Optimizing transit service with highway expansion when the two modes are imperfect substitutes

Fangni Zhang\textsuperscript{a,*} Hai Yang\textsuperscript{a} and Robin Lindsey\textsuperscript{b}

Research background

The scale economies of public transit (or the “Mohring effect”) suggests that when demand rises the service improvement will reduce the individual travel cost (Mohring, 1972). However, this loop may not sustain if the drop of transit demand is caused by the expansion of highway, which serves as a substitute to the transit mode. Instead of reducing its service, the transit operator may possibly improve the service to attract users. How a transit operator reacts to highway policy and the resulting ridership/travel cost changes are obscure when the alternative (highway) is improved and the bimodal interaction comes into play. Efforts have been extensively devoted to investigating transit service optimization in a single, isolated transit system; yet the impact of the bimodal interaction on transit policy needs to be unveiled.

The standard remedy to urban congestion, i.e., expanding highway network, has sparked a lot of controversy in the urban transportation field. Downs (1962) and Thomson (1977) suggested that an increase in highway capacity, by causing demand shifts from public transit to private transport, could lead to a new traffic equilibrium where the travel costs by both modes are higher, i.e., the “Downs-Thomson paradox (D-T paradox)”. The occurrence of the paradox highlights the necessity and importance of trans-modal integrated transport planning; however, few studies have showed how transit responds to highway expansion.

Recently Zhang et al. (2014) explored the occurrence of the D-T paradox considering transit responses with different economic objectives. In spite of the progress they have achieved, their model is built based on the D-T equilibrium, regarding the two modes as perfect substitutes. Nonetheless, all the cross-price elasticity estimates from empirical evidence are small in magnitude (between 0.08 and 0.7, according to Hensher (1998); 0.05 in the short run, and increase to 0.3 to 0.4 in the long run by Dargay and Hanly (1999)). This suggests, not surprisingly, that automobile and transit trips are highly imperfect substitutes. As a feature of the system with imperfect substitutability, utilities obtained from different alternatives are affected inconsistently by the changes of the system characteristics (e.g., highway capacity, transit frequency and fare). Therefore, the D-T equilibrium analysis which requires the generalized travel costs for the both modes to be equal at any equilibrium state, used in Zhang et al. (2014), does not apply. And the system-wide impact of highway expansion with bimodal interaction is more intriguing but relevant when the two modes are imperfectly substitutable.

This paper formulates the bimodal equilibrium problem with imperfect substitutability; explores how transit operator responds to highway expansion under alternative economic regimes; examines the impacts of the bimodal interaction on the equilibrium mode share and individual costs; verifies the impact of the Mohring effect; and establishes conditions for the

\textsuperscript{a} Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.

\textsuperscript{b} Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, British Columbia V6T 1Z2, Canada.
existence of the D-T paradox in the bimodal context. Also we intend to take into account external demand elasticity by allowing the total demand to be elastic when formulating the bimodal equilibrium with imperfect substitutability. Besides, to understand the impact of economic regime on transit policy, different settings are considered for the transit operation, in which the transit operator aims to maximize profit or social welfare, or maximize consumers’ surplus while keeping breakeven. Rather than solely focusing on the changes of individual cost (as in Zhang et al., 2014), this paper extends the scope of the system-wide welfare effect of the bimodal interaction to the aggregate level, i.e., in terms of the impact on consumers’ surplus, operator’s profit and the overall social welfare.

**Model setup**

The demand for travel with automobile is \( v_a = g(p_a, p_b) \), and public transit \( v_b = h(p_a, p_b) \), where \( p_i \) is the generalized cost of mode \( i \), \( i = a, b \) with ‘\( a \)’ stands for auto mode and ‘\( b \)’ for transit. The the direct demand elasticities with respect to the generalized costs are negative, and \( \frac{\partial v_a}{\partial p_a} < 0, \frac{\partial v_b}{\partial p_b} < 0 \). The cross demand elasticities are positive, meaning that \( \frac{\partial v_a}{\partial p_b} > 0, \frac{\partial v_b}{\partial p_a} > 0 \).

And the travel demand is more sensitive to its own cost rather than the alternative one, i.e., \( \frac{\partial v_a}{\partial p_a} - \frac{\partial v_a}{\partial p_b} \geq 0 \). The generalized cost is the sum of monetary cost and travel time cost:

\[
\begin{align*}
  p_a &= \tau_a + c_a(v_a, s), \\
  p_b &= \tau_b + c_b(v_b, f),
\end{align*}
\]

where \( \tau_a \) and \( \tau_b \) represent the monetary costs associated with driving a private car and taking public transit, respectively. \( c_a(v_a, s) \) is the travel time cost by auto mode, \( \frac{\partial c_a}{\partial s} > 0 \) and \( \frac{\partial c_a}{\partial v_a} < 0 \). \( c_b(v_b, f) \) is the transit travel time cost. Similarly, \( \frac{\partial c_b}{\partial s} > 0 \) and \( \frac{\partial c_b}{\partial v_b} < 0 \). For the transit mode, we assume the transit service is provided by a single operator, who selects its service frequency \( f \) and fare \( \tau_b \) for its economic objective. The operational cost associated with the frequency, \( k(f) \), is an increasing, convex and twice differentiable function. For the auto mode, highway capacity \( s \) is regarded as a continuous variable. We assume the change of \( s \) is exogenously determined, and the investment cost is not taken into account. Besides, the monetary cost \( \tau_a \) is assumed to be a constant.

**The impact of highway expansion in a bimodal system with imperfect substitutability**

The “Downs-Thomson paradox (D-T paradox)” occurs if:

\[
\frac{dp_i}{ds} > 0, \ i = a, b.
\]

where \( \frac{dp_i}{ds} \) is the marginal generalized cost of mode \( i \) with respect to highway capacity \( s \).

To represent the marginal transit cost imposed by service adjustments, we denote

\[
\delta = \frac{\partial c_b}{\partial f} \frac{df}{ds} + \frac{d\tau_b}{ds},
\]

- 2 -
where \( \frac{df}{dx} \) and \( \frac{d^n s}{dx} \) representing the marginal adjustments of \( f \) and \( \tau_b \) with respect to \( s \), respectively. Then \( \delta = \frac{\delta s}{\delta f} + \frac{d^n s}{dx} \) is the total marginal cost imposed on a transit user that is directly caused by the frequency and fare adjustments. A negative \( \delta \) means that the direct transit cost caused by the adjustments is negative (i.e., the transit service is generally improved), while \( \delta \geq 0 \) implies the adjustments directly enlarge the transit travel cost. Since \( \delta \) characterizes the direction and rate of transit adjustment, it is referred to as the “Impact Factor of Transit Adjustment (IFTA)” in the following analysis.

We intend to verify the existence of the cyclic phenomenon described in Mohring (1972) when the other alternative mode is improved in the bimodal context. In our setting, the impact of transit service adjustment on the ridership is dominated by the counteractive influence of highway expansion if

\[
\delta \cdot \frac{d_\nu}{ds} > 0 ,
\]

meaning that improving transit service (\( \delta < 0 \)) cannot increase transit travel demand (\( v_b \downarrow \)); instead, demand can rise (\( v_b \uparrow \)) when the direct marginal cost is positive (\( \delta > 0 \)).

Table 1 summarizes the changes of travel cost and demand in different intervals of the IFTA when there is a highway expansion.

<table>
<thead>
<tr>
<th>( \delta )</th>
<th>(-\infty, \delta_1 )</th>
<th>( \delta_1, \delta )</th>
<th>( \delta, 0 )</th>
<th>( 0, \delta_2 )</th>
<th>( \delta_2, \delta )</th>
<th>( \delta, +\infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_a )</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
</tr>
<tr>
<td>( p_b )</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
</tr>
<tr>
<td>( v_a )</td>
<td>\downarrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
<td>\uparrow</td>
</tr>
<tr>
<td>( v_b )</td>
<td>\uparrow</td>
<td>\uparrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
<td>\downarrow</td>
</tr>
<tr>
<td>( \delta \cdot \frac{d_\nu}{ds} )</td>
<td>( - )</td>
<td>( - )</td>
<td>( + )</td>
<td>( - )</td>
<td>( - )</td>
<td>( - )</td>
</tr>
<tr>
<td>D-T paradox</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \checkmark )</td>
</tr>
</tbody>
</table>

Note: \( \delta_1 = -\frac{\omega_v}{\omega_p} \left( 1 - \frac{\omega_v}{\omega_p} \right) / \frac{\omega_v}{\omega_p} \), \( \delta_2 = -\frac{\omega_v}{\omega_p} \left( 1 - \frac{\omega_v}{\omega_p} \right) / \frac{\omega_v}{\omega_p} \), \( \delta_1 = -\frac{\omega_v}{\omega_p} \left( 1 - \frac{\omega_v}{\omega_p} \right) / \frac{\omega_v}{\omega_p} \).

Then we proceed to look at how the transit operator responds differently to highway expansion with different economic objectives: (1) profit-maximization, (2) social welfare-maximization, (3) consumers’ surplus maximization with a budget constraint.

**Concluding remarks**

This paper investigates the system-wide impacts of highway expansion with responsive transit service when the two modes (auto and transit) are imperfect substitutes for system-wide elastic demand. The major findings are:

- The imperfect substitutability diverts the changes of the travel costs associated with
different alternatives; auto commuters are found to be more likely to benefit from the highway expansion compared to transit riders.

- The combined impact of highway expansion and transit response may lead to a new equilibrium where the travel costs of both modes increase (the so-called “Downs-Thomson paradox”).

- Improving transit service is not a necessary condition for the highway expansion to reduce travel costs of both modes, but it is necessary for raising transit ridership or reducing auto traffic.

- It is possible that improving service cannot attract more passengers; instead, traffic may be extracted from transit because of the highway expansion.

- The “Impact Factor of Transit Adjustment (IFTA)” is introduced to characterize the rate and direction at which the transit service is adjusted with the highway expansion.

- For given marginal demand and marginal cost, the IFTA is tightly related with the increasing rate of the marginal operational cost under profit/social welfare-maximizing transit policies; while a revenue-neutral transit operator is more concerned about the absolute value of the prevailing operational cost.

- The combined effect of highway expansion and transit response (i) reduces transit profit under profit-maximizing transit policy, (ii) enhances social welfare under social welfare-maximizing transit policy, and (iii) downgrades transit service under consumers’ surplus maximizing transit policy with a breakeven constraint.

- Highway expansion and responsive transit policies give rise to higher auto traffic and lower transit ridership under each of the economic regimes concerned.

Apart from refining our fundamental understanding of the two-mode transportation system, the results provide insights for the sustainable planning and policy-making on integrated transportation system.

**References**


Zhang, F., Yang, H., Liu, W., 2014. The Downs-Thomson Paradox with responsive transit service. Transportation Research Part A: Policy and Practice 70, 244-263.