Welfare implications of highway capacity expansion with responsive transit operator when the two modes are imperfect substitutes

Fangni Zhang, Hai Yang and Wei Liu

Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.

Abstract

This paper is concerned with the welfare issues arise when the highway capacity is being expanded. In the presence of the public transit as an alternative travel mode, commuters’ modal choices would be affected by: the change of the highway characteristics, the transit operating strategies and the degree of substitutability between the two travel modes. The purpose of this paper is to understand how the transit operator reacts to highway expansion under various operating regimes, and recognize the combined impacts of the interaction between transit responsive strategies and highway expansion when the two modes are imperfect substitutes. Three types of transit operating regimes are considered: the first-best scenario, profit-maximizing monopoly and the second-best scenario with budget constraint. Interestingly, auto commuters are found to be more likely to benefit from the highway expansion compared to the transit commuters in all the scenarios. Boundary conditions are well established for the Pareto-improving outcome (all commuters are better-off), the occurrence of the Downs-Thomson Paradox (all commuters are worse-off), and other welfare imbalance situations. Furthermore, the results substantiate the crucial roles of the transit operation cost and the degree of substitutability in determining transit responsive strategies. The theoretical analysis provides useful insights for public policy on highway improvement and transit service design for the enhancement of social welfare.

Key words: Highway expansion; transit responsive strategy; market regime; imperfect substitution; Downs-Thomson Paradox.

1. Introduction

There is a large body of literature concerning the interaction of the two travel modes, i.e., automobile and public transit, in the transportation system. Regarding the impact of highway capacity expansion, Downs (1962) and Thomson (1977) first claimed that the improvement does not necessarily increase the commuters’ travel utility when the public transit serves as a substitute. The direct consequence of the improvement is to increase the traffic volume on highway by the effect of modal split. Meanwhile, the responsive transit operator could manipulate the service frequency and fare as reactions to patronage loss. Mohring (1972)
noted that normally the transit operator would provide less frequent service when the demand is lower, which leads to even fewer passengers and activate a vicious cycle (Reinhold, 2008; Bar-Yosef et al., 2013). In parallel, to maintain the operation in the short run, transit operator could raise the ticket price, which may induce a similar vicious cycle. Many researchers have investigated the occurrence of the paradox (e.g., Mogridge, 1987; Holden, 1989; Jara-Díaz, 1986; Arnott and Small, 1994; Abraham and Hunt, 2001; Denant-Boémont and Hammiche, 2009; Bell and Wichiensin, 2012; Hsu and Zhang, 2014), but the precise and quantitative conditions for the occurrence of the paradox considering both transit scheduling and pricing strategies have not been established.

As a comparison, expansionary strategies such as increasing frequency or reducing fare, which aim to improve the service quality and attract more passengers, could be adopted and be beneficial to both the transit operator and the commuters. Therefore, the eventual impact of highway capacity expansion on the commuters of the both travel modes highly depends on the transit operator’s responsive strategies, which are substantially subject to the market regime of the transit service. Following Evans (1987) and Bell and Wichiensin (2012), three major types of transit regimes are considered in this paper: unconstraint maximum social welfare (first-best scenario), profit-maximizing monopoly and breakeven maximum social welfare (second-best scenario).

The optimal design of transit service has been well studied by the existing literature (e.g. Mohring, 1972; Jansson, 1979, 1980; Evans, 1987; Jansson, 1993; Jara-Díaz and Gschwender, 2003), especially the transit frequency strategies under various market regimes (e.g. van Reeven, 2008; Savage and Small, 2010; Basso and Jara-Díaz, 2010; Karamychev and van Reeven, 2010; Nilsson et al., 2013; Gómez-Lobo, 2013). However, most of the models are built to deal with the design issues within the transit system itself. In the context of the two-mode system, commuters are choosing either the transit service or the highway system; therefore, the demand for transit should no longer be modeled as simple elastic ones; instead, any change of the characteristics in the two-mode system makes influences on the equilibrium volume allocation. Therefore, to figure out how would transit operator reacts to highway expansion, the modal choice behavior is modeled as a binary Logit model, and accordingly commuters make the modal choice based on the utility difference between the two modes (following Kitamura et al., 1998; Abraham and Hunt, 2001; Basso and Jara-Díaz, 2012).

Another concern of this paper is the degree of substitutability between the two modes. Recently, Zhang et al. (2014) studied the welfare impacts of highway expansion and responsive transit based on the assumption that the two modes are perfect substitutes. However, except for the similar function that both modes carry the commuters from the origin to destination, they differ in all the other characteristics, such as convenience, comfort
and reliability; therefore, there is imperfect competition between the two modes (Arnott and Yan, 2000; de Palma et al., 2013). Conceptually, there are two levels of imperfect substitutability, i.e., within the two-mode system (fixed total demand), or with other alternatives available (the total demand for the two modes is elastic). This paper would focus on the lower level of imperfect substitutability and employ partial equilibrium analysis. There are some papers in the literature that have considered the two-mode problem under the assumption of imperfect substitutability (e.g. Sherman, 1971; Rouwendal and Verhoef, 2004; de Palma and Proost, 2006; Ahn, 2009; Basso and Jara-Díaz, 2012; van den Berg and Verhoef, 2013; Clark et al., 2014), but rarely analytically investigated its impact of the interaction between highway expansion and responsive transit service. The Logit model considered in this paper is tractable and could effectively capture the impacts of imperfect substitutability with the elasticity coefficient representing the degree of substitutability.

The remainder of this paper is organized as follows. Next section introduces the basic model framework and defines the occurrence of the Downs-Thomson Paradox. Section 3 analyzes the impact of highway expansion under different transit operating strategies and market regimes, which includes the optimal design of transit fare and frequency and the scenarios with the transit operator as an unconstraint social welfare-maximizer, a monopolistic profit maximizer and a breakeven social welfare-maximizer. Conclusions and discussions are given in the last section.

2. Model formulation

2.1 Problem settings

We consider a single origin–destination transportation network to represent a corridor with a congested highway running in parallel to an exclusive transit line, linking the residential area and the central business district (CBD), as shown in Figure 1.

![Figure 1. A simple two-mode network](image_url)
proposed by Arnott and Small (1994), and the modal choice is formulated as the Logit model (following Kitamura et al., 1998; de Palma and Proost, 2006; de Palma et al., 2013).

For a typical day, the total number of commuters traveling from home to CBD is fixed at $D$. Commuters choose their travel mode fully based on the indirect utility conditional on the use of a particular mode:

$$u_a(\cdot) = U - \alpha v_a + \gamma c - \tau_a$$  \hspace{1cm} (1)

$$u_t(\cdot) = U - mv_t + nf - \tau_t$$  \hspace{1cm} (2)

where $u_a(\cdot)$, $u_t(\cdot)$ denote the indirect utility functions by auto and transit (subscript ‘a’ and ‘t’), respectively. The first term $U$ represents the utility one commuter obtains from the trip, which is regarded as a constant and the same for all commuters. Disutilities associated with congestion or crowdedness are given by the second terms, which are assumed to be linear functions of the number of commuters using each mode, $v_a$ and $v_t$; $\alpha$ and $m$ are the coefficients represent the constant marginal disutility incurred by every one more commuter using that mode. The third terms characterize the utility related to the capacity provision, i.e., highway capacity $c$ and transit frequency $f$, and $\gamma$ and $n$ are the corresponding coefficients. (All the above coefficients, $\alpha$, $m$, $\gamma$, $n$ are strictly positive.) The final terms, $\tau_a$ and $\tau_t$, are the monetary costs associated with auto and transit, respectively; and $\tau_a$ is assumed to be a constant. Particularly, the value of utility is measured in the same unit of the monetary cost.

The number of commuters choosing mode $i$ ($i = a, t$) is given by the following Logit model:

$$v_i = \frac{e^{\lambda_i}}{\sum_i e^{\lambda_i}} D$$  \hspace{1cm} (3)

where $\lambda$ is a strictly positive scale parameter that defined by the degree of substitutability of the two alternative travel modes, and a larger value of $\lambda$ indicates higher degree of substitutability and vice versa (de Palma et al., 2013). Evidently, the equilibrium volume allocation is determined by the utility difference between the two alternative modes. For any given $f$, $\tau$, and $c$, there exists a unique user equilibrium as shown in Appendix A.

Transit service is provided by a single transit operator, with frequency $f$ and fare $\tau$, as its decision variables; however, under different market regimes, the operating strategies are formulated for different objectives. The transit service could be running by the government for the sake of maximizing social welfare, or operated by a private sector whose objective is to maximize its profit. This paper would focus on the different responses of the transit operator to the change of the highway capacity, and the resulted impact on the two-mode system under different market regimes: first-best, profit-maximizing and revenue-neutral. Throughout the analysis of this paper, it is assumed that the combination of transit fleet size
and service frequency is large enough to carry all the commuters waiting on the platform. The transit operation cost, $k(f)$, is an increasing, convex and second-order differentiable function of frequency with a positive fixed cost $k_0$.

The study originates from the increase of the highway capacity provision. The assumption of continuity is made on the highway capacity, which is standard in the theoretical literature (e.g., Arnott and Yan, 2000; Light, 2009; Zhang et al., 2014). To focus on the impact of transit scheduling and pricing strategies, neither highway investment nor operation cost would be taken into account. Besides, the following analysis only applies for interior equilibrium states.

2.2 The Downs-Thomson Paradox

The well-known “Downs-Thomson Paradox” refers to the phenomenon that every individual commuter obtains less utility from the trip by either of the travel mode after the highway is expanded in the presence of the transit service. As stated by Downs (1962), the improved capacity firstly reduces automobile travel time on highway, and some of the informative transit passengers would immediately give up their original choices and drive private cars. The transit operator would take measures to deal with the shrinking patronage, whose actions would make influences on the equilibrium in return. Therefore, the changes of the utility levels by both modes are determined by the changes of the $(f, \tau, c)$ combination.

The Downs-Thomson Paradox occurs when the individual utility levels of the both modes decrease with the highway capacity, i.e., the marginal utility changes incurred by the marginal increase of capacity is negative for a specific volume allocation (Mogridge, 1987; Abraham and Hunt, 2001; Basso and Jara-Díaz, 2012; Bell and Wichiensin, 2012):

$$\frac{du_i(\cdot)}{dc} < 0, \; i = a, t$$

Apart from the demand elasticity with respect to highway capacity, the responses of the transit operator play critical roles in determining the marginal changes in the utilities. Mohring (1972) shows that if the transit operator tends to improve the level of service when more riders come, it would generate user benefits in the form of reduced waiting time that included in the total travel cost and even more riders, which is often noted as the Mohring Effect (e.g., van Reeven, 2008; Parry and Small, 2009; Savage and Small, 2010; Basso and Jara-Díaz, 2010, 2012). The Downs-Thomson Paradox refers to the opposite side of the Mohring Effect, which works as a vicious cycle resulted from the transit operator’s contractionary responses to the falling demand in the background of the two-mode system. The properties and impacts of different transit’s responsive strategies in terms of service
frequency and fare under various operating regimes are the main focuses of this paper.

3. The impact of highway expansion with responsive transit and imperfect substitutability under different transit operating regimes

When the highway capacity is being expanded, the transit’s responsive strategies and the degree of substitutability between modes are both essential in determining the changes in the travel utility of the commuters; therefore, the properties and interacting effect of the two factors are discussed and compared in three distinct scenarios: first-best, profit-maximizing and revenue-neutral.

We begin with the dissection of the transit strategy and modal split mechanism. Substitute Eqs.(1) and (2) into (3), the equilibrium volume allocation \((v_a, v_r)\) is determined by:

\[
\begin{align*}
  v_a &= g(v_a) \\
  v_r &= D - v_a
\end{align*}
\]

where \(g(v_a) = \frac{\ln(D-v_a) - \ln v_a - \lambda_n(a'-p) + \tau_n - mD)}{\lambda(m+\alpha)}.\)

An equilibrium volume allocation \((v_a, v_r)\) is locally stable if

\[
|g'(v_a)| < 1 \iff \lambda > \lambda_s = \frac{1}{\alpha + m} \rho
\]

where \(\rho = \frac{D}{(D-v_a)v_a}.\)

Eq.(7) provides a lower-bound of \(\lambda\), the degree of substitutability between the two travel modes, for the stability of an equilibrium volume allocation. An interpretation is that when the two modes can hardly substitute between each other, i.e., \(\lambda\) approaches 0, the modal choice becomes a random choice behavior and cannot stably stand; therefore, \(\lambda\) must be large enough. \(\rho\), on the other hand, is representing the degree of diverse of the traffic volume in the sense that the value of \(\rho\) increases with the difference of the traffic volume of the two modes.

Now we look at the relation between \((v_a, v_r)\) and the \((f, \tau, c)\) combination. According to Eq.(5), the marginal changes of the highway traffic volume \(v_a\) with respect to the change of \(f\) and \(\tau\) are as follows:

\[
\frac{\partial v_a}{\partial f} = -\frac{\lambda_n(D - v_a)v_a}{\lambda(m+\alpha)(D - v_a)v_a + D}
\]

\[
\frac{\partial v_a}{\partial \tau} = \frac{\lambda_n(D - v_a)}{\lambda(m+\alpha)(D - v_a)v_a + D}
\]
\[
\frac{\partial v_a}{\partial \tau_i} = \frac{\lambda (D - v_a) v_a}{\lambda (m + \alpha)(D - v_a) v_a + D}
\]  

(9)

Then it is easily checked that the following property holds:

\[
\frac{\partial v_a}{\partial f} \frac{\partial v_a}{\partial \tau_i} = \frac{\partial u_a}{\partial f} \frac{\partial u_a}{\partial \tau_i} = -n
\]

(10)

Eq.(10) indicates that the ratio of the marginal change in the equilibrium volume allocation incurred by one unit increase in frequency to that due to fare is equal to the ratio of the marginal utility by transit associated with the increase in frequency to that with fare. In fact, \( n \) is the marginal utility obtained by the transit commuter when the service frequency is increased by one unit, and the marginal utility loss caused by one unit increase in fare is \(-1\). Since the travel utility is proportional to the volume, the ratio of the marginal changes in equilibrium volume allocation remains to be \(-n\).

Meanwhile, the total change in the equilibrium volume allocation incurred by the change in highway capacity is given by:

\[
\frac{dv_a}{dc} = \delta \left( -n \frac{df}{dc} + \frac{d\tau_i}{dc} + \gamma \right)
\]

(11)

where \( \delta = \frac{\lambda v_a (D - v_a)}{\lambda v_a (m + \alpha)(D - v_a) + D} \).

Note that one of the purposes of this paper is to examine the transit operator’s strategies under different regimes in terms of their responses to highway expansion. In particular, the changes in service frequency and fare with respect to the highway capacity are of our interest. Therefore, Eq.(11) is shown in the form of the total derivative of highway traffic volume with respect to highway capacity. Therefore, the aggregate marginal change in travel utilities by both modes with respect to highway capacity are as follows:

\[
\frac{du_a}{dc} = \alpha \delta \left( n \frac{df}{dc} - \frac{d\tau_i}{dc} \right) + \gamma (1 - \alpha \delta)
\]

(12)

\[
\frac{du_t}{dc} = (1 - m \delta) \left( n \frac{df}{dc} - \frac{d\tau_i}{dc} \right) + m \gamma \delta
\]

(13)

Evidently, the ultimate impact of the highway capacity change on commuters’ travel utilities of both modes are dominated by the marginal change in transit frequency and fare due to capacity change, i.e., \( \frac{df}{dc} \) and \( \frac{d\tau_i}{dc} \), and also subject to the specific equilibrium state. Now we are ready to analyze the properties and impacts of transit’s strategies under various operating regimes.
3.1 The first-best transit strategies

In the first-best scenario, the transit service is optimally designed by the central planner to achieve the socially optimal volume allocation. Generally, the aggregate social welfare is the summation of consumer surplus and producer’s surplus. According to the assumptions above, the aggregate commuter’s travel utility constitutes the consumer surplus, while the producer’s surplus refers to transit operator’s net profit:

\[ SW(f, \tau_i) = u_a v_a + u_t v_t + \tau_i v_t - k(f) \]  \hspace{1cm} (14)

For given \( c \), the optimal \((f^*, \tau_i^*)\) combination is chosen to maximize the aggregate social welfare and the only constraint is the feasibility of the volume allocation. Since the ticket revenue, \( \tau_i v_t \), represents the welfare reallocation within the system, this term would be eliminated from the aggregate social welfare; instead, the optimal \( \tau_i^* \) is implicitly chosen to achieve the optimal volume allocation. The first-order conditions of Eq.(14) with respect to \((f^*, \tau_i^*)\) are:

\[ f^* : \frac{2(\alpha + m)}{n} k'(f^*) - nf^* + \gamma c - \tau_a - 2\alpha D = 0 \]  \hspace{1cm} (15)

\[ \tau_i^* : v_a^* = D - \frac{1}{n} k'(f^*) \]  \hspace{1cm} (16)

The optimal frequency \( f^* \) is pinned down by Eq.(15) and the optimal fare \( \tau_i^* \) is implicitly determined through \( v_a^* \) by Eq.(16). Besides, to ensure the feasibility of the volume allocation, one condition must be imposed:

\[ k'(f^*) \in (0, nD) \]  \hspace{1cm} (17)

According to Eqs.(12), (13), (15) and (16), the marginal changes in the travel utilities of both modes due to the highway capacity change are as follows:

\[ \frac{du_a^*(\cdot)}{dc} = \frac{n^2 - (\alpha + 2m) k^*(f^*)}{n^2 - 2(\alpha + m) k^*(f^*)} \cdot \gamma \]  \hspace{1cm} (18)

\[ \frac{du_t^*(\cdot)}{dc} = \left(\frac{1}{\delta} - m\right) \frac{\gamma k''(f^*)}{n^2 - 2(\alpha + m) k^*(f^*)} + \gamma \]  \hspace{1cm} (19)

where \( k''(f^*) \neq \frac{n^2}{2(\alpha + m)} \). Obviously, for the given equilibrium volume allocation, the properties of utility changes are determined by the degree of substitutability and the property of the operation cost.

By denoting \( \kappa_1 = \frac{n^2}{2(\alpha + m)} \), \( \kappa_2 = \frac{n^2}{\alpha + 2m} \), \( \kappa_3 = \frac{\beta^2}{m} \) and \( \lambda_u = \frac{k'(f^*)}{(\alpha + 2m) k^*(f^*) - nD^*} \), the properties of utility
change according to the ranges of $k''(f^*)$ for the first-best scenario under conditions (7) and (17) are summarized in Proposition 1:

**Proposition 1.** Under the first-best transit responsive strategies,

1. If $k^*(f^*) \in [0, \kappa_1]$, then $\frac{d\mu_c^i(\lambda)}{d\lambda} \geq 0$, $\frac{d\mu_t^i(\lambda)}{d\lambda} \geq 0$, implying that both the auto and transit commuter would benefit from the marginal growth of the highway capacity.

2. If $k^*(f^*) \in (\kappa_1, \kappa_2)$, then $\frac{d\mu_c^i(\lambda)}{d\lambda} < 0$, $\frac{d\mu_t^i(\lambda)}{d\lambda} < 0$, implying that commuter’s travel utility by either mode strictly decreases with the highway capacity. Any marginal highway expansion would result in the occurrence of the **Downs-Thomson Paradox** defined in Eq. (4).

3. If $k^*(f^*) \in [\kappa_2, \infty)$, then $\frac{d\mu_c^i(\lambda)}{d\lambda} \geq 0$, implying that auto commuters would surely benefit from the marginal growth of the highway capacity. Furthermore,

   (3a) If $k^*(f^*) \in [\kappa_2, \kappa_1]$ and $\lambda \in (\bar{\lambda}, \bar{\lambda}_u)$, then $\frac{d\mu_c^i(\lambda)}{d\lambda} < 0$; if $\lambda \in [\bar{\lambda}_u, \infty)$, then $\frac{d\mu_t^i(\lambda)}{d\lambda} \geq 0$.

   (3b) If $k^*(f^*) \in (\kappa_3, \infty)$, and $\lambda \in [\bar{\lambda}_z, \infty)$, then $\frac{d\mu_c^i(\lambda)}{d\lambda} \geq 0$.

**Proof.** The proof for Proposition 1 is relayed to Appendix B.

Proposition 1 implies that under the first-best transit strategies, there are intervals of $k^*(f^*)$ in which commuter’s travel utility decreases with the highway capacity for both auto and transit modes; however, such an interval for auto mode lies within the interval of transit. Therefore, in the joint interval where both utilities diminishes, $k^*(f^*) \in (\kappa_1, \kappa_2)$, any marginal increase of the highway capacity would result in the occurrence of the Downs-Thomson Paradox. Meanwhile, in the two-end of the interval of $k^*(f^*)$, any marginal highway expansion would be a Pareto-improvement.

Figure 2 is shown to further illustrate the results stated in Proposition 1. The domain spanned by the range of $k^*(f^*)$ and $\lambda$ is divided into four regions by the blue lines: In Region I, i.e., $k^*(f^*) \in [0, \kappa_1]$, the travel utilities by both modes are increasing; Region II stands for the area where $k^*(f^*) \in (\kappa_1, \kappa_2)$, and indicates the occurrence of the Downs-Thomson Paradox in terms of commuter’s travel utility; different directions of the travel utility changes of the two modes appear in Region III where $k^*(f^*) \in [\kappa_2, \infty)$ and $\lambda \in [0, \bar{\lambda}_u)$, and only the auto commuters benefit from the highway capacity expansion; Region IV represents the area
where $k^*(f^*) \in [k_2, \infty)$, but $\lambda \in [\lambda_n, \infty)$, and both the auto commuters and transit users benefit from the highway expansion. Each of the region is separated into two minor parts by a red line standing for the boundary condition for the stability of the equilibrium state, and all the feasible equilibrium states lie in the sub-regions (b) on the right hand-side of the red line are stable while those on the left hand-side turn out to be unstable, i.e., the sub-regions (a).

![Figure 2. Division of the feasible domain for the first-best scenario](image)

The division of the feasible domain and the corresponding boundaries are illustrated in Table 1 (the region number corresponds to Figure 2):
Table 1. Division of the feasible domain in Figure 2

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>([0, \lambda_s])</td>
<td>((\lambda_s, \infty))</td>
<td>([0, \lambda_s])</td>
<td>((\lambda_s, \lambda_i))</td>
</tr>
<tr>
<td>(k''(f^*))</td>
<td>([0, \kappa_t])</td>
<td>((\kappa_t, \kappa_2))</td>
<td>([\kappa_2, \kappa_3])</td>
<td>([\kappa_2, \infty))</td>
</tr>
</tbody>
</table>

Note: \(\lambda_m = \max \{\lambda_s, \lambda_i\}\).

Besides the marginal utilities, Table 2 further summarizes the properties of the transit responsive strategies on frequency and fare with respect to the highway capacity change in each region (proof see Appendix B):

Table 2. Properties of different regions in Figure 2

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>(df^*)  (dc)</td>
<td>+ +</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>(d\kappa^*)  (dc)</td>
<td>- +</td>
<td>+ -</td>
<td>+ -</td>
<td>+ -</td>
</tr>
<tr>
<td>(d\kappa^*(t))  (dc)</td>
<td>+ +</td>
<td>- -</td>
<td>+ +</td>
<td>+ +</td>
</tr>
<tr>
<td>(d\kappa^*(t))  (dc)</td>
<td>+ +</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
</tr>
<tr>
<td>stability</td>
<td>(\times \sqrt{\quad})</td>
<td>(\times \sqrt{\quad})</td>
<td>(\times \sqrt{\quad})</td>
<td>(\times \sqrt{\quad})</td>
</tr>
</tbody>
</table>

Generally, the change directions of frequency and both utilities are consistent (“+” or “−”) in each big region, i.e., I, II, III, IV, and different in between them; the marginal fare changes exhibit opposite signs between the sub-regions (a) and (b) within the same big region. This implies that when the highway capacity is being expanded, the welfare impact is the combined action of the transit responsive strategies, which is the consequence of trade-off between high revenue and low expense: more frequent service and lower fare attract more passengers, but lower frequency requires less investment.

The first-best transit frequency firstly increases with the highway capacity when \(k''(f^*)\) is small and tends to be decreasing thereafter. Since the marginal increment of the operation cost grows rapidly with frequency when the second-order marginal operation cost \(k''(f^*)\) becomes greater; therefore, it is economically unsustainable to further increase frequency when \(k''(f^*)\) is large. It is worth noting that the marginal change of the first-best transit fare is tightly related to the degree of substitutability in the sense that changing the direction at the boundary of \(\lambda = \lambda_s\) in each big region. As previously mentioned, the optimal \(\tau^*\) is implicitly
chosen to realize the optimal volume allocation \((v^*_a, v^*_r)\) and the relation between them is controlled by the degree of substitutability as Eqs. (5) and (6) indicate.

The above results also substantiate the crucial role of the degree of substitutability between the modes in influencing the welfare impacts of highway expansion. Given the equilibrium volume allocation, the degree of traffic volume diverse is determined and so as the value of \(\rho^*\). From Figure 2, it can be observed that the area of Region II is independent of \(\lambda\) and the area of Region IV depends on the value of \(\lambda\). This implies that under the first-best transit regime, the likelihood of the Downs-Thomson Paradox occurrence is independent of the degree of substitutability; but when the marginal operation cost grows rapidly with frequency, the capacity expansion is more likely to be a Pareto-improvement with higher degree of substitutability.

Furthermore, it is worth noting that Proposition 1 demonstrates the possibility of unintendedly impairing every commuter’s welfare by expanding the highway capacity and maintaining the first-best transit scheduling and pricing strategies (the occurrence of the Downs-Thomson Paradox under the first-best scenario).

### 3.2 Profit-maximizing transit strategies

Suppose the transit operator behave as a profit-maximizing monopoly, for given \(c\), the scheduling and pricing strategies are determined by:

\[
\max_{\tau} \pi(f, \tau) = \tau v_i - k(f) \tag{20}
\]

where \(\pi(f, \tau)\) denotes the transit operator’s profit, which is a function of frequency and fare. The first-order conditions for the optimal \(f^*\) and \(\tau^*_i\) are:

\[
f^*: \tau_i, \frac{\partial v_i}{\partial f} - k'(f) = 0 \tag{21}
\]

\[
\tau^*_i: v_i + \tau_i \frac{\partial v_i}{\partial \tau_i} = 0 \tag{22}
\]

According to Eqs.(8), (9), (12), (13), (21) and (22), the marginal changes in the travel utilities of both modes due to the highway capacity change are as follows:

\[
\frac{du^*_a}{dc} = \frac{\alpha \delta^* g k''(f^*)}{n^1 \delta^* - n \delta^*(f^*)\left[\lambda D(D - 2v^*_a)\tau^*_i + 2\mu^*_i\right]/\mu^*} + \gamma \tag{23}
\]

\[
\frac{du^*_r}{dc} = \frac{(1 - m \delta^*) g k''(f^*)}{n^1 \delta^* - n \delta^*(f^*)\left[\lambda D(D - 2v^*_a)\tau^*_i + 2\mu^*_i\right]/\mu^*} + \gamma \tag{24}
\]
where \( \mu^* = \left[ \lambda (m+a) \left( D - v^*_a \right) \nu^*_a + D \right]^2 \). Evidently, for given equilibrium volume allocation, the combined influence of the degree of substitutability and the operation cost determines the properties of \( u^*_a \) and \( u^*_t \). By denoting \( \theta_1 = \frac{n^2 \nu^*_a / p}{2(m+a) \kappa_2 / p + D} \), \( \theta_2 = \frac{n^2 \nu^*_a / p}{2(m+a) \kappa_2 / p + D} \) and

\[
\theta_3 = \frac{n^2 \kappa_2 \nu^*_a / p}{2(m+a) \kappa_2 / p + D - 1} \quad (0 < \theta_1 < \theta_2 < \theta_3),
\]

we have the following proposition for this scenario:

**Proposition 2.** Under profit-maximizing transit responsive strategies:

1. If \( k^*(f^*) \in [0, \theta_1] \cup [\theta_3, \infty] \), then \( \frac{du^*_a}{dk} \geq 0 \), \( \frac{du^*_t}{dk} \geq 0 \), implying that both the auto and transit commuter would benefit from the marginal growth of the highway capacity;

2. If \( k^*(f^*) \in [\theta_2, \theta_3] \), then \( \frac{du^*_a}{dk} \geq 0 \), \( \frac{du^*_t}{dk} \leq 0 \), implying that only the auto commuters benefit from the marginal growth of the highway capacity while the transit commuters suffer;

3. If \( k^*(f^*) \in (\theta_1, \theta_2) \), then \( \frac{du^*_a}{dk} < 0 \), \( \frac{du^*_t}{dk} < 0 \), implying that commuter’s travel utility by either mode strictly diminishes when there is a marginal growth of the highway capacity. In this case, any marginal highway expansion would result in the occurrence of the Downs-Thomson Paradox.

Proof. The proof for Proposition 2 is relayed to Appendix C.

Proposition 2 implies that under the profit-maximizing transit strategies, there are intervals of \( k^*(f^*) \) in which commuter’s travel utility decreases with the highway capacity for both auto and transit modes; however, such an interval for auto mode lies within the interval of transit, i.e., \( (\theta_1, \theta_2) \subseteq (\theta_1, \theta_3) \). Therefore, in the joint interval where both utilities diminishes, \( k^*(f^*) \in (\theta_1, \theta_2) \), any marginal increase of the highway capacity would result in the occurrence of the Downs-Thomson Paradox. Meanwhile, in the two-end of the interval of \( k^*(f^*) \), any marginal highway expansion would be a Pareto-improvement.

To further illustrate the welfare impacts under each case, Figure 3 is employed to show the divisions of the feasible domain and the corresponding properties of each region. The domain spanned by the range of \( k^*(f^*) \) and \( \lambda \) is divided into four regions by the blue curves. The directions of commuter’s utility changes are marked for each region. As it is shown, each region, i.e., I, II, III, IV, is separated into two minor parts by a red line standing for the boundary condition for the stability of the equilibrium state, and all the feasible equilibrium
states lie in the sub-regions (b) on the right hand-side of the red line are stable while those on the left hand-side turn out to be unstable, i.e., the sub-regions (a).

**Figure 3.** Division of the feasible domain for the profit-maximizing scenario

The division of the feasible domain and the corresponding boundaries are illustrated in Table 3 (the region number corresponds to Figure 3):

**Table 3.** Division of the feasible domain in Figure 3

<table>
<thead>
<tr>
<th>Region</th>
<th>I (a)</th>
<th>II (a)</th>
<th>III (a)</th>
<th>IV (a)</th>
<th>I (b)</th>
<th>II (b)</th>
<th>III (b)</th>
<th>IV (b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>$[0, \lambda_z]$</td>
<td>$(\lambda_z, \infty)$</td>
<td>$[0, \lambda_z]$</td>
<td>$(\lambda_z, \infty)$</td>
<td>$[0, \lambda_z]$</td>
<td>$(\lambda_z, \infty)$</td>
<td>$[0, \lambda_z]$</td>
<td>$(\lambda_z, \infty)$</td>
</tr>
<tr>
<td>$k^<em>(f^</em>)$</td>
<td>$[0, 0]$</td>
<td>$(0, 0)$</td>
<td>$(0, 0)$</td>
<td>$(0, 0)$</td>
<td>$[0, 0]$</td>
<td>$(0, 0)$</td>
<td>$[0, 0]$</td>
<td>$(0, 0)$</td>
</tr>
</tbody>
</table>

Besides the marginal utilities, Table 4 further summarizes the properties of the transit responsive strategies on frequency and fare with respect to the highway capacity change in each region (proof see Appendix C):
Table 4. Properties of different regions in Figure 3

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a b</td>
<td>a b</td>
<td>a b</td>
<td>a b</td>
</tr>
<tr>
<td>(\frac{df^*}{dc})</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\frac{dc^*}{dc})</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\frac{df_c^* (t)}{dc})</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(\frac{dc_c^* (t)}{dc})</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Stability</td>
<td>(\times \sqrt{\cdot})</td>
<td>(\times \sqrt{\cdot})</td>
<td>(\times \sqrt{\cdot})</td>
<td>(\times \sqrt{\cdot})</td>
</tr>
</tbody>
</table>

It is shown that the change directions of frequency and fare are consistent, but the marginal changes of commuters’ utilities exhibit different properties even with the same frequency and fare responses, e.g., Region III and IV. This again demonstrates that when the highway capacity is being expanded, the welfare impact is the combined action of the transit responsive strategies. When \(k^*(f^*)\) is relatively small, the optimal frequency increases with the highway capacity. With the growth of the increase rate of the marginal operation cost \(k^*(f^*)\), the marginal increment of the operation cost grows rapidly with frequency; therefore, the frequency is reduced when \(k^*(f^*)\) grows. Under the profit-maximizing transit strategies, marginal revenue should compensate the marginal increase of the operation cost; therefore, the optimal fare changes in accord with the optimal frequency.

The degree of substitutability between the modes plays an indispensable role in determining the commuter’s utility because of the nonlinear boundary conditions for each regions. Given equilibrium volume allocation, Figure 3 shows that the area of Region II increases with the value of \(\lambda\), implying that the likelihood of the Downs-Thomson Paradox occurrence grows with higher degree of substitutability between the modes.

3.3 Revenue-neutral transit strategies

Now we consider a second-best scenario where the transit operator optimizes the service frequency and fare to maximize commuters’ welfare with the budget constraint. This scenario is more applicable and sustainable to maintain the transit service and provide benefits to the commuters simultaneously. The problem for the transit operator becomes:

\[
\begin{align*}
\max_{f, \tau} & \quad SW(f, \tau) = u_a v_a + u_t v_t \\
\text{s.t.} & \quad \tau v_t - k(f) = 0
\end{align*}
\]

(25)

The first-order conditions for the optimal \(f^*\) and \(\tau^*_v\) are:
\[ f^* : (-2\alpha v_a + \gamma c - \tau_a + 2mv_v - nf + \tau_v - \varepsilon \tau_v)\delta - v + \frac{1}{n}ek^*(f) = 0 \]  
(26)

\[ \tau^* : (-2\alpha v_a + \gamma c - \tau_a + 2mv_v - nf + \tau_v - \varepsilon \tau_v)\delta - v + \varepsilon v = 0 \]  
(27)

where \( \varepsilon \) is the Lagrangian multiplier of the constraint. According to Eqs. (8), (9), (12), (13), (26) and (27), the marginal changes in the travel utilities of both modes due to the highway capacity change are as follows:

\[ \frac{du_a^*}{dc} = \gamma \left[ \frac{k(f^*)}{k^{\alpha}(f^*)} - \frac{(m+\varepsilon \tau_v)}{n^2} \right] \left[ \frac{k(f^*)}{k^{\alpha}(f^*)} - \frac{m+\alpha + \varepsilon v}{n^2} \right]^{-1} \]  
(28)

\[ \frac{du_t^*}{dc} = \gamma \left[ \frac{k(f^*)}{k^{\alpha}(f^*)} - \frac{m}{n^2} \right] \left[ \frac{k(f^*)}{k^{\alpha}(f^*)} - \frac{m+\alpha + \varepsilon v}{n^2} \right]^{-1} \]  
(29)

where \( k(f^*)/k^{\alpha}(f^*) \neq m+\alpha + \varepsilon v/n^2 \). As expected, for given equilibrium volume allocation, the properties of \( u_a^* \) and \( u_t^* \) are dominated by the degree of substitutability and the operation cost. Denote \( \phi_1 = m/n^2, \phi_2 = (m+\varepsilon \tau_v)/n^2 \) and \( \phi_3 = (m+\alpha + \varepsilon v)/n^2 \) (\( 0 < \phi_1 < \phi_2 < \phi_3 \)), then we have the following proposition:

**Proposition 3.** Under the revenue-neutral scenario, the transit operator is considered as a social welfare maximizer under the zero-profit constraint, and:

(1) If \( k(f^*)/k^{\alpha}(f^*) \in [0,\phi_1] \cup (\phi_3,\infty) \), then \( \frac{du_a^*}{dc} \geq 0, \frac{du_t^*}{dc} \geq 0 \), implying that both the auto and transit commuter would benefit from the marginal growth of the highway capacity;

(2) If \( k(f^*)/k^{\alpha}(f^*) \in [\phi_1,\phi_2] \), then \( \frac{du_a^*}{dc} \geq 0, \frac{du_t^*}{dc} \leq 0 \), implying that only the auto commuters benefit from the marginal growth of the highway capacity while the transit commuters suffer;

(3) If \( k(f^*)/k^{\alpha}(f^*) \in (\phi_2,\phi_3) \), then \( \frac{du_a^*}{dc} < 0, \frac{du_t^*}{dc} < 0 \), implying that commuter’s travel utility by either mode strictly diminishes when there is a marginal growth of the highway capacity. In this case, any marginal highway expansion would result in the occurrence of the **Downs-Thomson Paradox.**

**Proof.** The proof for Proposition 3 is relayed to Appendix D.

Interestingly we found that under the second-best transit regime, the properties of commuters’ utilities are determined by the ratio of operation cost and the squared marginal cost. This implies that the revenue-neutral transit operator would if either the absolute value of the
operation cost or the marginal increase rate of it is relatively large, the combination effects of the capacity expansion and responsive transit strategies are socially preferable. In the medium intervals of $k(f^*)/k^2(f^*)$, commuter’s travel utility decreases with the highway capacity for both auto and transit modes; and as expected, such an interval for auto mode lies within the interval of transit, i.e., $(\phi_2, \phi_3) \subseteq (\phi_1, \phi_3)$.

Figure 4 is employed to show the divisions of the feasible domain and the corresponding properties of each region. The domain spanned by the range of $k(f^*)/k^2(f^*)$ and $\lambda$ is divided into four regions by the blue curves. As what we previously did, each region is separated into two minor parts by a red line standing for the boundary condition for the stability of the equilibrium state, and all the feasible equilibrium states lie on the right hand-side of it are stable while those on the left hand-side turn out to be unstable. Note that the vertical distance between $\phi_2$ and $\phi_3$ is constant.

Figure 4. Division of the feasible domain for the second-best scenario
The division of the feasible domain and the corresponding boundaries are illustrated in Table 5 (the region number corresponds to Figure 4):

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ</td>
<td>[0,λa)</td>
<td>(λa,∞)</td>
<td>[0,λb)</td>
<td>(λb,∞)</td>
</tr>
<tr>
<td>k(f*)/k(2)k(f*)</td>
<td>[0,φ1]</td>
<td>[φ1,φ2]</td>
<td>(φ2,φ3)</td>
<td>(φ3,∞)</td>
</tr>
</tbody>
</table>

Besides the marginal utilities, Table 6 further summarizes the property of the marginal change of transit frequency with respect to the highway capacity change in each region (proof see Appendix D):

<table>
<thead>
<tr>
<th>Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>dτ*/dc</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>dσc(·)/dc</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>dσc(·)/dc</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>stability</td>
<td>×√</td>
<td>×√</td>
<td>×√</td>
<td>×√</td>
</tr>
</tbody>
</table>

Note: If k(f*)/k(2)k(f*) < \min\{1/k*(f*),φ3\}, or k(f*)/k(2)k(f*) > \max\{1/k*(f*),φ3\}, then dτ*/dc < 0; otherwise, dτ*/dc ≥ 0.

As is well known, if the marginal operation cost is high, the absolute value of operation cost grows rapidly with the frequency. Therefore, the optimal frequency decreases with the highway capacity in the ranges where the increasing rate of operation cost is relatively large compared to the absolute value of it, i.e., Region I, II and III; while increases in the ranges with lower marginal cost, i.e., Region IV. However, the optimal fare increases with the highway capacity only in the range where neither the operation cost nor its increasing rate is disproportionately large. It is reasonable because with small k(f*) and relatively large k(·), the operation cost is low and the transit passenger volume is high, and the revenue-neutral transit operator would reduce the fare; while with large k(f*) and relatively small k(·), the absolute value of operation cost is high but growing slowly, and the transit operator has no incentive to raise the fare.
In summary, the revenue-neutral transit operator makes decisions on frequency and fare considering both the absolute value of operation cost and its increasing rate. Noting with this, it is not difficult to understand the properties of commuters’ utilities. With small \( k(f^*) \) and relatively large \( k'(f^*) \), both frequency and fare would be reduced; since the increase of transit fare should cover the rapid growth of the operation cost, the negative impact of reducing frequency would be dominated by the positive impact of reducing fare. With large \( k(f^*) \) and relatively small \( k'(f^*) \), transit operator would provide more frequent service while cut the fare, and both strategies would generate benefits to the commuters of both modes.

The influence of the degree of substitutability between the modes is evident by observing Figure 4: given equilibrium volume allocation, the highway expansion is more likely to be a Pareto-improvement with higher degree of substitutability when the absolute value of operation cost is relatively high compared to the increasing rate of it. The boundary condition of Downs-Thomson Paradox occurrence region is related to the degree of substitutability; but not for the likelihood of its occurrence since the vertical distance between \( \phi_2 \) and \( \phi_3 \) is constant.

4. Conclusion and discussions

This paper presents an analytical study on the impact of highway expansion with responsive transit and imperfect substitutability under different transit operating regimes. Three types of market regimes of the transit service are considered: the first-best scenario, the profit-maximizing monopoly and the second-best scenario with budget constraint.

Generally, in the presence of imperfect substitutability, the changes of the commuters’ utilities of the two alternative modes do not always accord with each other. Interestingly we found that in all scenarios, auto commuters are more likely to benefit from the highway expansion compared to the transit commuters, and transit commuters’ utility would always decrease with the highway capacity as long as the auto commuters’ utility diminishes. Boundary conditions are established for the Pareto-improving outcome (all commuters are better-off), the occurrence of the Downs-Thomson Paradox (all commuters are worse-off), and other welfare imbalance situations.

More importantly, this paper characterizes how the transit operator reacts to the highway expansion differently under distinct market regimes and its influences on the commuters’ welfare. Without the budget constraint (i.e., in the first-best and profit-maximizing scenarios), transit profit acts one component or even the whole part of the objective of the maximizing problems; and accordingly, the properties of transit responsive strategies and commuters’
welfare changes are found to be closely related to the increasing rate of the marginal operation cost. Particularly, the optimal frequency increases with the highway capacity when the marginal operation cost grows relatively slowly with the frequency, while decreases thereafter. The changes of the optimal fare are fairly different: in the first-best scenario, the optimal fare is indirectly determined through its relation with the optimal volume allocation, and therefore it tightly relies on the degree of substitutability between the modes; in the profit-maximizing scenario, the optimal fare changes consistently with the optimal frequency since the marginal revenue should cover the marginal increase of the transit operation cost.

In the second-best scenario, the transit operator is considered as a social welfare maximizer with budget constraint. As reactions to highway expansion, the transit frequency and fare should be designed to achieve a new equilibrium volume allocation such that the ticket revenue compensates the operation cost. The foregoing analysis shows that the revenue-neutral transit operator makes decisions on frequency and fare considering both the absolute value of operation cost and its increasing rate. The optimal frequency decreases with the highway capacity in the ranges where the increasing rate of operation cost is relatively large compared to the absolute value of it, while increases in the ranges with lower marginal cost. Likewise, the optimal fare increases with the highway capacity only in the range where neither the operation cost nor its increasing rate is disproportionately large.

In parallel, the results also reveal the role of substitutability in determining social welfare under the interaction of highway capacity expansion and transit responsive strategies. First of all, the boundary conditions, which categorized the properties of the commuters’ welfare changes, rely on the relations between the transit operation cost and the degree of substitutability, especially when the relations are nonlinear. Besides, the possibilities of the Pareto-improving outcome, the Downs-Thomson Paradox occurrence and other welfare imbalance situations are closely related to the degree of substitutability between the travel modes. To be specific, the likelihood of the Pareto-improving outcome increases with it in all the scenarios, and the possibility of the Downs-Thomson Paradox occurrence also increases with it under profit-maximizing transit strategies while independent of it under social welfare-maximizing transit strategies with or without budget constraint.

The results provide insights for public policy especially where public transit serves as an irreducible substitute to the private automobile. In most of the densely populated countries or areas, improving the road capacity and enhancing transit service are equally valued to the central planner for the sake of social benefits. This paper analytically shows that to avoid paradoxical outcomes, comprehensive preliminary studies on commuters’ choice behavior, current network performance and transit operation conditions are necessary.
Appendix A. Proof of the uniqueness of the user equilibrium volume allocation.

Substitute Eqs.(1), (2) into Eq.(3), then for given $c$, $f$, and $\tau$, the volume allocation $(v_a, v_t)$ at the user equilibrium state is pinned down by:

\[
\begin{cases}
h(v_a) = 0 \\
v_t = D - v_a
\end{cases}
\]

where $h(v_a) = \lambda \left[ (m + \alpha) v_a + nf - c + \gamma c + \tau - \tau - mD \right] + \ln v_a - \ln(D - v_a)$ is a function of $v_a$, which is a monotonically increasing function of $v_a$ with:

\[
h'(v_a) = \lambda (m + \alpha) + \frac{D}{v_a (D - v_a)} > 0
\]

Therefore, for given $c$, $f$, and $\tau$, there would be one and only one solution to $h(v_a) = 0$, which completes the proof for the uniqueness of the user equilibrium volume allocation.

Appendix B. Proof of Proposition 1.

According to Eqs.(15) and (16), the marginal change in the optimal frequency, fare and the corresponding total change in the highway traffic volume due to the change in the highway capacity are:

\[
\frac{df^*}{dc} = \frac{n \gamma}{n^2 - 2(\alpha + m) k^*(f^*)}
\]

\[
\frac{d\tau^*_t}{dc} = \left[ n - \frac{1}{n \delta} k^*(f^*) \right] \frac{df^*}{dc} - \gamma
\]

\[
\frac{dv^*_a}{dc} = -\frac{1}{n} k^*(f^*) \frac{df^*}{dc}
\]

Thereby, the first-best responsive strategies are as follows:

\[
\frac{df^*}{dc} \geq 0 \iff \frac{dv^*_a}{dc} \geq 0 \iff k^*(f^*) \in \left[ 0, \frac{n^2}{2(\alpha + m)} \right]
\]

\[
\frac{df^*}{dc} < 0 \iff \frac{dv^*_a}{dc} < 0 \iff k^*(f^*) \in \left( \frac{n^2}{2(\alpha + m)}, \infty \right)
\]

\[
\frac{d\tau^*_t}{dc} \geq 0 \iff \lambda \in [\lambda_s, \infty), k^*(f^*) \in \left[ 0, \frac{n^2}{2(\alpha + m)} \right]
\]

or $\lambda \in [0, \lambda_s], k^*(f^*) \in \left( \frac{n^2}{2(\alpha + m)}, \infty \right)$
\[
\frac{d\tau_i^*}{dc} < 0 \iff \lambda \in [0, \lambda_s) \cup \left(\frac{n^2}{2(\alpha + m)}, \infty\right) \\
\text{or } \lambda \in (\lambda_s, \infty) \cap \left[0, \frac{n^2}{2(\alpha + m)}\right]
\]

Then the marginal change in the auto travel utility with respect to the changed highway capacity is characterized as:

\[
\frac{du^*_a(\cdot)}{dc} = n^2 - (\alpha + 2m)k^*(f^*) - 2(\alpha + m)k^*(f^*).\gamma
\]

Therefore, the conditions for the positive or negative changes are given by:

\[
\frac{du^*_a(\cdot)}{dc} \geq 0 \iff k^*(f^*) \in \left[0, \frac{n^2}{2(\alpha + m)}\right] \cup \left(\frac{n^2}{\alpha + 2m}, \infty\right)
\]

\[
\frac{du^*_a(\cdot)}{dc} < 0 \iff k^*(f^*) \in \left(\frac{n^2}{2(\alpha + m)}, \frac{n^2}{\alpha + 2m}\right)
\]

As for the marginal transit travel utility change with respect to the highway capacity change and the corresponding conditions are given as follows:

\[
\frac{du^*_t(\cdot)}{dc} = \frac{\lambda v^*_a(D - v^*_o)\left[n^2 - (\alpha + 2m)k^*(f^*) + Dk^*(f^*)\right]}{\lambda v^*_o(D - v^*_o)\left[n^2 - 2(\alpha + m)k^*(f^*)\right]}\gamma
\]

\[
\frac{du^*_t(\cdot)}{dc} \geq 0 \iff k^*(f^*) \in \left[0, \frac{n^2}{2(\alpha + m)}\right] \cup \left(\frac{n^2}{\alpha + 2m}, \infty\right), \text{ or } k^*(f^*) \in \left(\frac{n^2}{\alpha + 2m}, \infty\right), \text{ and } \lambda \in [\lambda_u, \infty).
\]

\[
\frac{du^*_t(\cdot)}{dc} < 0 \iff k^*(f^*) \in \left(\frac{n^2}{2(\alpha + m)}, \frac{n^2}{\alpha + 2m}\right) \cup \left(\frac{n^2}{\alpha + 2m}, \infty\right), \text{ and } \lambda \in (0, \lambda_u).
\]

where \(\lambda_u = \left(\frac{\gamma_t(f^*)}{(\alpha + 2m)k^*(f^*)}\right)^{-1}B^*\).

Note that the stability of the equilibrium volume allocation requires \(\lambda > \lambda_s\) (see Eq.(7)), and the relations with \(\lambda_u\) are:

\[
\lambda_u > \lambda_s \iff k^*(f^*) \in \left(\frac{n^2}{\alpha + 2m}, \frac{n^2}{m}\right)
\]

\[
\lambda_u \leq \lambda_s \iff k^*(f^*) \in \left[0, \frac{n^2}{\alpha + 2m}\right] \cup \left[\frac{n^2}{m}, \infty\right)
\]

This completes the proof.
Appendix C. Proof of Proposition 2.

According to Eqs.(21) and (22), the marginal change in the optimal frequency, fare and the corresponding total change in the highway traffic volume due to the change in the highway capacity are:

\[
\frac{df^*}{dc} = \frac{n\gamma\delta^*\mu^*}{n^2\delta^*\mu^* - [2\mu^* + \lambda D(D-2v_a^*)]k^*(f^*)}
\]

\[
\frac{d\tau^*_i}{dc} = \frac{\lambda(m+\alpha)(D-v_a^*)v_a^* + \frac{D-v_a^*}{v_a^*}D}{n\lambda v_a^*(D-v_a^*)}k^*(f^*)\frac{df^*}{dc}
\]

\[
\frac{dv_a^*}{dc} = -\frac{1}{n}k^*(f^*)\frac{df^*}{dc}
\]

Then the marginal change in the auto travel utility with respect to the changed highway capacity is characterized as:

\[
\frac{du^*_a}{dc} = \frac{[2\mu^* + \lambda D(D-2v_a^*)]\tau_i^* - \alpha\delta^*\mu^*]k^*(f^*) - n^2\delta^*\mu^*}{[2\mu^* + \lambda D(D-2v_a^*)]\tau_i^* - n^2\delta^*\mu^*} . \gamma
\]

\[
\frac{du^*_i}{dc} = \frac{[2\mu^* + \lambda D(D-2v_a^*)]v_a^* - (1-m\delta^*)\mu^*]k^*(f^*) - n^2\delta^*\mu^*}{[2\mu^* + \lambda D(D-2v_a^*)]v_a^* - n^2\delta^*\mu^*} . \gamma
\]

Denote \( \theta_1 = \frac{n^2\delta^*\mu^*}{2\mu^* + \lambda D(D-2v_a^*)} \), \( \theta_2 = \frac{n^2\delta^*\mu^*}{2\mu^* + \lambda D(D-2v_a^*)} - \alpha\delta^*\mu^* \) and \( \theta_3 = \frac{n^2\delta^*\mu^*}{2\mu^* + \lambda D(D-2v_a^*)} - (1-m\delta^*)\mu^* \), then we have:

\[
\frac{du^*_a}{dc} \geq 0 \iff k^*(f^*) \in [0, \theta_1) \cup [\theta_2, \infty)
\]

\[
\frac{du^*_a}{dc} < 0 \iff k^*(f^*) \in (\theta_1, \theta_2)
\]

\[
\frac{du^*_i}{dc} \geq 0 \iff k^*(f^*) \in [0, \theta_1) \cup [\theta_3, \infty)
\]

\[
\frac{du^*_i}{dc} < 0 \iff k^*(f^*) \in (\theta_1, \theta_3)
\]

\[
\frac{df^*}{dc} \geq 0 \iff \frac{d\tau_i^*}{dc} \geq 0 \iff k^*(f^*) \in [0, \theta_1)
\]

\[
\frac{df^*}{dc} < 0 \iff \frac{d\tau_i^*}{dc} < 0 \iff k^*(f^*) \in (\theta_1, \infty)
\]

Note that:
\[ \theta_1 = \frac{n^2 \lambda v^*_a / \rho^*}{2(m + \alpha) \lambda v^*_a / \rho^* + D} \]
\[ \theta_2 = \frac{n^2 \lambda v^*_a / \rho^*}{(2m + \alpha) \lambda v^*_a / \rho^* + D} \]
\[ \theta_3 = \frac{n^2 \lambda v^*_a / \rho^*}{(2m + \alpha) \lambda v^*_a / \rho^* + D - 1} \]

This completes the proof.

**Appendix D.** Proof of Proposition 3.

According to Eqs. (26) and (27), the marginal change in the optimal frequency, fare and the corresponding total change in the highway traffic volume due to the change in the highway capacity are:

\[
\frac{df^*}{dc} = \frac{\gamma}{nk^*(f^*)} \left[ \frac{k(f^*)}{k^2(f^*) - \frac{m + \alpha + \frac{\nu}{\lambda}}{n^2}} \right]^{-1}
\]
\[
\frac{d\tau^*_i}{dc} = -\gamma \left[ \frac{k(f^*)}{k^2(f^*) - \frac{1}{k^*(f^*)}} \right] \left[ \frac{k(f^*)}{k^2(f^*) - \frac{m + \alpha + \frac{\nu}{\lambda}}{n^2}} \right]^{-1}
\]
\[
\frac{dv^*_a}{dc} = -\frac{1}{n} \frac{k^*(f^*)}{d\tau^*_i / dc}
\]

Then the marginal change in the auto travel utility with respect to the changed highway capacity is characterized as:

\[
\frac{du^*_a}{dc} = \gamma \left[ \frac{k(f^*)}{k^2(f^*) - \frac{m + \alpha + \frac{\nu}{\lambda}}{n^2}} \right]^{-1}
\]
\[
\frac{du^*_i}{dc} = \gamma \left[ \frac{k(f^*)}{k^2(f^*) - \frac{m + \alpha + \frac{\nu}{\lambda}}{n^2}} \right]^{-1}
\]

Denote \( \phi_1 = m / n^2 \), \( \phi_2 = (m + \alpha + \frac{\nu}{\lambda}) / n^2 \) and \( \phi_3 = (m + \alpha + \frac{\nu}{\lambda}) / n^2 \), then we have:

\[
\frac{du^*_a}{dc} \geq 0 \iff k(f^*) / k^2(f^*) \in [0, \phi_2] \cup (\phi_3, \infty)
\]
\[
\frac{du^*_i}{dc} < 0 \iff k(f^*) / k^2(f^*) \in (\phi_2, \phi_3)
\]
\[
\frac{du^*_i}{dc} \geq 0 \iff k(f^*) / k^2(f^*) \in [0, \phi_1] \cup (\phi_3, \infty)
\]
\[
\frac{du^*}{dc} < 0 \iff k\left(f^*\right)/k^\alpha\left(f^*\right) \in \left(\phi_1, \phi_3\right)
\]

\[
\frac{df^*}{dc} \geq 0 \iff k\left(f^*\right)/k^\alpha\left(f^*\right) \in \left[\phi_3, \infty\right)
\]

\[
\frac{df^*}{dc} < 0 \iff k\left(f^*\right)/k^\alpha\left(f^*\right) \in \left[0, \phi_3\right)
\]

The marginal change of the transit fare depends not only on the operation cost itself, but also the second-order marginal operation cost: If \(k\left(f^*\right)/k^\alpha\left(f^*\right) < \min\left\{1/k^\alpha(f^*), \phi_3\right\}\) or \(k\left(f^*\right)/k^\alpha\left(f^*\right) > \max\left\{1/k^\alpha(f^*), \phi_3\right\}\), then \(d\tau^*_i/dc < 0\); otherwise, \(d\tau^*_i/dc \geq 0\).
References


