and commuters desire to make maximum use of their time at both the base and destination within an exogenously given time window, denoted by D, which can be interpreted as the traveling time and the two periods of time adjacent to it. The achieved utility within the time window is denoted by $u(d,t)$, where d and t are respectively the time when the travel starts and ends. Suppose that the users have completely no gain during the time spending on the move and could only achieve utility gain before d and after t. We consider that the road is allowed to be tolled by τ measured in utility unit, and that the toll is not returned to the user. Then an individual user can attain utility of $u(d,t)-\tau$ during D. Since the travel demand is assumed to be fixed, we then define a social welfare function as the total utilities of all users. All other utility losses are assumed to be constant and thus able to be ignored without affecting the result qualitatively.

The marginal utility of time at the base is assumed to be $h(s)=\beta_0+\beta_1 s$, and that at the destination to be $w(s)=\gamma_0+\gamma_1 s$. Then u can be described as

$$\begin{aligned} u(d,t) &= \int_{T_1}^{d}(\beta_0+\beta_1 s)ds+\int_{t}^{T_2}(\gamma_0+\gamma_1 s)ds \\ &= H(d)+W(t) \end{aligned}$$

where $T_1, T_2 \in D$ are constants, $H(d)$ and $W(t)$ are the cumulative utility at the base and the destination respectively. We assume that $H'(d)=\beta_0+\beta_1 d>0, H''(d)<0$, and that $W'(t)=-\gamma_0-\gamma_1 t<0, W''(t)<0$. The achieved utility $u(d,t)$ of a single user is therefore strictly concave.

The users have scheduling preferences that are only heterogeneous in terms of marginal utility of time at the destination. The increasing rate of marginal utility at the destination γ_1 follows a distribution with density f and cumulative distribution F, both of which have support on $\Gamma=[\gamma_L,\gamma_H]$. With the assumptions on $H'(d)$ and $W'(t)$, there must be an intersection of the two marginal utility functions. We define the time when $H'(d)$ and $W'(t)$ intersect as the ideal arrival time(IAT), which is the optimal arrival time when the travel time being reduced to zero. Given we only consider γ_1 varies among users, their IATs are identical.

In order to examine the impacts of tolling on users with heterogeneous MUT, we assume that all the users have to travel the same T time units before arriving at the bottleneck. The arrival time at the bottleneck is denoted by a, and the arrival rate at the bottleneck by $\rho(a)$. When $\rho(a)$ is larger than φ, a vertical queue is generated at the entrance of bottleneck. Let a_0 and a_1 be the time when

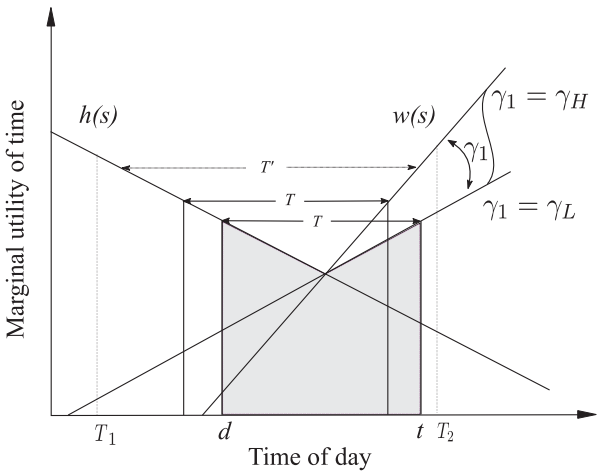

Figure 1. Graphical solution to the utility maximization problem of users with different marginal utility increasing rates

the first and the last user join the queue respectively, then the cumulative arrival $R(a)=\int_{a_0}^{a}\rho(s)da$, and the total travel time of the user who arrives at the bottleneck at time a will be $a_0+\frac{R(a)}{\varphi}-(a-T)$.

A single user with fixed γ_1 travels through the corridor and the bottleneck without congestion can achieve utility $u(a-T,a)$, where a is the time he arrives at the bottleneck as well as the destination. To maximize the utility, he will choose the optimal arrival time $a_*(\gamma_1)$ which solves the utility maximization problem $\arg\max_a u(a-T,a)$. A graphical solution to the maximization problem is illustrated in Figure 1. The shaded area gives the minimum utility loss when the user γ_L schedule his arrival time t at the optimal arrival time $a_*(\gamma_1)$. The optimal arrival time $a_*(\gamma_1)$ is found to have the following property.

Proposition 1 *If* $H'(d)=\beta_0+\beta_1 d>0, W'(t)=-\gamma_0-\gamma_1 t<0$ *and* $H''(d)<0, W''(t)<0$ *hold for all* $t,d\in D, d\le t$, *and if, but not necessarily if,* $\gamma_0-\beta_0\le 0$ *holds, then the optimal arrival time* $a_*(\gamma_1)$ *uniquely exists and* $a'_*(\gamma_1)<0$.

This proposition identifies the feature that travelers sort their optimal arrival times according to the increasing rates of their marginal utilities. The negative $a'_*(\gamma_1)$ indicates that the traveler whose MUT at the destination increases faster will prefer to arrive earlier when traveling without congestion.

In particular, a freight carrier is more likely to gain a higher increasing rate of marginal utility at the destination than a single commuter, owing to the dense agglomeration of payloads in freight vehicles. In the meanwhile, the receivers of the payload are not likely to arrive at the city center simultaneously. Therefore, the agglomeration process of receivers leads to an increase to the marginal utility of shippers. The proposition shows us that the higher increasing rate of marginal utility at the destination is one of the reason that drives carrier to arrive earlier at the bottleneck than commuters. [1] evidences that the peak hour for freight vehicles is typically one hour earlier than for all traffic in the morning with similar travel distance from suburbs to the urban area.

3 No-toll equilibrium

Consider each user schedules his travel by taking the decisions of all other users as given. The conditions when Nash equilibrium exists and the basic properties can be described by the following propositions.

Proposition 2 *Assume that* $a_*(\gamma_1) < \frac{N-F(\gamma_1)}{\varphi} + a_*(\gamma_H)$ *holds for all* $\gamma_1 \in \Gamma, \gamma_1 < \gamma_H$. *When users taking Nash equilibrium strategy, the queue begins and ends during an interval* $[a_0, a_1]$, *where* $a_1 = a_0 + N/\varphi$, *such that* $a_0 \le a_*(\gamma_H)$ *and* $a_1 \ge a_*(\gamma_L)$.

The condition ensures that if the user with the highest increasing rate of marginal utilities arrive the bottleneck at his optimal arrival time, all the other users with lower increasing rate of marginal utilities arrive later than their optimal arrival time. The queue which starts at a_0 and ends at a_1 always exists during $[a_0, a_1]$ where the first and last user experience exactly no queue. The first user arrives at the bottleneck earlier and the last arrives later than his optimal arrival time. This phenomenon can be visualized in Figure 1, when travel time increased from T to T'. We have so far shown the basic queueing property when the heterogeneity of marginal utility exists. The following proposition shows the properties of arrival time, and the analytical solution of scheduling utility.

Proposition 3 *Assume that* $a_*(\gamma_1) < \frac{N-F(\gamma_1)}{\varphi} + a_*(\gamma_H)$ *holds for all* $\gamma_1 \in \Gamma, \gamma_1 < \gamma_H$. *The no-toll equilibrium exists and is unique. Users who derive increasing marginal utility at rate* γ_1 *arrive at the bottleneck at time* $a(\gamma_1)$ *and its derivative satisfies*

$$a'(\gamma_1) = -\frac{f(\gamma_1)}{\varphi}\frac{\gamma_0 + \gamma_1\left(a_0 + \frac{N-F(\gamma_1)}{\varphi}\right)}{\beta_0 + \beta_1(a-T)} < 0. \qquad (1)$$

The scheduling utility under equilibrium is

$$u\left(a - T, a_0 + \frac{N-F(\gamma_1)}{\varphi}\right)$$

The proposition shows that the arrival time a at the destination is a function of the increasing rate γ_1 of marginal utility at the destination under no-toll equilibrium. The derivative of a is strictly negative which indicates that those who gain higher increasing rate of marginal utility at the destination arrive earlier. Given the assumption of FIFO(first-in-first-out) discipline, the bottleneck preserves the arriving order at the entrance. The time of their arrival at the destination is also in the order of their γ_1 at the destination.

According to the arrival schedule, the first user who arrives at bottleneck as well as the destination at time $a(\gamma_H) = a_0$ has the highest increasing rate γ_H. The user derives an increasing rate γ_1 of his marginal utility at the destination arrives at the destination after $N - F(\gamma_1)$ users at the bottleneck at time $a_0 + (N - F(\gamma_1))/\varphi$, while those who gain lower increasing rate of the marginal utility arrive later. Lastly, the user with the lowest increasing rate γ_L of marginal utility joins the queue, and arrives at the destination at time $a_1 = a_0 + N/\varphi$.

4 Optimal tolling

In this section, we first find the socially optimal toll $\tau(a)$, then discuss the welfare effects by the optimal toll on users with different increasing rates of MUT.

Proposition 4 *A socially optimal toll exists if, but not necessarily if,* $a'_*(\gamma_1) < -\frac{f(\gamma_1)}{\varphi}$. *The arrival at the bottleneck begins at* $a_{\tau 0}$ *which solves*

$$0 = \int_{a_{\tau 0}}^{a_{\tau 0}+N/\varphi} [\beta_0 + \beta_1(a-T) - \gamma_0 - \gamma_1 a]\, da.$$

The optimal toll completely removes the queue, which satisfies

$$\tau'(a_\tau(\gamma_1)) = \beta_1(a_\tau(\gamma_1) - T) + \beta_0 - \gamma_1 a_\tau(\gamma_1) - \gamma_0,$$

and

$$\tau(a_{\tau 0}) = \tau(a_{\tau 1}).$$

$\tau(a_{\tau 0}) = 0$ *gives one of the optimal tolls. The users arrive the bottleneck and the destination simultaneously at*

$$a_\tau(\gamma_1) = \frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0}.$$

The order of the arrival at the destination remain unchanged relative to no-toll equilibrium.

Infinitely number of optimal tolls can exist when the demand elasticity is ignored, where $\tau(a_{\tau_0}) = 0$ is one of them. The proof shows that the optimal toll keeps the sorting property of arrival under no-toll equilibrium unchanged. The formulation of arrival time at the destination is also identical to the one under no-toll equilibrium.

Along with the commuters, the carriers are also facing the optimal toll that varies with the increasing rate of MUT at the destination γ_1. The toll now enters the marginal cost when delivering an additional order to a receiver. It is then of great interest to discuss if anyone is getting better or worse off by tolling when this important barrier for freight users to accept the toll is removed. The next proposition also answers the question if the arrival interval is able to be shifted by the optimal toll.

Proposition 5 *With the conclusion of , consider that an socially optimal toll can be achieved by* $\tau(a_{\tau 0}) = \tau(a_{\tau 1}) = 0$. *The arrivals in social optimum are the same as that under no-toll equilibrium:* $a_\tau(\gamma_1) = a(\gamma_1)$. *Relative to the no-toll equilibrium, the socially optimal toll does not change each user's indirect utility, when the toll revenues are not redistributed to the users.*

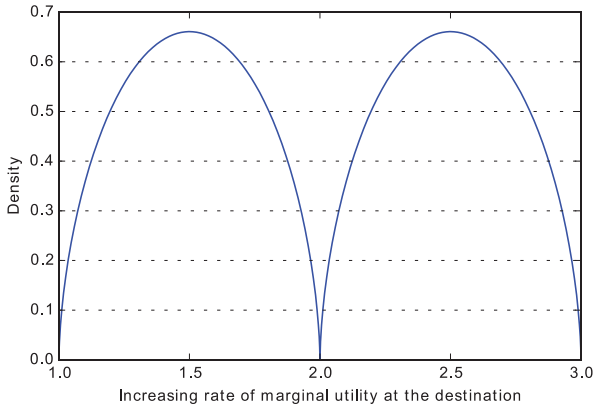

Figure 2. The density of increasing rates of marginal utility at the destination.

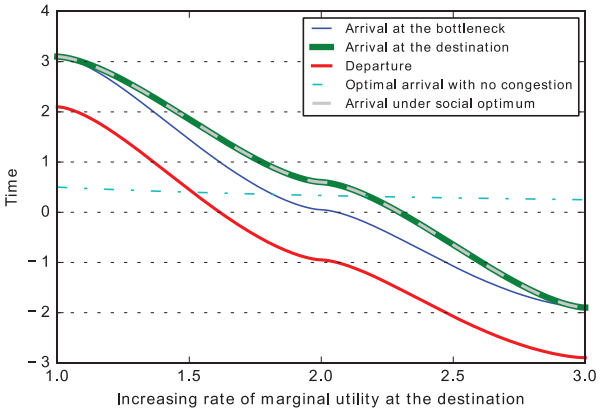

Figure 3. Departure and arrival times under no-toll equilibrium and social optimum

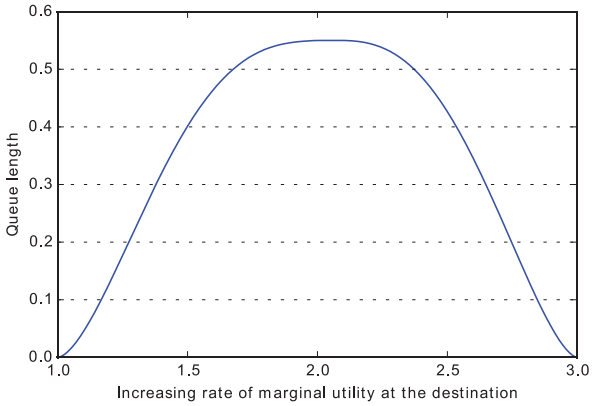

Figure 4. Queue length

Under the no-toll equilibrium, the user with highest MUT increasing rate at the destination departs and arrives first to avoid the queue, and the one with lowest MUT increasing rate arrives last. If the queue were removed, the first user could arrive later, and the last one earlier. However, a social optimal toll that removes the queue requires the travelers pay exactly the same monetary cost as the scheduling utility gain. This explains the reason why the arrival interval under socially optimal tolling remains unchanged. This observation is very similar to the conventional bottleneck model with homogeneous user [23] and the one with heterogeneous users whose cost of arriving late differs [6].

It is also worthy to relate the current finding with that of [16]. The heterogeneity of distance to the bottleneck in the model of [16] causes unequal gain and loss under the first-best tolling scheme. The first-best pricing scheme under their setting does not equal the time saving benefit. User who travels longer distance to the bottleneck whose MUT at the destination is higher than those short distance traveler, therefore they benefit more from the elimination of the queue.

5 Numerical example

A numerical simulation is presented here to visualize the theoretical model by freight and passenger users with continuously distributed increasing rates of MUT at the destination. The users are assumed to be a continuum with mass 1. The schedule preference is specified by the following achieved utility:

$$u(d,t) = \int_{T_1}^{d} (\beta_0 + \beta_1 s)ds + \int_{t}^{T_2} (\gamma_0 + \gamma_1 s)ds,$$

where β_0 is assumed to be equal to γ_0 without losing generality. T_1 and T_2 are arbitrarily set to 0 without affecting the result in qualitative. The capacity of bottleneck φ is 0.02 users per minute, which means all users can pass the bottleneck in 50 minutes.Free flow travel time T before arriving at the bottleneck is arbitrarily set to 1 minute. The increasing rates γ_1 of MUT at the destination follow a bimodal distribution, composed of two beta distributions, each with mass 0.5. One of the two has support on [1,2] and the other on [2,3]. The density of γ_1 is shown by Figure 2.

The simulation calculates for each γ_1, then performs a search of a_0 by solving $a_1 = a_0 + N/\varphi$. Under social optimum, $a_{\tau 0}$ is found numerically to maximize mean scheduling utility as well as social welfare. Simulation result of arrival times are illustrated in Figure 3.

The dot-dash line shows the situation when all users travel without congestion where only free flow travel time T is needed. As states, the optimal arrival times are sorted by the descending order γ_1. Since β_0 is assumed to be equal to γ_0, then $a_*(\gamma_1) = \frac{\beta_1 T}{\beta_1 - \gamma_1}$. The first user with the highest γ_1 schedules his individual optimal arrival time at $a_*(\gamma_H) = 0.25$.

Under no-toll equilibrium, as shown by the upper thick line, the user with the highest increasing rate of MUT at the destination shifts his arrival time to a_0 to avoid the queue, which is earlier than his optimal arrival time $a_*(\gamma_H)$. He arrives at the destination as soon as he gets to the bottleneck, as there is no queue for him. At the other end of the queue, the user gains the lowest increasing rate γ_L of MUT arrives last at time $a_1 > a_*(\gamma_L)$. He also experiences exactly no queue either. The thin line is the arrival time at the bottleneck, while the lower thick line gives departure time from their bases. Since all user are supposed to travel the same time before arriving at the bottleneck, the departure time function has the same derivative as the arrival time at the bottleneck, and the traveler with lower γ_1 always departs later.

Figure 4 gives a closer look at the queue length as a function of γ_1. The density of γ_1 is lower than the bottleneck capacity near $\gamma_1 = 2$. But the two curves $a_0 + \frac{N-F(\gamma_1)}{\varphi}$ and $a_*(\gamma_1)$ still intersect only once, which means the density is not sparse enough to evolve another peak.

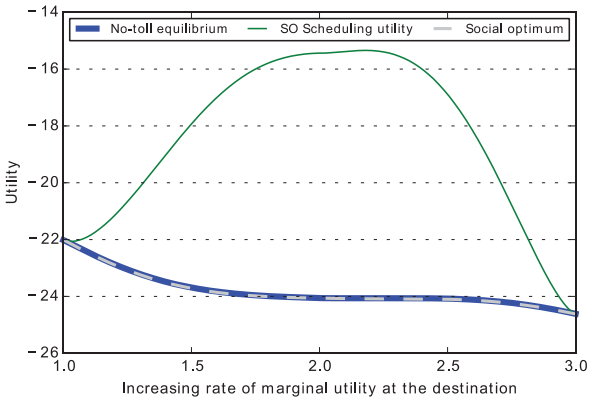

Figure 5. Utilities of users with different increasing rate of marginal utilities at the destination

Next consider the situation under social optimum. Since the functional form $\frac{N-F(\gamma_1)}{\varphi} + a_{\tau 0}$ remain unchanged, the order of arrival at the destination is identical to that in no-toll equilibrium. Moreover, the constant $a_{\tau 0}$ does not shift the arrival interval either earlier or later, compared with that of no-toll equilibrium. Therefore, the arrival time under social optimum shown by the upper dash line in Figure 3 coincides with the arrival time under no-toll equilibrium. Consequently, the departure time in social optimum does not move.

Figure 5 presents the achieved utilities of users with different increasing rates of MUT at the destination. The indirect utility under no-toll equilibrium shown by the lower thick curve consists only scheduling utility. The scheduling utility in social optimum is shown by the upper thin line. The lower dash line gives the indirect utility in social optimum, which exactly coincides with the indirect utility under no-toll equilibrium. The difference between them is thus the socially optimal toll. As stated in Proposition 5, the welfare gain in terms of scheduling utility is completely offset, and no user can benefit from the tolling, when the toll revenues are not returned to the user.

It is noteworthy that the indirect utilities under both no-toll equilibrium and social optimum are a monotonically decreasing function of γ_1. The users with higher productivity increasing rate achieve less utility instead of more, as they incur higher scheduling cost. It implies that a higher rebate to those more productive users can motivate them to improve marginal utility.

6 Conclusion

This paper introduces the heterogeneity of preference to the time-varying marginal utility scheduling model. The model is then applied to treat the dynamics of urban traffic congestion incurred by both freight and passenger traffic. By directly relating the productivity of freight carriers to the amount of toll, the proposed optimal tolling removes the fundamental barrier for freight carriers to accept congestion pricing. The first-best pricing strategy without rebate is found inducing no welfare gain or loss to any of the users, in spite of the heterogeneity of marginal utilities. This finding gives an important policy implication that an equal revenue rebate of a social optimal toll can generate a Pareto improvement, relative to the no-toll equilibrium.

Our interesting result is that the user who gains higher productivity increasing rate failed to achieve higher utility, which indicates that a differentiated redistribution of toll revenues can be an incentive to improve productivity. This shows the possibility of two interesting extensions to our current model. The first would be to examine the second-best pricing policy to freight and passenger traffic, for example in the form of freight dedicated lane [24] or pricing a portion of lanes [25]. The second is to make the productivity increasing rate in our model endogenous. Compared with commuters, it is more possible for freight carriers to decide their marginal utilities at the destination, as they could choose an appropriate load per trip. By capturing this responsive behavior, the dynamics of congestion and the effect of tolling policy can be more comprehensively examined.

Appendix A. Main notations

Symbol	Description
PAT	Exogenous preferred arrival time
N	Total number of road users
φ	Capacity of the bottleneck in the city center
MUT	Marginal utility of time
D	A specific period of time within a day
T	Travel time from the base to the bottleneck
u	Achieved utility
τ	Time-varying toll at the bottleneck
h	Marginal utility of time at the base
w	Marginal utility of time at the destination
β_0	Constant of the marginal utility of time at the base
β_1	Increasing rate(slope) of MUT at the base
γ_0	Constant of the marginal utility of time at the destination
γ_1	Increasing rate(slope) of MUT at the destination
H	Cumulative utility at the base
W	Cumulative utility at the destination

IAT	Ideal arrival time when travel time is zero
a	Arrival time at the bottleneck without tolling
d	Departure time from the base
t	Arrival time at the destination
ρ	Arrival rate at the bottleneck
R	Cumulative arrival at the bottleneck
a_*	Optimal arrival time without queueing
$f(\gamma_1)$	Density of γ_1
$F(\gamma_1)$	Cumulative distribution of γ_1
γ_H	Upper bound of γ_1
γ_L	Lower bound of γ_1
a_0	Time when the queue begins
a_1	Time when the queue ends
a_τ	Arrival time at the bottleneck under socially optimal tolling

Appendix B. Lemmas and proof of propositions

Proof of Proposition 1 The strictly convexity assumption on u assures that the optimal arrival time $a^*(\gamma_1)$ uniquely exists. Here the the first order condition for the utility maximization is that

$$0 = \beta_0 + \beta_1(a - T) - \gamma_0 - \gamma_1 a. \qquad (2)$$

Differentiate the first order condition with respect to γ_1 to find that

$$a'_*(\gamma_1) = \frac{\beta_1 T + \gamma_0 - \beta_0}{(\beta_1 - \gamma_1)^2}.$$

Note that $H''(d) = \beta_1 < 0$, and $\gamma_0 - \beta_0 \le 0$, then $a'_*(\gamma_1)$ must be negative.

Lemma 1 *If $a_0 \le a_*(\gamma_H), a_0 + N/\varphi = a_1 \ge a_*(\gamma_L)$, and if there is no queue at time a_1, then taking the behavior of all other users as given, any user will choose to arrive at the bottleneck in the interval $[a_0, a_1]$.*

Consider an arbitrary queue starts at a_0 and ends at a_1, and a user with an increasing marginal utility at the destination at rate $\gamma_1 \in \Gamma$, where $a_0 \le a_*(\gamma_H) \le a_*(\gamma_1) \le a_*(\gamma_L) \le a_1$. Then the user prefers to arrive at the bottleneck at a_0 to any other earlier time, since earlier arrival time will not reduce the queueing time anymore but decrease the utility. He also prefers a_1 to any other time later, since the queue ends at time a_1. Therefore there will not be any one choosing to arrive outside the interval $[a_0, a_1]$.

Lemma 2 When $a_*(\gamma_1) < \frac{N - F(\gamma_1)}{\varphi} + a_*(\gamma_H)$ holds for all $\gamma_1 \in \Gamma, \gamma_1 < \gamma_H$, the equilibrium arrive time as a function of γ_1 is given by (1) and

$$-\int_{\gamma_L}^{\gamma_H} a'(\gamma_1) d\gamma_1 = \frac{N}{\varphi}. \qquad (3)$$

Then $a(\gamma_1)$ satisfies $a(\gamma_H) \le a_*(\gamma_H)$ and $a(\gamma_L) \ge a_*(\gamma_L)$.

Given the conclusion of , $a_*(\gamma_1) < 0$. Then if any user arrives later than his optimal arrival time, his marginal utility at the destination when he arrives must be larger than that at the base when he departs, which implies $a'(\gamma_1) > -\frac{f(\gamma_1)}{\varphi}$. If $a(\gamma_1) > a_*(\gamma_1)$ for all γ_1, then $a_1 - a_0 > N/\varphi$, which contradicts (3).

Suppose $a(\gamma_H) > a_*(\gamma_H)$. There exists $\xi < \gamma_H$ such that $a_(\xi) = a_*(\xi)$. Then for all $\gamma_1 > \xi$, $a(\gamma_1) > a_*(\gamma_1)$ and $a'(\gamma_1) > -\frac{f(\gamma_1)}{\varphi}$ hold. Therefore by $a_*(\gamma_1) < \frac{N - F(\gamma_1)}{\varphi} + a_*(\gamma_H)$, we have $a(\xi) > a_0 + \frac{N - F(\gamma_1)}{\varphi} > a_*(\gamma_H) + \frac{N - F(\gamma_1)}{\varphi} > a_*(\xi)$, which is a contradiction to $a_(\xi) = a_*(\xi)$.

Then assume that $a_1 < a_*(\gamma_L)$. If $a_0 + N - F(\gamma_1)/\varphi < a_*(\gamma_1)$ holds for all γ_1, then $a'(\gamma_1) < -\frac{\gamma_1}{\varphi}$ for all γ_1, which contradicts (3). So there exists ζ such that $a_0 + \frac{N - F(\zeta)}{\varphi} = a_*(\zeta)$. Then by $a_1 - a_1 = N/\varphi$ and $a_*(\gamma_1) < \frac{N - F(\gamma_1)}{\varphi} + a_*(\gamma_H)$, we have $\frac{N}{\varphi} - \frac{N - F(\zeta)}{\varphi} = a_1 - a_*(\zeta) < a_*(\gamma_L) - a_*(\zeta) < \frac{N}{\varphi} - \frac{N - F(\zeta)}{\varphi}$, which is a contradiction.

Therefore $a(\gamma_H) \le a_*(\gamma_H)$ and $a(\gamma_L) \ge a_*(\gamma_L)$ must hold.

Proof of Proposition 2 Let $[a_0, a_1]$ be the minimum arrival interval required by all N users in no-toll equilibrium. By Proposition 1 , $a_*(\gamma_1) < 0$ for all γ_1, then $a_0 \le a_*(\gamma_1)$ for all γ_1, otherwise there could be some users who are supposed to arrive within interval $[a_0, a_1]$ can shift their arrive to an earlier time at their optimal arrival time without queueing and then increase utility. At the end of the queue, one could similarly have $a_1 \ge a_*(\gamma_1)$ for all γ_1. Therefore, $a_0 \le a_*(\gamma_L)$ and $a_1 \ge a_*(\gamma_H)$.

Next we show that the length of the interval when no-toll equilibrium must be N/φ. If $a_1 - a_0 > N/\varphi$, there must be a period of time when the bottleneck is not fully utilized. Given $a_*(\gamma_1) < \frac{N - F(\gamma_1)}{\varphi} + a_*(\gamma_H)$, someone who is supposed to arrive later in the period can shift to an earlier arrival time when the bottleneck was not fully utilized, and this will lead to strict increase of utility and contradict the equilibrium. If $a_1 - a_0 < N/\varphi$, there will be a residual queue at a_1 (see [26] for further discussion on the condition when the property of no residual queue holds). The user in the residual queue can arrive later at the bottleneck to reduce queueing time and arrive the destination at the same time as they wait in the residual queue.

Proof of Proposition 3 To prove $a'(\gamma)$ has a unique solution, we first assume the equilibrium exists for the

ease of discussion. As shown by Proposition 2, the queue starts at a_0 and ends at a_1. Given there will always be queue during $[a_0, a_1]$, for a user with increasing rate of marginal utility γ_1, the achieved utility can be represented by

$$u\left(a - T, \frac{R(a)}{\varphi} + a_0\right) = \int_0^{a-T} (\beta_0 + \beta_1 s)ds + \int_{\frac{R(a)}{\varphi}+a_0}^{0} (\gamma_0 + \gamma_1 s)ds$$

Then the first order condition of maximizing the utility of the individual is

$$0 = \beta_1(a - T) + \beta_0 - \left[\gamma_0 + \gamma_1\left(\frac{R(a)}{\varphi} + a_0\right)\right]\frac{\rho(a)}{\varphi} \quad (4)$$

and the second order condition is

$$\beta_1 - \gamma_0\frac{\rho'(a)}{\varphi} - \gamma_1\frac{R(a)\rho(a)}{\varphi^2} - \gamma_1\left(\frac{\rho(a)}{\varphi}\right)^2 - \gamma_1\frac{\rho'(a)}{\varphi}a_0 < 0$$

Then the equilibrium utility $u\left(a(\gamma_1) - T, \frac{R(a(\gamma_1))}{\varphi} + a_0\right)$ can be achieved when the arrival time at the bottleneck $a(\gamma_1)$ solves the utility maximization problem.

To see how equilibrium utility varies with the increasing rate of marginal utility, we could differentiate u with respect to γ_1, which yields

$$\frac{\partial}{\partial\gamma_1}u\left(a(\gamma_1) - T, \frac{R(a(\gamma_1))}{\varphi} + a_0\right) = a'(\gamma_1)\beta_1(a(\gamma_1) - T) - \gamma_1\left(a_0 + \frac{R(a)}{\varphi}\right)\frac{\rho(a)}{\varphi} - \frac{1}{2}\left(a_0 + \frac{R(a)}{\varphi}\right)^2 - \gamma_0\frac{\rho(a)}{\varphi}. \quad (5)$$

By the concavity assumptions on u, (5) can vary from negative to positive, which relates with cumulative arrival $R(a)$.

To examine the property of $a'(\gamma_1)$, we differentiate the first order condition (4) with respect to γ_1, which yields

$$0 = a' \cdot u''(a) - \frac{\rho(a)}{\varphi}\left(\frac{R(a)}{\varphi} + a_0\right).$$

Since u is assumed to be strictly concave, we have $u''(a) < 0$, and by $\rho(a) > 0$, therefore $a'(\gamma_1) < 0$. Note that $a(\gamma_1)$ has an inverse $\gamma_1(a)$, which gives $a(\gamma_1(a)) = a$, we then have $\gamma_1'(a) = 1/a'(\gamma_1) < 0$. Therefore $R(a) = N - F(\gamma_1)$. Differentiate $R(a)$ with respect to a shows that

$$\rho(a) = -\frac{f(\gamma_1)}{a'(\gamma_1)}. \quad (6)$$

Substitute $\rho(a)$ into the first order condition (4), we find that

$$a'(\gamma_1) = -\frac{f(\gamma_1)}{\rho(a)} = -\frac{\gamma_0 + \gamma_1\left(a_0 + \frac{N - F(\gamma_1)}{\varphi}\right)}{\beta_0 + \beta_1(a - T)}\frac{f(\gamma_1)}{\varphi}.$$

Two conditions must hold for the existence of equilibrium: (a) No one can improve his utility by unilaterally changing his arrival time at the bottleneck, and (b) all users must arrive. Since the user's individual scheduling utility u is assumed to be concave, a user has a unique maximum scheduling utility taking all others behavior as given. In the meanwhile, the arrival rate at the bottleneck given by (1) and $\int_{\gamma_H}^{\gamma_L} a'(\gamma_1)d\gamma_1 = N/\varphi$ has the unique solution, which does not depend on the existence of equilibrium. If someone is able to increase individual utility by changing his arrival time at the bottleneck, then $a'(\gamma_1)$ cannot have an unique solution, which contradicts to (1). Therefore, condition (a) must hold. Condition (b) requires that $a(\gamma_L) = a_0 + N/\varphi$. If a_0 exists and is not unique, $a'(\gamma_1)$ will either change non-monotonically or remain unchanged when a_0 changes. Then by differentiating $a'(\gamma_1)$ with respect to a_0, we have

$$\frac{\partial a'(\gamma_1)}{\partial a_0} = \left(-\frac{f(\gamma_1)}{\varphi}\right)\left[\frac{\gamma_1[\beta_1(a - T) + \beta_0] - \beta_1\frac{\partial a(\gamma_1)}{\partial a_0}\left[\gamma_1\left(a_0 + \frac{N - F(\gamma_1)}{\varphi}\right) + \gamma_0\right]}{[\beta_1(a - T) + \beta_0]^2}\right]. \quad (7)$$

By the assumptions on u, $\frac{\partial a\prime(\gamma_1)}{\partial a_0}$ is strictly negative if $\frac{\partial a(\gamma_1)}{\partial a_0} \geq 0$. Note that $\frac{\partial a(\gamma_1)}{\partial a_0} = \frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} + 1$, and that

$$\frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} = \int_{\gamma_H}^{\gamma_1}\frac{\partial a'(\gamma_1)}{\partial a_0}d\gamma_1.$$

If $\frac{\partial(a(\xi) - a_0)}{\partial a_0} \geq 0$ holds for all $\xi > \gamma_1$ can be proved, then $\frac{\partial a\prime(\xi)}{\partial a_0} < 0$ would hold for all $\xi > \gamma_1$ by the above strictly negative condition of (7). Note that $a'(\gamma_1)$ is continuous, since $[a_0, a_1]$ is the smallest interval that contains all the arrivals at the bottleneck, then $\frac{\partial a\prime(\xi)}{\partial a_0}$ must be continuous as well. At point $\xi = \gamma_1$, $\frac{\partial a\prime(\gamma_1)}{\partial a_0} \leq 0$ must hold. Therefore we could have $\frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} = \int_{\gamma_H}^{\xi}\frac{\partial a\prime(\xi)}{\partial a_0}d\xi + \int_{\xi}^{\gamma_1}\frac{\partial a\prime(\gamma_1)}{\partial a_0}d\gamma_1 > 0$.

Substitute $\frac{\partial(a(\gamma_H) - a_0}{\partial a_0} = 0$ into , we have $\frac{\partial a\prime(\gamma_H)}{\partial a_0} < 0$. Then there exist a small neighborhood $V(\gamma_H)$ around γ_H such that $\frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} \geq 0$ for all $\gamma \in V(\gamma_H)$. By the above discussion, therefore, $\frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} > 0$ holds for all $\gamma_1 < \gamma_H$, where $\frac{\partial(a(\gamma_L) - a_0)}{\partial a_0} > 0$. So there is only one a_0 that satisfies $a(\gamma_L) - a_0 = N/\varphi$, which ensures that only one equilibrium can exist.

Next we show there indeed exists a_0 such that $a(\gamma_L) = a_0 + N/\varphi$, then condition b holds. By Proposition 2, $a_0 \leq a_*(\gamma_H)$, $a_1 \geq a_*(\gamma_L)$, and $a_1 - a_0 = N/\varphi$ hold under no-toll equilibrium. Then $a(\gamma_1) \leq \frac{N - F(\gamma_1)}{\varphi}$ must holds, since there is always a queue during $[a_0, a_1]$. Here we then discuss the situation when a_0 varies outside the above range.

Consider first $a_0 + N/\varphi < a_*(\gamma_H)$. By the above discussion, $\frac{\partial(a(\gamma_1) - a_0)}{\partial a_0} = 1 + \frac{\partial(a(\gamma_1 - a_0))}{\partial a_0} > 1$ holds for all $\gamma_1 < \gamma_H$, and $\frac{\partial a(\gamma_H)}{\partial a_0} = 1 + \frac{\partial(a(\gamma_H) - a_0)}{\partial a_0} = 1$. Therefore when $a_0 < a_*(\gamma_H) - N/\varphi$, we have

$$a'(\gamma_1) = -\frac{\gamma_0 + \gamma_1\left(a_0 + \frac{N - F(\gamma_1)}{\varphi}\right)}{\beta_0 + \beta_1(a - T)}\frac{f(\gamma_1)}{\varphi} < -\frac{f(\gamma_1)}{\varphi}$$

and

$$\frac{\gamma_0 + \gamma_1\left(a_0 + \frac{N - F(\gamma_1)}{\varphi}\right)}{\beta_0 + \beta_1(a - T)} > 1,$$

then

$$\int_{\gamma H}^{\gamma_L} a'(\gamma_1)d\gamma_1 > \int_{\gamma H}^{\gamma_L} -\frac{f(\gamma_1)}{\varphi}d\gamma_1 = \frac{N}{\varphi}.$$

If $a_0 > a_*(\gamma_L)$, we could otherwise obtain $\int_{\gamma_H}^{\gamma_L} a'(\gamma_1)d\gamma_1 < N/\varphi$. Then $a(\gamma_L) - a(\gamma_H)$ can achieve both larger and smaller values than N/φ, which establishes the existence of equilibrium.

By Lemma 2, when a_0 attains $a_1 - a_0 = N/\varphi$, $a_0 < a_*(\gamma_H)$ and $a_1 > a_*(\gamma_L)$ hold. Then by Lemma 1, no one will choose the arrival time outside the interval $[a_0, a_1]$. Together with Lemma 1 and the previous discussion that (7) is strictly negative, it is now sufficient to conclude that (1) defines a unique equilibrium that no one can improve his utility by unilaterally changing arrival time at the bottleneck within the interval $[a_0, a_1]$.

Proof of pproposition 4 The socially optimal toll exactly removes the queue. If there is a queue at some point of time, the user will delay the arrival and social welfare is then improved. Therefore we have $R_\tau(a) = \varphi(a - a_{\tau 0})$ and $\rho_\tau = \varphi$. Under the optimal tolling, a user with the increasing rate γ_1 of marginal utility arrives at the bottleneck at time $a_\tau(\gamma_1)$. The first order condition for utility maximization is

$$\tau'\big(a_\tau(\gamma_1)\big) = \beta_1(a_\tau - T) + \beta_0 - \gamma_1 a_\tau - \gamma_0, \qquad (8)$$

and the corresponding second order condition is

$$\tau''\big(a_\tau(\gamma_1)\big) > \beta_1 - \gamma_1.$$

Differentiating the first order condition with respect to γ_1 yields

$$0 = \tau'' - \beta_1 a_\tau' + \gamma_1 a_\tau' + a_\tau.$$

By the second order condition and the assumptions of $\gamma_1 > 0 > \beta$, $a_\tau'(\gamma_1)$ must be negative, which means the order of arrival will remain unchanged compared with the one under no-toll equilibrium. Then we have

$$R_\tau\big(a_\tau(\gamma_1)\big) = N - F(\gamma_1) = \varphi(a_\tau(\gamma_1) - a_{\tau 0}),$$

where $a_\tau(\gamma_1) = \frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0}$. To find the solution to an optimal toll, note first that the inverse of a_τ is $\gamma_{1\tau}(a) = F^{-1}(N - \varphi(a - a_{\tau 0}))$. Substituting $\gamma_{1\tau}$ into the first order condition (8) yields the derivative of an optimal toll $\tau'(a)$ which satisfies

$$\tau'(a) = \beta_1(a - T) + \beta_0 - F^{-1}(N - \varphi(a - a_{\tau 0}))a - \gamma_0,$$

such that

$$\tau(a) - \tau(a_{\tau 0}) = \int_{a_{\tau 0}}^{a} \tau'(s)ds = \int_{a_{\tau 0}}^{a} \beta_1(s - T) - F^{-1}(N - \varphi(s - a_{\tau 0}))sds.$$

Any toll satisfies (9) will exactly remove the queue. Users under such queue arrives at the bottleneck as well as at the destination within the interval $[a_{\tau 0}, a_{\tau 1}]$. Then we have the mean scheduling utility under optimal tolling

$$Eu = \int_{\gamma_L}^{\gamma_H} u\Big(\frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0} - T, \frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0}\Big) f(\gamma_1)d\gamma_1$$

Differentiate the mean scheduling utility with respect to a_0 yields

$$\frac{\partial Eu}{\partial a_0} = \int_{\gamma_L}^{\gamma_H} \begin{bmatrix} \beta_0 + \beta_1\left(\frac{N - F(\gamma_1)}{\varphi} + a_\tau 0 - T\right) \\ -\gamma_0 - \gamma_1\left(\frac{N - F(\gamma_1)}{\varphi} + a_\tau 0\right) \end{bmatrix} f(\gamma_1)d\gamma_1$$

$$= \varphi \int_{a_{\tau 0}}^{a_{\tau 1}} (\beta_0 + \beta_1(a - T) - \gamma_0 - \gamma_1 a)\, da, \qquad (10)$$

and the second order derivative is

$$\frac{\partial^2 Eu}{\partial (a_0)^2} = \int_{\gamma_L}^{\gamma_H} (\beta_1 - \gamma_1) f(\gamma_1)d\gamma_1.$$

Since $H' > 0 > W'$, there must exist γ_1 such that (10) equals zero, and (11) is strictly negative. Therefore Eu is strictly concave as a function of a_0. When $\frac{\partial Eu}{\partial a_0} = 0$, Eu reaches a global maximum, where the start time of the arrival under an optimal toll $a_{\tau 0}$ is the unique solution. Then we rewrite $\frac{\partial Eu}{\partial a_0} = 0$ as

$$0 = \int_{a_{\tau 0}}^{a_{\tau 0} + \frac{N}{\varphi}} (\beta_0 + \beta_1(a - T) - \gamma_0 - F^{-1}(N - \varphi(a - a_{\tau 0}))a)\, da.$$

Substitute (9) in to the above equation, and denote $a_{\tau 0} + N/\varphi$ by $a_{\tau 1}$, then we have $\tau(a_{\tau 0}) = \tau(a_{\tau 1})$.

If the toll satisfies $\tau(a_{\tau 0}) = \tau(a_{\tau 1}) = 0$, then by Lemma 1 in the appendix, we could eliminate the possibility that users may choose an arrival time outside the tolling interval. To see the reason, note first that $\beta_1 - \gamma_1 < 0$ by the assumption of the model. Then by (2) and (8), it is clear that $a(\gamma_1) \gtreqless a_*(\gamma_1)$ is a sufficient and necessary condition of $\tau'(a) \lesseqgtr 0$. If $'_*(\gamma) < -\frac{f(\gamma)}{\varphi}$, there must exist only one $\zeta \in \Gamma$ such that $a(\zeta) = a_*(\zeta)$.Therefore $a_{\tau 0} \le a_*(\gamma_H), a_*(\gamma_L) \le a_{\tau 1}$.

Proof of Proposition 5 The utility difference between the case with and without optimal tolling of a user with MUT increasing rate at γ_1 can be written as

$$\Delta u(\gamma_1) = u(a_\tau(\gamma_1) - T, a_\tau(\gamma_1)) - \tau(a_\tau(\gamma_1)) - u\left(a(\gamma_1) - T, \frac{N - F(\gamma_1)}{\varphi} + a_0\right).$$

By differentiating $\Delta u(\gamma_1)$ with respect to γ_1, and substiute the first order condition under no-toll equilibrium and under socially optimal tolling, we have

$$\frac{\partial \Delta u(\gamma_1)}{\partial \gamma_1} = -\frac{1}{2}\Big(\frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0}\Big)^2 + \frac{1}{2}\Big(\frac{N - F(\gamma_1)}{\varphi} + a_0\Big)^2.$$

Suppose that $a_{\tau 0} < a_0$. Then if $\frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0} < \frac{N - F(\gamma_1)}{\varphi} + a_0 < 0$ or $\frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0} < -\left(\frac{N - F(\gamma_1)}{\varphi} + a_0\right) < 0$, for all γ_1, $\frac{\partial \Delta u(\gamma_1)}{\partial \gamma_1} < 0$ must hold. Then the two curve $\frac{N - F(\gamma_1)}{\varphi} + a_{\tau 0}$ and $\frac{N - F(\gamma_1)}{\varphi} + a_0$ must intersect at some

$\xi \in \Gamma$. Therefore for any $\gamma_1 > \xi$, $\frac{N-F(\gamma_1)}{\varphi} + a_{\tau 0} > \frac{N-F(\gamma_1)}{\varphi} + a_0$ must hold, which is a contradiction. If $0 < \frac{N-F(\gamma_1)}{\varphi} + a_{\tau 0} < \frac{N-F(\gamma_1)}{\varphi} + a_0$, or $\frac{N-F(\gamma_1)}{\varphi} + a_{\tau 0} < 0, \frac{N-F(\gamma_1)}{\varphi} + a_0 > 0$ and $\frac{N-F(\gamma_1)}{\varphi} + a_{\tau 0} > -\left(\frac{N-F(\gamma_1)}{\varphi} + a_0\right) < 0$, then $\frac{\partial \Delta u(\gamma_1)}{\partial \gamma_1} < 0$ must hold for all γ_1. Therefore $a_{\tau 1} < a_1$, and both $\Delta u(\gamma_L)$ must be negative, which contradicts Proposition 2.

The contradictions can be similarly found if $a_{\tau 0} > a_0$. This then ensures $a_{\tau 0} = a_0$ must hold.

References

1. P. Vilain, P. Wolfrom, Transp. Res. Rec. J. Transp. Res. Board, **1707**, 64–72 (2000)
2. T. Kenmochi, H. Oka, R. Tani, D. Fukumoto, Y. Hagino, T. Hyodo, *IBS Annual Report* (2016).
3. MDS Transmodal Limited, Chester, UK., DG MOVE European Commission: Study on urban freight transport, Final report (2012) [Accessed: 17-May-2016]
4. Beijing Municipal Commission of Transport, China, Institute of Transportation Science, Ministry of Transport, China, and Beijing Municipal Bureau of Statistics, China, Report on urban freight demand in central Beijing (in Chinese) (2011) [Accessed: 17-May-2016]
5. W. S. Vickrey, *Am. Econ. Rev.*, **59**, 2, 251–260 (1969)
6. R. Arnott, A. de Palma, and R. Lindsey, *J. Transp. Econ. Policy*, **28**, 2, 139–161 (1994)
7. V. A. C. van den Berg, E. T. Verhoef, *Transp. Res. Part B Methodol.*, **45**,1, 60-78 (2011)
8. R. Lindsey, *Transp. Sci.*, **38**, 3, 293–314, (2004)
9. K. A. Small, *Am. Econ. Rev.*, **72**, 3, 467–479, (1982)
10. S. Mun, M. Yonekawa, *J. Transp. Econ. Policy*, **40**, 3, 329–358 (2006)
11. W. Vickrey, *Highw. Res. Rec.*, 476, 36–48 (1973)
12. Y.-Y. Tseng, E. T. Verhoef, *Transp. Res. Part B Methodol.*, vol. 42, no. 7–8, pp. 607–618, Aug. 2008.
13. M. Börjesson, J. Eliasson, J. P. Franklin, *Transp. Res. Part B Methodol.*, **46**, 7, 855–873 (2012)
14. K. Hjorth, M. Börjesson, L. Engelson, M. Fosgerau, *Transp. Res. Part B Methodol.*, **81**, 1, 230–251 (2015)
15. M. Fosgerau, L. Engelson, *Transp. Res. Part B Methodol.*, **45**, 1, 1–8 (2011)
16. M. Fosgerau, A. de Palma, *J. Urban Econ.*, **71**, 3, 269–277 (2012)
17. K. A. Small, *Econ. Transp.*, **1**, 1–2, 2–14 (2012)
18. M. Fosgerau, K. Small, University of California-Irvine, Department of Economics, Working Paper 131403 (2014)
19. Department for Transport, UK., WebTAG: TAG data book, December 2015 - Publications - GOV.UK, (2015) [Accessed: 09-Feb-2016]
20. J. V. Henderson, *J. Urban Econ.*, **9**, 3, 349–364, (1981)
21. J. Holguín-Veras, *Transp. Res. Part Policy Pract.*, **42**, 2, 392–413 (2008)
22. J. Holguín-Veras, *Transp. Res. Part Policy Pract.*, **45**, 8, 802–824 (2011)
23. R. Arnott, A. de Palma, R. Lindsey, *J. Urban Econ.*, **27**, 1, 111–130 (1990)
24. A. de Palma, M. Kilani, R. Lindsey, *J. Urban Econ.*, **64**, 2, 340–361 (2008)
25. J. D. Hall, Pareto improvements from Lexus lanes: the effects of pricing a portion of the lanes on congested highways, Working paper (2013) [Accessed: 18-May-2016]
26. A. de Palma, M. Fosgerau, *Eur. J. Oper. Res.*, **230**, 2, 313–320 (2013)