Existence of self-financing and Pareto-improving congestion pricing: Impact of value of time distribution

Yu (Marco) Nie *, Yang Liu

Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, United States

**ABSTRACT**

This paper considers a static congestion pricing model in which travelers select a mode from either, driving on highway or taking public transit, to minimize a combination of travel time, operating cost and toll. The focus is to examine how travelers’ value of time (VOT), which is continuously distributed in a population, affects the existence of a pricing-refunding scheme that is both self-financing (i.e. requiring no external subsidy) and Pareto-improving (i.e. reducing system travel time while making nobody worse off). A condition that insures the existence of a self-financing and Pareto-improving (SFPI) toll scheme is derived. Our derivation reveals that the toll authority can select a proper SFPI scheme to distribute the benefits from congestion pricing through a credit-based pricing scheme. Under mild assumptions, we prove that an SFPI toll always exists for concave VOT functions, of which the linear function corresponding to the uniform distribution is a special case. Existence conditions are also established for a class of rational functions. These results can be used to analyze more realistic VOT distributions such as log-normal distribution. A useful implication of our analysis is that the existence of an SFPI scheme is not guaranteed for general functional forms. Thus, external subsidies may be required to ensure Pareto-improving, even if policy-makers are willing to return all toll revenues to road users.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The study of congestion pricing is traced back to the seminal work of Pigou (1920) and Knight (1924). From then on, the underlying economic model has been developed (e.g. Beckmann et al., 1956; Walters, 1961; Vickrey, 1963, 1969). The essential idea behind congestion pricing is to charge road users a fee corresponding to the cost of congestion they cause. This charge, as a price signal, dampens travel demand (hence alleviate traffic congestion) by “taxing out” motorists whose benefits from making their trips on the "tolled" road do not justify the payment of the toll. Road pricing in various forms has been practiced in a number of cities (e.g. Small and Gomez-Ibanez, 1998; Nakamura and Kockelman, 2002; May and Sumalee, 2003; Eliasson and Mattsson, 2006; Eliasson et al., 2009), and existing empirical evidence (e.g. Button, 1986; Small, 1993) indicates that pricing a road does reduce demand for its use and consequently alleviate congestion.

Despite congestion pricing is theoretically appealing and becoming increasingly easy to implement, the public appears reluctant to embrace this policy. Among many arguments against road pricing, the equity issue is frequently cited. Marginal cost pricing, for example, is known to produce higher user cost and lead to inequities among users (see e.g. Yang and Zhang, 2002; Hau, 2005). It has been argued that congestion pricing is easier to draw support from the public if it is Pareto-improving, that is, alleviating traffic congestion while increasing nobody's travel cost (Lawphongpanich et al., 2004; Hau, 2005). Lawphongpanich and Yin (2008) proposed a network model that includes the above Pareto-improving requirement as
constraints, but ignores the heterogeneity in the value of time. On the other hand, many (e.g. Evans, 1992; Arnott et al., 1994; Hau, 1998) observed that if toll revenues are not refunded, road pricing is regressive in that its benefit rises with income. Namely, because those with higher incomes usually have higher value of time (VOT), they benefit more from congestion relief promised by road pricing. To resolve this equity problem calls for redistributing toll revenues among users. This paper is focused on the revenue redistribution schemes and the role the VOT distribution plays in these schemes.

Small (1992) showed that congestion pricing may be progressive if a lump sum refunding is implemented, i.e. an equal travel allowance for all commuters. Similar schemes have been examined by others, see e.g. Goodwin (1989), Poole (1992) and DeCorla-Souza (1995). Recently, Kockelman and Kalmanje (2005) proposed a credit-based congestion pricing strategy, in which all drivers receive a monetary travel allowance (as a portion of toll revenues) to use on the roads. The strategy subsidizes those who would like to cash out the credit by staying off the road, with toll revenues collected from those who are willing to pay out-of-pocket for using the road.

Redistribution schemes have also been studied using various mathematical models. Adler and Cetin (2001) proposed a redistribution model based on a two-route bottleneck model (Arnott et al., 1990), in which toll revenues collected from users on a more desirable route are used to compensate users on a less desirable route. Without considering heterogeneity of users, they showed that the approach will eliminate queuing time and reduce the travel cost of all users. For a single Origin–Destination (O–D) pair connected by a number of parallel routes, Eliasson (2001) showed that a tolling and refunding system that reduces system travel time and refunds the toll revenues equally to all users will make everyone better off, regardless of traffic flow model or VOT distribution. Similar results were obtained in Yang and Guo (2005), which concluded that a congestion pricing scheme that reduces system travel time and redistributes the toll revenues to all users is always Pareto-improving. Heterogeneous VOT is considered using a multiple-class network equilibrium model in Yang and Guo (2005). Arnott et al. (1994) studied the welfare effects of congestion tolls using Vickrey’s model (Vickrey, 1969), in which commuters differ from each other not only in VOT, but also in their value of late/early arrival costs as well as desirable arrival time. Arnott et al. (1994) found that Pareto-improving is not guaranteed with an equal lump sum return. In particular, drivers with low VOTs could remain worse off if they are minority. Recently, Liu et al. (2008) examined the existence of Pareto-improving and revenue-neutral pricing schemes in a bi-modal network. The model charges a toll on a more desirable mode (automobile) and subsidizes the less desirable one (transit) using the revenues. With a general VOT distribution of commuters, an anonymous toll scheme is sought to: (1) reduce system travel time, (2) ensure no individual user to be worse off, and (3) guarantee revenue-neutral (all toll revenues are redistributed). Based on a static user equilibrium model, a general condition is proposed to ensure the existence of a Pareto-improving scheme when all toll revenues are returned to users. Liu et al. (2008) also showed that the condition is guaranteed for uniformly distributed VOT functions.

It is well-known that actual VOT distributions are not uniform. For example, Ben-Akiva et al. (1993) suggested that VOT in a population follow a log-normal distribution. Is the general condition given in Liu et al. (2008) satisfied for a more general VOT distribution? In other words, if all toll revenues are to be returned to road users, is it always possible to find a toll scheme to alleviate congestion (i.e. reduce system travel time) while making nobody worse off? The present paper aims to address precisely this question.

A main finding of the paper is that whether Liu et al.’s condition is satisfied highly depends on the shape of the VOT distribution. We prove that any concave VOT distribution function satisfies the condition, and develop existence conditions for a class of first-order rational distribution functions. These distributions are useful because they could provide better approximations to realistic VOT distributions such as log-normal (see Section 4.3 for details). Nonetheless, general VOT distributions do not guarantee Liu et al.’s condition, which implies that an equal lump sum return does not always guarantee Pareto-improving without external subsidy. This result is interesting and, apparently, disagrees with the previous results such as Eliasson (2001) and Yang and Guo (2005). We note that the reason for the discrepancy is that our model imposes a mode-specific operating cost (see Section 2 for details).

We also provide an alternative derivation of Liu et al.’s existence condition. The derivation relaxes the requirement for revenue neutral and thereby reveals that a Pareto-improving toll scheme can generate positive net revenues. In turn, this implies that a toll authority could withhold all or a portion of these net revenues without making anybody worse off. To be more specific (see Section 3 for details), a Pareto-improving toll scheme can be either revenue-neutral (i.e. returning all net revenues to users, which is considered in Liu et al. (2008)) or revenue-maximization (i.e. the toll authority withholds all net revenues). To some extent this is a good news because it gives policy-makers flexibility to utilize toll revenues for other investments (such as the road construction and maintenance), without undermining the feasibility of the toll scheme. It follows that a self-financing and Pareto-improving (SFPI) toll scheme is unique only when the net revenue is zero.

This paper is organized as follows. The next section reviews the bi-modal pricing model and its user equilibrium solution. In Section 3, a condition that guarantees the existence of an SFPI toll scheme is derived and its relationship with the one proposed in Liu et al. (2008) is discussed. In Section 4, the existence of an SFPI toll is considered for several special VOT functions, including concave, linear and a class of rational functions. We present numerical results in Sections 5 and 6 concludes the study.

2. The bi-modal pricing model and its solutions

Consider a network with one O–D pair, which has a fixed demand d and is connected by two routes: a highway and a transit line. Denote the travel time on the highway by \(\tau(f)\), a strictly increasing and convex function of the number of people
who decide to drive \( f \). The number of transit users is thus \( d - f \). The travel time of transit service is assumed to be a constant \( \gamma \). Let \( c_A \) and \( c_T \) denote the operating costs associated with using highway (e.g. cost of owning and driving a car) and transit line (e.g. transit fare), respectively. Without loss of generality, assume that the highway provides shorter travel time in the absence of congestion. That is, \( \gamma > \tau(0) \), \( c_A > c_T \). Travelers are heterogeneous in the sense that their value of time (denoted by \( \beta \)) are different. Let \( F(\beta) \) describe the distribution of \( \beta \) among travelers. We define \( F(\beta_0) \) as the total number of users whose value of time \( \beta \geq \beta_0 \). Note that by this setting \( F(\beta_1) = d \) and \( F(\beta_0) = 0 \), where \( \beta_1 \) and \( \beta_0 \) are minimum and maximum VOT among all users, and \( \beta_1 \geq 0 \). For simplicity, we shall also assume that \( F(\cdot) \in \mathcal{F} \), where \( \mathcal{F} \) is a class of continuous and strictly decreasing function defined on \( [\beta_L, \beta_U] \). Thus, any individual can be identified according to \( \beta \) or a ranking in the population \( f = F(\beta) \). Conversely, for any given \( f \), \( F^{-1}(f) \) identifies a unique \( \beta \).

The total travel cost of an individual is a combination of travel time, operating cost and toll, and is measured in monetary unit. When the highway flow is \( f_p \), the travel cost of an individual \( f \) is defined by

\[
\begin{align*}
J_p(f) &= \left\{ \begin{array}{ll}
\beta(f)\tau(f_p) + c_A + \mu_h & \text{if } f \leq f_p \\
\beta(f)\gamma + c_T + \mu_t & \text{if } f > f_p
\end{array} \right.
\]

(1)

where \( \mu_h \) and \( \mu_t \) are the tolls charged on highway and transit line, respectively. At present we consider the no-toll equilibrium so both \( \mu_h \) and \( \mu_t \) are assumed to be zero. When all users choose a mode to minimize their own travel costs, the bi-modal network admits a user equilibrium flow pattern at which a traveler \( f_e \) is subject to an identical cost whether s/he chooses transit or highway. Namely,

\[
F^{-1}(f_e)\tau(f_e) + c_A = \gamma F^{-1}(f_e) + c_T
\]

(2)

Let \( \beta_e = F^{-1}(f_e) \), \( \tau_e = \tau(f_e) \), \( \Delta c = c_A - c_T \) (note that \( \Delta c > 0 \) by assumption). We rewrite the above equation as

\[
(\gamma - \tau_e)\beta_e = \Delta c
\]

(3)

The traveler \( f_e \) is hereafter referred to as the indifferent user at no-toll equilibrium. People with \( \beta > \beta_e \) will use the highway because \( \beta > \beta_e \) implies \( \tau_e \beta + c_A < \gamma \beta + c_T \). Similarly, people with \( \beta < \beta_e \) will select the transit line. For a given VOT distribution, when \( \Delta c \) or \( \gamma \) is large enough, corner solutions may emerge in which all travelers choose transit or highway. For simplicity we shall rule out these possibilities and focus on interior solutions in this paper. To this end, the following condition is imposed:

\[
(\gamma - \tau(d))F^{-1}(d) < \Delta c < (\gamma - \tau(0))F^{-1}(0)
\]

(4)

Note that \( F^{-1}(d) = \beta_L \), \( F^{-1}(0) = \beta_U \). Fig. 1 illustrates that corner solutions may occur when VOT is overly concentrated on high or low ends.

It is clear that one may influence the equilibrium flow \( f_e \) by changing \( \Delta c \), in the form of a toll on either mode or both. We define a tolled equilibrium by \( f_e \), in contrast with a no-toll solution \( f_e \). In fact, any \( f_p \in [0, d] \) can be attained at equilibrium by selecting a proper \( \Delta \mu = \mu_h - \mu_t \). As demonstrated in (4) and Fig. 1, one can always find a \( \Delta \mu \) (for given \( F(\cdot), \Delta c, \gamma \) and \( \tau(\cdot) \)) to achieve either corner solution, and hence any solution in between (due to the continuity of the VOT distributions and travel time functions).

**Proposition 1.** \( f_p \) decreases with \( \Delta \mu \), where \( f_p \) satisfies \( f_p = F\left(\frac{\Delta c + \Delta \mu}{\gamma - \tau(d)}\right) \).

**Proof.** If \( \Delta \mu \uparrow \), then \( \beta(f_p) = \frac{\Delta c + \Delta \mu}{\gamma - \tau(d)} \uparrow \). Suppose \( f_p \uparrow \), then \( F(\beta(f_p)) \) is an increasing function. A contradiction with the assumption that \( F \in \mathcal{F} \). \( \square \)

Intuitively, as \( \Delta \mu \) increases the expense of highway use, people tend to be tolled off the highway.

---

**Fig. 1.** An illustration of corner solutions.
3. Self-financing and Pareto-improving toll scheme

3.1. Definitions

We first note that the system travel time, denoted as $G$, and the system cost, denoted as $\hat{G}$, can be defined as

$$G \equiv \tau(f) + \gamma(d - f); \quad \hat{G} \equiv \int_0^d \beta(\omega)\tau(\omega)d\omega + \int_f^d \beta(\omega)\gamma d\omega + c_f + c_T(d - f)$$

(5)

Note that, from a planner’s viewpoint, the toll revenue is considered a way of transferring funds and therefore is not included in the system travel cost. The self-financing and Pareto-improving (SFPI) toll is characterized as follows.

**Definition 1** (SFPI). A toll scheme is SFPI if it produces a flow pattern $f_p$ such that (1) $\mu_p f_p + \mu_T (d - f_p) \geq 0$, (2) $G_p < \hat{G}$, and (3) $u_p(f) \leq u_c(f), \forall f \in [0, d]$, where $G_p$ and $\hat{G}$ are the system travel times corresponding to tolled and untolled equilibria.

A toll is **Pareto-improving** when it satisfies Conditions (2) and (3) in the above definition. It merits elaborations why the reduction of system travel time, instead of system cost, is required in our definition of Pareto-improving. Let $TR$ be the total toll revenue, and we have the following identity:

$$\int_0^d u(f)df = \hat{G} + TR.$$  

(6)

which states that the summation of the all individual costs should equal to system cost $\hat{G}$ plus toll revenue. Since by Condition (3) in **Definition 1** (no individual is worse off in terms of cost), $\int_0^d u(f)df$ in Eq. (6) will decrease (or remain constant) after the toll is implemented. Moreover, self-financing implies $TR \geq 0$. Consequently, any toll scheme that satisfies Conditions (1) and (3) in **Definition 1** will either lower system cost or keep it constant. Therefore, by requiring strict reduction of total travel time, we effectively define a more restrictive set of toll schemes which entails another dimension of “improvement”. Decision-makers may be interested in reducing total system travel time (hence alleviating congestion) even when the reduction of system cost is guaranteed, because, among other reasons, it is beneficial to lowering emissions and fuel consumption.

We proceed to show under what conditions the strict travel time reduction is possible, since it is required by Pareto-improving. Note that $G$ as defined in (5) is minimized when

$$\frac{dG}{df} = \tau(f) + f\tau(f)' - \gamma = 0,$$

or equivalently, $\gamma - \tau(f) = f\tau(f)'$.

Denote the flow pattern that satisfies the above equation by $f_s$, and the corresponding system travel time by $G_s$. Let $\beta(f_s) = \beta_s, \tau(f_s) = \tau_s$, the tolled equilibrium implies $(\gamma - \tau_s)\beta_s = \Delta c + \Delta \mu$. It follows that if the toll is such set that

$$\Delta \mu = \beta_s f_s \tau_s' - \Delta c,$$

(7)

toled equilibrium solution minimizes the system travel time. Thus, the system travel time can be reduced by a toll if (1) the highway is overused (i.e. $f_s > f_e$) and (2) $\Delta \mu$ defined in (7) is positive. We formally state this as

**Assumption 1.** In the bi-modal pricing model, it is assumed that $f_e > f_s$ and $\Delta c < \beta(f_s) f_s \tau(f_s)'$ where $f_s$ and $f_e$ are the system optimal and no-toll user equilibrium flows, respectively.

The above assumption ensures that the system travel time can be reduced from $G_e$ (the total travel time at user equilibrium) when $\Delta \mu \in (0, \Delta \mu_0)$, where $\Delta \mu_0$ corresponds to either the toll that forces all people leave the highway, or the toll that drives the system travel time back to $G_e$.

Finally, it is worth mentioning that the optimum toll that minimizes the total system cost $\hat{G}$ has a different form from (7). Nevertheless, the difference should not concern us since the SFPI toll examined in the following is not a marginal toll, and thus need to minimize neither system travel time nor system cost.

3.2. An alternative derivation of the general existence condition

We now provide an alternative derivation of the general existence condition given in Liu et al. (2008), without imposing the revenue-neutral requirement. We first examine how travel cost changes when $\Delta \mu$ is imposed to reduce highway flow from $f_e$ to $f_s$. There always exists an $f_p$ such that $G_p < G_e$ by **Assumption 1**. For those who use transit before the toll (i.e. $f > f_e$), their cost will be affected only by $\mu_p$. Pareto-improving thus implies $\mu_p \leq 0$. Accordingly, self-financing requires $\mu_A \geq 0$. For drivers who are tolled off the highway, (i.e. $f_p < f < f_e$), the cost change is

$$\delta u(f) = \beta(f)\gamma + c_T + \mu_T - \beta(f)\tau(f_e) - c_A = \beta(f)(\gamma - \tau(f_e)) - \Delta c + \mu_T$$

:\footnote{Whether this will occur depends on the shape of the function $G$ as well as VOT distribution.}
Recalling Eq. (3), the above relation is simplified as
\[
\delta u(f) = (\beta(f) - \beta_s)(\gamma - \tau(f_s)) + \mu_T
\]  
(8)
Since when \( f < f_e, \beta(f) > \beta_s, \delta u(f) \leq 0 \) implies \( \mu_T \leq (\beta_s - \beta(f))(\gamma - \tau(f_s)) \). Because the largest \( \beta(f) \) among those who switch from highway to transit is \( \beta_p \) (the VOT of the indifferent user at the tolled equilibrium), all these users are not worse off as long as the indifferent user receives enough subsidy to compensate the increased travel time, namely,
\[
\mu_T \leq (\beta_s - \beta_p)(\gamma - \tau(f_s)) < 0
\]  
(9)
For users who stay on the highway \( (f \leq f_p) \), the cost change is
\[
\delta u(f) = \beta(f)(\tau(f_p) - \tau(f_s)) + \mu_A
\]  
(10)
They are not worse off if the toll does not exceed the benefits they receive from travel time savings, namely when \( \delta u(f) \leq 0 \Rightarrow \beta(f) \geq \mu_A/(\tau(f_s) - \tau(f_p)) \). In this case, the gain of the indifferent user is smallest because its VOT is smallest among those who stay on highway. Therefore, nobody is worse off if
\[
\beta_p \geq \mu_A/(\tau(f_s) - \tau(f_p))
\]  
(11)
To summarize, we have the following result.

**Proposition 2.** A toll scheme makes everyone better off if \( 0 < \mu_A \leq \beta_p(\tau_e - \tau_p) \) or \( \mu_T \leq (\beta_s - \beta_p)(\gamma - \tau_e) \).

It is worth emphasizing that Conditions (9) and (11) are not independent since \( \mu_A \) and \( \mu_T \) can be determined from each other for a given \( f_p \). In particular, when \( \mu_A \) attains its upper bound in (9), \( \mu_T \) must also attain the upper bound. This is essentially because the indifferent user will have zero gain in either case. We proceed to examine when a Pareto-improving toll scheme is also self-financing. Note that self-financing implies that the net toll revenue \( R = \mu_A f_p + \mu_T f_t \geq 0 \). Consequently, as long as the maximum possible \( R \) is non-negative, an SFPI scheme always exists. This observation gives rise to the following important result.

**Proposition 3.** A Pareto-improving and self-financing toll scheme exists for a targeted flow pattern \( f_p \) when
\[
\Delta G = G_e - G_p \geq \Delta c \left( \frac{d - f_p}{\beta_s} - \frac{d - f_e}{\beta_p} \right) = \Delta c(\sigma_e - \sigma_p)
\]  
(12)

**Proof.** \( R \) achieves its maximum possible value \( R_m \) when \( \mu_A \) takes the maximum, which also corresponds to the minimum subsidy \(-\mu_T\) for transit users. Thus, \( R_m \) for a given \( f_p \) is given by
\[
R_m = \beta_p(\tau_e - \tau_p)f_p + (\beta_e - \beta_p)(\gamma - \tau_e)(d - f_p)
\]
Setting \( R_m \geq 0 \) and using (5), we have the following result.
\[
R_m = \beta_p(\tau_e - \tau_p)f_p + (\beta_e - \beta_p)(\gamma - \tau_e)(d - f_p)
\]
\[
= \beta_p(\gamma - \tau_e)(d - f_p) + \beta_e[\beta_e(\tau_e - \tau_p) - (\gamma - \tau_e)(d - f_p)]
\]
\[
= \Delta c(d - f_p) + \beta_p[-G_e + d\tau_e]
\]
\[
= \Delta c(d - f_p) + \beta_p[-G_e + G_e + (d - f_e)(\tau_e - \gamma)]
\]
\[
= \Delta c(d - f_p) + \beta_p\left(-G_e + G_e - \Delta c\frac{d - f_e}{\beta_e}\right)
\]
\[
= \beta_p\left[G_e - G_p + \Delta c\left(\frac{d - f_p}{\beta_p} - \frac{d - f_e}{\beta_e}\right)\right] \geq 0
\]
\[\iff G_p \leq G_e + \Delta c\left(\frac{d - f_p}{\beta_p} - \frac{d - f_e}{\beta_e}\right) \iff \Delta G \geq \Delta c(\sigma_e - \sigma_p)\]
This completes the proof. \( \square \)

According to Condition (12), we define maximum equivalent system travel time (MESTT) as
\[
G_M \equiv G_e + \Delta c(\sigma_p - \sigma_e) = G_p + \frac{R_m}{\beta_p}
\]  
(13)
Note that \( G_p \) is the system travel time at \( f_p \) and \( R_m/\beta_p \) can be regarded as an extra system travel time corresponding to maximum toll revenues (weighted by the value of time of the indifferent user). Thus, \( G_M \) measures the maximum total system cost (including both toll and travel time) to guarantee that nobody is worse off. Only when \( G_M \) does not lie below \( G_e \) is the toll scheme also self-financing. That is, Condition (12) equals \( G_M \geq G_p \).

Condition (12) is identical to the one given in Liu et al. (2008). The latter is derived by forcing the net revenue to be zero, and thereby maximizing the gain of the indifferent user. To see this, first recall that at the tolled equilibrium
\[ \Delta \mu = \beta_p (\gamma - \tau_p) - \Delta c \] (14)

Noting that \( \Lambda_s = \mu_s - \mu_t = \Delta \mu \), and that the zero-revenue implies \( f_p \mu_A + (d - f_p) \mu_t = 0 \), \( \mu_A \) can be solved as

\[ \mu_A = \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c] \]

Plugging this into Eq. (10), the cost change for those who stay on highway after toll is

\[ \delta u(f) = \beta(f)(\tau(f_p) - \tau(f_e)) + \frac{d - f_p}{d} (\beta_p (\gamma - \tau_p) - \Delta c) \]

Again, the indifferent user suffers more loss than anyone else. Thus, if the indifferent user’s cost increase is non-positive, an SFPI toll can be guaranteed. That is,

\[ \delta u(f_p) = \beta_p (\tau(f_p) - \tau(f_e)) + \frac{d - f_p}{d} (\beta_p (\gamma - \tau_p) - \Delta c) = \frac{\beta_p}{d} [d \tau_p - d \tau_e + (d - f_p) (\gamma - \tau_p)] - \frac{d - f_p}{d} \Delta c \]

\[ = \frac{\beta_p}{d} [d (\gamma - \tau_e) - f_p (\gamma - \tau_p) - \frac{d - f_p}{\beta_p} \Delta c] = \frac{\beta_p}{d} [\Delta c \frac{d - f_p}{\beta_p} + \Delta c \frac{f_p}{\beta_p} + G_p - d \gamma - \frac{d - f_p}{\beta_p} \Delta c] \]

\[ = \frac{\beta_p}{d} [\Delta c (\sigma_e - \sigma_p) + G_p - f_p (\gamma - \tau_e) - d \gamma] = \frac{\beta_p}{d} [\Delta c (\sigma_e - \sigma_p) + G_p - G_e] = \frac{R_m}{d} \leq 0 \] (15)

Not surprisingly, the above condition is identical to (12). To summarize, there are two extreme SFPI toll schemes: one is cost-neutral, in which the indifferent user in the presence of toll is subject to zero cost change \( \delta u(f_p) = 0 \); another is revenue-neutral, which means that the total revenue from toll is zero \( (R_m = 0) \). Specifically, for a given \( f_p \), if the interval \( \Lambda_s = \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c], \beta_p (\tau_p - \tau_p)] \) is not empty, any \( \mu_s \in \Lambda_s \) is SFPI provided that \( G_p < G_e \). The toll authority thus enjoys a certain flexibility to distribute the benefit between users and itself, by selecting \( \mu_s \) and \( \mu_t \). The feasible range of \( \mu_t \) can be established from that of \( \mu_s \) according to Eq. (14). Note that

\[ \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c] \leq \mu_t \leq \beta_p (\tau_e - \tau_p) \]

\[ \Rightarrow \quad \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c] - (\beta_p (\gamma - \tau_p) - \Delta c) \leq \mu_t \leq \beta_p (\tau_e - \tau_p) - (\beta_p (\gamma - \tau_p) - \Delta c) \]

\[ \Rightarrow \quad \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c] - \Delta c \leq \mu_t \leq \Delta c 1 - \frac{\beta_p}{\beta_p} < 0 \]

Therefore, \( \Lambda_t = \left[ \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c], \Delta c 1 - \frac{\beta_p}{\beta_p} \right] \). When \( R_m = 0 \), the SFPI toll scheme is unique and is both cost-neutral and revenue-neutral. Whenever \( R_m < 0 \), no SFPI scheme exists for the given \( f_p \).

As shown in the last equality of (15), if all \( R_m = \beta_p [\Delta G + \Delta c (\sigma_p - \sigma_e)] \) is redistributed to users as an equal lump sum, the gain of the indifferent user is \( R_m/d \). Also, whenever \( R_m > 0 \), a credit-based congestion pricing such as suggested in Kockelman and Kalmanje (2005) can be implemented by providing a credit equal to the lower bound of \( \mu_t \) and setting \( \mu_A \) to be the sum of the credit and its lower bound \( \left( \frac{d - f_p}{d} [\beta_p (\gamma - \tau_p) - \Delta c] \right) \). In this case, the credit replaces the negative toll on the transit line, so \( \mu_t = 0 \).

4. VOT distribution functions

Assumption 1 guarantees \( G_p < G_e \), i.e. \( \Delta G > 0 \). Thus, an immediate observation from Condition (12) is that an SFPI scheme exists when \( \sigma_e < \sigma_p \). If \( \sigma_e > \sigma_p \), it is possible that (12) is not satisfied.² Fig. 2a gives a geometrical interpretation of \( \sigma_e \) and \( \sigma_p \), which shows that the function \( \sigma(\cdot) \) of \( f \) denotes the slope of the straight line connecting \( [\beta, f] \) to \( [\beta, d] \) on the VOT distribution function. Based on this observation, we now examine the existence of SFPI toll schemes for several special VOT functions.

4.1. Concave functions

Theorem 1. Given Assumption 1, an SFPI scheme always exists for the bi-modal system if the VOT distribution function \( F \in \mathcal{F} \) is concave on \( [\beta_L, \beta_U] \).

Proof. Note that it suffices to show that \( \sigma_e < \sigma_p \) whenever \( 0 < f_p < f_e \). This implies that \( \sigma(\cdot) \) is a decreasing function within the range. We first introduce an auxiliary function \( h(\cdot) \) which denotes the slope of the line connecting \( [\beta, f] \) to \( [\beta, d] \) (see Fig. 2b). It is evident from the figure that \( \tau_p > \theta_e \). To see this, let us define

² In previous studies such as Eliasson (2001) and Yang and Guo (2005), \( \Delta c \) is always assumed to be zero. Because \( \Delta G \geq \Delta c (\sigma_e - \sigma_p) \) as long as \( \Delta c = 0 \) and \( \Delta G > 0 \), regardless of the relationship between \( \sigma_e \) and \( \sigma_p \), it is concluded that \( \Delta G > 0 \) always implies the existence of an SFPI toll. This need not be the case when \( \Delta c > 0 \).
4.2. First-order rational functions

Given Assumption 1, an SFPI scheme always exists for the bi-modal system if VOT is uniformly distributed on $[\beta_L, \beta_U]$. A uniform distribution function can be written as $F(\beta) = \frac{d-f}{\beta_0-\beta_L} (\beta - \beta_L) + d$. Let $\alpha = \frac{\beta_0-\beta_L}{\beta_0-\beta} \quad \alpha < 0$ and the inverse of the $F(\beta)$ is $F^{-1}(f) = \alpha(f-d) + \beta_L$. So we can write $\sigma(f) \mid \beta_L$ as a function of $f$, i.e.

$$\sigma(f) = \frac{d-f}{\beta(f)} = \frac{d-f}{\alpha(f-d) + \beta_L}$$

This is a non-increasing and concave function because

$$\sigma(f)' = \frac{-\beta_L}{(\alpha(f-d) + \beta_L)^2} \leq 0; \sigma(f)'' = 2\alpha\beta_L(\alpha(f-d) + \beta_L)^{-3} \leq 0$$

Fig. 3 illustrates the relationship between the system travel time function ($G_\beta$) as well as the corresponding MESTT function ($G_M$, cf. Condition (13)). Clearly, for any $f$ such that $G_M$ lies above $G_\beta$ and $G_\beta < G_\sigma$, an SPFI toll scheme exists. Trajectories of $G_M$ corresponding to two different uniform distributions are plotted in the figure, namely, homogeneous defined as $\beta_L = 0$ and inhomogeneous defined as $\beta_L > 0$. When $\beta_L = 0, \sigma(f)' = 0$ (cf. (17)). Since $G_M$ does not change with $f$ in this case, its trajectory is the horizontal line in Fig. 3. Moreover, one may design a self-financing toll scheme that makes everyone better off but increases the total system travel time. For the inhomogeneous case, this occurs when $f$ falls into the range $[\beta_0, \beta_1]$.

4.2. First-order rational functions

We now examine another special case in which the VOT distribution takes the form of a first-order rational function as follows:
\[ F(\beta) = \frac{d(\beta_U - \beta)}{\rho \beta + \beta_U} \]  

where \( \rho > -1 \) is a parameter. It is easy to verify that \( F(0) = d \) and \( F(\beta_U) = 0 \). Also, a function of form (18) is convex when \( \rho > 0 \), concave when \(-1 < \rho < 0\), and reduced to a linear function if \( \rho = 0 \) (see Fig. 4a).

**Proposition 4.** If Assumption 1 holds and \( F(\cdot) \) is a first-order rational function of form (18), then

1. an SFPI toll scheme exists only when \( \frac{\partial G_M}{\partial f} \leq \tau(f_p) + f_p \tau(f_p)' - \gamma \); and
2. whenever an SFPI scheme exists, a target flow \( f_p \) maximizes maximum toll revenues if \( \frac{\partial G_M}{\partial f} = \tau(f_p) + f_p \tau(f_p)' - \gamma \).

**Proof.** Note that the function \( \sigma(f) \) is defined as

\[ \sigma(f) = \frac{d - f}{\beta} = \frac{(d - f)(\rho f + d)}{\beta_U (d - f)} = \frac{1}{\beta_U} (\rho f + d) \]

Thus, \( G_M = G_e + \Delta \sigma(f_p - f_e) \). Recalling Proposition 3, the existence of an SFPI scheme requires that \( G_M - G_e > 0 \). Taking the first-order derivative of \( G_M - G_e \) with respect to \( f_p \) and recalling that \( G_e \) is constant,

![Fig. 3. An illustration of Gp and G0 for a uniformly distributed VOT function.](image)

![Fig. 4. An analysis for the rational function.](image)
\[
\frac{d(G_M - G_p)}{df_p} = \frac{\rho \Delta c}{\beta_0} - [\tau(f_p) + f_p \tau(f_p)' - \gamma]
\]

\(G_M - G_p\) attains the maximum when its gradient equals zero, which proves the second part of the proposition. To prove the first part, note that the convexity of \(G_p\) and Assumption 1 imply that \(\tau(f_p) + f_p \tau(f_p)' - \gamma > 0\) when \(f_p = f_e\) and that for any \(f_p < f_e:\)

\[G_e + (\tau(f_e) + f_e \tau(f_e)' - \gamma)(f_p - f_e) < G_p\]

Thus, if \(\frac{\rho \Delta c}{\beta_0} > \tau(f_e) + f_e \tau(f_e)' - \gamma > 0\), then \(G_M = G_e + \frac{\rho \Delta c}{\beta_0} (f_p - f_e) < G_p\) (cf. Fig. 4-(b)).

Note that \(\tau(f_e) + f_e \tau(f_e)' - \gamma\) represents the difference between the marginal travel times of highway and transit. The larger the difference is, the more system travel time reduction is achieved by tolling the highway, which in turn makes it easier to realize an SFPI toll scheme.

**Corollary 2.** For the bi-modal system, if \(F(\cdot)\) is a first-order rational function of form (18), then (1) an SFPI toll scheme always exists if \(-1 < \rho \leq 0\); and (2) if \(\rho = 0\), the system optimal flow pattern \(f_e\) maximizes the maximum toll revenue.

**Proof.** The first part directly follows Theorem 1 by noting that the rational function (4) is concave when \(\rho < 0\) and linear when \(\rho = 0\). The second part follows from Proposition 4 by noting that \(\tau(f_e) + f_e \tau(f_e)' - \gamma = 0\).

Fig. 4b plots \(G_M\) and \(G_p\) curves for different \(\rho\) values. All \(G_M\) curves are straight lines with a slope of \(\frac{\rho \Delta c}{\beta_0}\). As shown in the figure, when \(\rho \leq 0\), any \(f_p\) such that \(G_p < G_e\) (i.e. \(f_p \in (f_e, f_1)\)) corresponds to an SFPI toll. However, if \(\rho > 0\), the feasible range of \(f_p\) is reduced. When \(\rho\) is too large to satisfy the condition given in Proposition 4, the feasible range will become empty. Therefore, an SFPI toll may not exist for any target flow pattern even if it reduces the system travel time.

### 4.3. Log-normal functions

We discuss log-normal functions because real VOT distribution data can be reasonably fitted by them (Ben-Akiva et al., 1993). Unlike in previous cases, general conditions for the existence of an SFPI toll are hard to establish. Instead, the existence of such a toll depends on the shape of the log-normal function as well as no-toll equilibrium flow \(f_e\). Generally, a log-normal distribution consists of three portions: a concave region for low VOT values, a convex region for high VOT values, and an approximate linear region in between (see Fig. 5). According to Theorem 1 and its corollary, when \(\beta_e\) falls into either the concave or linear region, an SFPI is guaranteed so long as the target flow \(f_p\) does not go beyond the two regions. Three log-normal distribution functions with a same mean of 30 and different variances are plotted in Fig. 5. As shown, when variance is small in the population, the log-normal distribution can be approximated by a linear function. However, for large variance, the log-normal distribution is convex almost everywhere. This circumstance may arise, for example, when a small number of people are extremely wealthy (very high VOT) while the majority is poor (low VOT). In this case, it is possible that an SFPI toll does not exist.

![Fig. 5. An illustration of log-normal distribution function (mean = 30).](image-url)
5. Numerical results

5.1. First-order rational function

The analysis presented in the previous sections is tested with a set of data for a bi-modal pricing system. For simplicity, a set of first-order rational functions of form (4) are employed to calculate no-toll equilibrium $f_0$, the feasible range of target flow $f_p$ to guarantee SFPI, as well as revenue-neutral and revenue-maximization toll schemes.

Let the highway performance function be of the following BPR form:

$$\tau(f) = 0.75 \left(1 + 0.15\left(\frac{f}{500}\right)^5\right) \text{(hour)}$$

where 0.75 (h) is the free flow travel time and the capacity of the highway is 500 (vehicle per hour). The transit travel time is set to $\gamma = 1.5$ (h). The difference in the operating cost $\Delta c = 54$. The total demand $d = 1000$, which if all loaded onto the highway, will almost increase the free flow travel time by five times. The VOT function takes the form of (18), in which the upper bound $b_0 = 40$ ($/\text{hour}$), and $\rho$ varies between $(-1,3,1]$. Rational functions with different $\rho$ values are plotted in Fig. 6a. Fig. 6b compares $G_p$ (i.e. the system travel time curve as a function of target flow $f_p$, represented by the solid curve) with $G_M$ (i.e. the maximum equivalent system travel time, represented by dash lines) corresponding to five $\rho$ values. We note that the highway flow at no-toll equilibrium ($f_0$) are different for each different rational function (see Column 2 of Table 1).

Recall from Proposition 3 that an SFPI toll scheme exists when $G_M > G_p$. Also, the discrepancy between $G_M$ and $G_p$ gives the maximum possible revenue (in the unit of travel time) for a corresponding target flow $f_p$. Clearly, for each rational function, there exists a feasible range of $f_p$ within which $G_M > G_p$. Column 3 of Table 1 reports these ranges. Also there is a decreasing range of $f_p$ within which $G_p < G_M$, as shown in Column 4 of Table 1. The intersection of the two ranges defines an SFPI range which insures the existence of an SFPI toll. We note that the two ranges are identical when VOTs are uniformly distributed ($\rho = 0$). When $\rho = 3.1$, $G_M$ is tangent to $G_p$ at $f_p \left(\frac{w_n}{w_n} = \tau(f_p) + f_p\tau(f_p)' - \gamma\right)$ in Fig. 6b. Consequently, for all the first-order rational functions with $\rho \geq 3.1$, the SFPI range is empty.

Within the SFPI range, the flow pattern $f_p$ that maximizes the toll benefit is calculated according to Proposition 4, reported in Column 2 of Table 2. Table 2 also presents two extreme revenue distributions: revenue-maximization scheme which leaves the indifferent user to gain nothing; and revenue-neutral scheme which refunds all revenues to users so that the indifferent user’s gain is maximized. When $\rho = 0.8$, for example, the maximum toll revenue $788.3$/h is achieved at $f_p = 521.1$ (veh/h). The toll authority can charge a toll on freeway as high as $2.83$ to realize this revenue, or as low as $2.04$ (lower bound), the subsidy for transit riders is $2.23$. In this case, to implement a credit-based congestion pricing, all travelers should receive an equal lump sum of $2.23$ and the toll on highway will be set to $2.23 + 2.04 = $4.27.

5.2. Log-normal VOT distributions

We have mentioned in Section 4 that log-normal VOT distributions are convex almost everywhere when variance is large. Since rational functions are also convex when $\rho > 0$, a log-normal VOT distribution with large variance may be approximated using a proper rational function. Numerical results in this section test this conjecture.

In Fig. 7, the dashed lines represent log-normal VOT functions with the same mean of 21 ($/\text{hour}$) and different standard deviations ranging from 10.5 to 50. The parameters of the first distribution (with mean = 21 and standard devia-
tion = 10.5) are adopted from a study of commuters on State Route 91 located in Orange County, CA (Lam and Small, 2001). From the figure we can see that the two curves with larger variances (35 and 50) could be reasonably approximated by rational functions with \( q = 2.5 \) and 4, respectively, \( (\beta_U = 50) \) represented by the solid curves in the figure. In particular, these rational functions well fit the corresponding log-normal curves when \( \beta \in [0, 20] \). If the real VOT data can be fitted using a rational function, the analysis of SFPI toll schemes is simplified by using Proposition 4. However, when the variance of log-normal VOT distribution is small, rational functions no longer provide good fit. As mentioned before, log-normal functions may be approximated piecewisely in this case, using other functions for which analytical results are available (e.g. concave or linear).

6. Concluding remarks

We studied a static congestion pricing model in which users choose between highway and transit to minimize their own travel cost combining travel time, operating cost and toll. The purpose is to examine the effects of individuals’ value of time

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( f_p )</th>
<th>Feasible range</th>
<th>Decreasing range</th>
<th>SFPI range</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>541.0</td>
<td>[541.0, 541.0]</td>
<td>(477.7, 541.0)</td>
<td>0</td>
</tr>
<tr>
<td>1.5</td>
<td>598.7</td>
<td>[445.9, 598.7]</td>
<td>(396.2, 598.7)</td>
<td>(445.9, 598.7)</td>
</tr>
<tr>
<td>0.8</td>
<td>626.7</td>
<td>[376.3, 626.7]</td>
<td>(343.9, 626.7)</td>
<td>(376.3, 626.7)</td>
</tr>
<tr>
<td>0</td>
<td>661.2</td>
<td>[261.7, 661.2]</td>
<td>(261.7, 661.2)</td>
<td>(261.7, 661.2)</td>
</tr>
<tr>
<td>−0.5</td>
<td>684.6</td>
<td>[155.1, 684.6]</td>
<td>(190.6, 684.6)</td>
<td>(190.6, 684.6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( f_p )</th>
<th>Revenue-maximize</th>
<th>Revenue-neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5</td>
<td>529.6</td>
<td>( \mu_\beta = 1.33, \mu_t = -0.96 )</td>
<td>253.6</td>
</tr>
<tr>
<td>0.8</td>
<td>521.1</td>
<td>( \mu_\beta = 2.83, \mu_t = -1.44 )</td>
<td>788.3</td>
</tr>
<tr>
<td>0</td>
<td>510.6</td>
<td>( \mu_\beta = 6.46, \mu_t = -1.78 )</td>
<td>2427.5</td>
</tr>
<tr>
<td>−0.5</td>
<td>503.7</td>
<td>( \mu_\beta = 11.3, \mu_t = -1.53 )</td>
<td>4916.0</td>
</tr>
</tbody>
</table>

Fig. 7. Log-normal distributed VOT functions with different standard deviations (SD).
(VOT) on the policy of congestion pricing. In particular, we are interested in the existence of a combined pricing-refunding scheme that is both self-financing and Pareto-improving (SFPI). A condition that insures the existence of such a toll was derived in Liu et al. (2008) by assuming that all toll revenues will be returned to users as an equal lump sum. This paper derives the same condition using an alternative approach. Our analysis suggests that a revenue-neutral toll scheme is only one of many feasible SFPI schemes. Consequently, the toll authority can select a proper SFPI scheme to distribute the benefits from congestion pricing. Also, these SFPI toll schemes can be easily implemented using a credit-based pricing strategy in which the travel allowance simply equals the calculated subsidy for transit riders.

We proved that an SFPI toll always exists if the VOT function is concave. The linear function, which was discussed in Liu et al. (2008), is a special case of this more general result. However, the existence of SFPI schemes is not always guaranteed in the bi-modal pricing model. This finding disagrees with those reported in the literature essentially because existing studies often ignores the differences in the operating costs of different modes (or routes). We derived analytical existence conditions for a class of first-order rational VOT functions, which suggest that an SFPI may not exist in a population whose VOT distribution is highly skewed to the low end. Although deriving closed-form existence (or non-existence) conditions for log-normal functions is difficult, it is possible to approximate them piecewisely with concave, linear or convex functions. When variance is large, for instance, log-normal functions can be reasonably approximated by rational functions and thereby amenable to aforementioned analysis. Although a rigorous proof cannot be offered, it is unlikely that an SFPI toll scheme would always exist for a realistic log-normal VOT function. Consequently, simply returning all revenues to users may not solve the equity problem rooted in congestion pricing.

This paper is focused on the interaction between highway and public transit. However, the transit line in the bi-modal pricing system could also be another highway that is longer but with enough capacity to keep congestion-free. Thus, the results of this paper are directly applicable in such a route choice circumstance. The model can also handle more general situations where all routes are affected by congestion effects.

Although the simple model considered in this paper reveals interesting properties of the pricing-refunding system, it has a few noteworthy limitations that should be relaxed in future studies. First, since only one O–D pair is considered, the spatial inequity effect due to the interaction among O–D pairs is ignored. For a network with many O–D pairs, both social and spatial equity issues exist, which have different causes and need different solutions to offset the disadvantage. Second, although the demand is intermodal elastic in our model, the number of total trips is exogenous. If total demand is also elastic, the level of equity issues exist, which have different causes and need different solutions to offset the disadvantage. Finally, our model implicitly assumes constant returns to scale on transit, and thereby amenable to aforementioned analysis. Although a rigorous proof cannot be offered, it is unlikely that an SFPI toll scheme would always exist for a realistic log-normal VOT function. Consequently, simply returning all revenues to users may not solve the equity problem rooted in congestion pricing.

Acknowledgements

We are grateful to two anonymous referees for comments on an earlier draft of the paper. We accept full responsibility for remaining shortcomings and errors. The research is partially funded by Louis Berger Junior Chair at Northwestern University.

References


