First-best dynamic assignment of commuters with endogenous heterogeneities in a corridor network

Minoru Osawa\textsuperscript{a,}\textsuperscript{*}, Haoran Fu\textsuperscript{b,}\textsuperscript{*}, Takashi Akamatsu\textsuperscript{b}

\textsuperscript{a}School of Engineering, Tohoku University, Aramaki Aoba 6–6, Aoba-ku, Sendai, Miyagi, 980–8579, Japan
\textsuperscript{b}Graduate School of Information Sciences, Tohoku University, Aramaki Aoba 6–6, Aoba-ku, Sendai, Miyagi, 980–8579, Japan

\textbf{A R T I C L E  I N F O}

\textbf{Article history:}
Received 5 September 2017
Accepted 7 September 2017
Available online 20 October 2017

\textbf{Keywords:}
Departure time choice
Job-location choice
Corridor problem
User heterogeneity
Dynamic system optimal assignment

\textbf{A B S T R A C T}

We study a parsimonious theory that synthesizes short-term traffic demand management (TDM) policies with long-term endogenous heterogeneities of demand. In a corridor network with multiple discrete bottlenecks, we study a model of system optimal assignment that integrates the short-term problem (departure time choice with tolling) and the long-term problem (job and residential location choice). For the short-term departure-time-choice equilibrium, under mild assumptions on schedule delay function, we derive analytical solutions under a first-best TDM scheme. Investigating properties of long-term equilibria, we found that the overall equilibrium pattern exhibits remarkable spatio-temporal sorting properties. It is further shown that a lack of integration of the short- and the long-term policy results in excessive investments for long-term road construction.

© 2017 Elsevier Ltd. All rights reserved.

\section{1. Introduction}

\subsection{1.1. Background and purpose}

Alleviating traffic congestion in the morning peak hours is one of the most significant challenges in transportation science. Parsimonious and rich in insights, the standard departure-time-choice equilibrium model with a single bottleneck (\textit{Vickrey, 1969; Hendrickson and Kocur, 1981; Arnott et al., 1990}) and its extensions have long been the workhorse for tackling the problem, yielding a large body of theoretical studies (\textit{e.g. Smith, 1984; Daganzo, 1985; Newell, 1987; Lindsey, 2004; van den Berg and Verhoef, 2011; Liu and Nie, 2011}). Even though researches have successfully provided deep implications for the short-term TDM policies, when it comes to the long term, there is room for further investigation. In the long term, what is given in the short term is endogenous. Origin-destination demand patterns might vary over time because of residential location choice of commuters; commuters can also switch their jobs, thereby affecting value-of-time (VOT) distribution over commuters. Such long-term endogenous variations in demand inevitably affect long-term effectiveness of short-term TDM policies. Extending the short-term framework to incorporate long-term demand dynamics thus seems to be a valuable direction of study.

The purpose of the present paper is to introduce a parsimonious theory that synthesizes long-term endogenous heterogeneities of traffic demands with the short-term TDM policy. To this end, keeping the model as compact as possible, we extend the standard bottleneck model to allow each commuter to choose his/her own residential location and job in the
long-run, so that endogenous commuter heterogeneities emerge. Also, our analysis is conducted in a corridor network with multiple bottlenecks à la Akamatsu et al. (2015) to allow rich spatial and temporal dynamics.

We first prove that, under the short-term first-best policy, any integrated equilibrium for the whole problem is Pareto-efficient. We also show that there is an equivalent optimization formulation that is a highly structured linear programming problem, whose structure yields a natural decomposition into the two components of the short- and the long-term. We further show that the short-term component (departure time choice equilibrium) is analytically solvable under mild assumptions on the schedule delay function.\footnote{The result contrasts to previous studies which usually assume piecewise-linear schedule delay. This generalization becomes possible partly because we consider dynamic system optimal assignment. Nonetheless, it is noted that the basic strategy presented in this paper is effective and can be used to tackle dynamic user equilibrium assignment in a corridor network (Fu et al., 2016).}

Interestingly, equilibrium job–location–departure-time choice patterns are shown to exhibit striking sorting regularities in both the short- and the long-term. In the short-term equilibrium traffic flow, we observe a job- and location-based temporal sorting. Specifically, (i) commuters who work for jobs with higher VOT arrive at the CBD at times closer to their desired arrival time, and (ii) commuters who reside at distant locations have strictly longer arrival time window. In the long-term equilibrium patterns, we observe a job-based spatial sorting, where commuters who work for jobs with higher VOT choose to live closer to the CBD.

As a major policy implication, we show that, without integration of the short- and the long-term TDM policy, the investment for bottleneck capacity is overestimated; without consideration of the long-term decision of commuters, the road manager results in excessive extension of road capacity.

1.2. Related literature

To date, only Arnott (1998), Gubins and Verhoef (2014), and Takayama and Kuwahara (2017) have developed dynamic bottleneck models that include location choice of commuters. Our study extends the scopes of these studies in two ways. First, without sacrificing tractability, we generalize the single bottleneck setup to a corridor with multiple bottlenecks. The setting allows us to examine richer spatial dynamics in both the short and the long term. Second, we allow endogenous job choices that determine commuter heterogeneities (regarding their VOT), which is novel.

For the short-term component, our study formulates a dynamic system optimal (DSO) assignment which is achieved as the equilibrium under a short-term TDM scheme that eliminates the bottleneck queues. A DSO assignment problem with similar network setting (a freeway corridor with multiple discrete on- and off-ramps) is studied by Shen and Zhang (2009). Tian et al. (2012) is also in related direction, albeit the model is formulated via a continuum approximation à la Arnott and de Palma (2011). The prevalent problem is that analytical solutions are hard to obtain even when one assume piecewise linear schedule delay and homogeneous commuters. We contribute to this literature by providing analytical solutions (i.e., equilibrium traffic flow, equilibrium commuting cost) under mild assumptions on the schedule delay function, as well as allowing commuter heterogeneity.

2. Model

Consider a freeway corridor that connects \( I \) residential locations to a central business district (CBD), as illustrated by Fig. 1. The residential locations are indexed sequentially from the CBD. We denote the set of locations by \( I = \{1, 2, \ldots, I\} \). There is a single bottleneck with capacity \( \mu_i \) just downstream of the on-ramp from the location \( i \in I \), which we call the “bottleneck \( i \)”. Dynamic queueing at the bottlenecks are modeled by the standard point queue model along with first-in-first-out; at each bottleneck, a queue is formed vertically when the inflow exceeds the capacity. Each residential location \( i \in I \) is endowed with \( A_i \) unit of land. At the CBD, there are \( J \) job opportunities. The labor demand and the VOT for each job \( j \in J = \{1, 2, \ldots, J\} \) are given by \( L_j \) and \( \alpha_j \), respectively.

We assume that the tradable network permits (TNP) scheme is implemented in the corridor network.

Assumption 2.1. The TNP scheme (Akamatsu et al., 2006; Akamatsu, 2007; Wada and Akamatsu, 2013; Akamatsu and Wada, 2017) is implemented by the road manager as the short-term TDM policy.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{corridor_network.png}
\caption{The corridor network.}
\end{figure}
In the TNP scheme, a permit allows its holder to pass through a pre-specified bottleneck during a pre-specified time period, and a perfectly competitive trading market for the network permits is launched. The road manager limits the number of permits to be either equal to or less than the capacity of the bottleneck. Then, the queue at each bottleneck is completely eliminated so that there is no queuing delay. This scheme internalizes congestion externalities and is mathematically equivalent to an optimal dynamic road pricing scheme that completely eliminates bottleneck congestion; in effect, equilibrium values of the permits coincide with optimal congestion toll—without any knowledge of users’ preference. For details of the scheme including comparisons with other TDM policies (e.g., congestion tolling), see Wada and Akamatsu (2013) as well as Akamatsu and Wada (2017).

There are Q ex-ante identical workers in the network that commute to the CBD. Each commuter makes three choices to maximize his/her own utility: residential location \( i \in \mathcal{I} \) and job \( j \in \mathcal{J} \) in the long term and the arrival time \( t \in \mathcal{T} \) at the CBD in the short term, where \( \mathcal{T} \) denotes a sufficiently long arrival time window. Commuters’ utility is assumed to be homogeneous and quasilinear, i.e., the utility function is linear with respect to the numéraire.

At every morning, each commuter chooses his/her arrival time \( t \) at the CBD to minimize his/her own travel cost. The commuting cost of a worker who resides at location \( i \in \mathcal{I} \) with job \( j \in \mathcal{J} \) (henceforth “an \((i, j)\)-commuter”) is expressed in terms of arrival time \( t \in \mathcal{T} \) at the CBD:

\[
C_{i,j}(t) = c_{i,j}(t) + \sum_{m=1}^{d} p_m(t) \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T},
\]

where \( c_{i,j}(t) \) is the \((i, j, t)\)-specific monetary commuting cost, \( p_m(t) \) is the price of a permit at the bottleneck \( i \) with CBD arrival time \( t \). Note that the second term is the total amount of the permit costs a commuter from location \( i \) has to pay to arrive at \( t \) at the CBD. Also note that because we assume the TNP scheme, the queuing delay cost is absent.

We assume that \( c_{i,j}(t) \) is a sum of schedule delay cost and free-flow travel cost.

**Assumption 2.2.** \( c_{i,j}(t) \) is expressed as \( c_{i,j}(t) = \alpha_j s(t) + d_i \), where \( \alpha_j \) is the VOT for job \( j \), \( s(t) \) is the job-independent schedule delay function measured in time unit, and \( d_i \) is the free-flow travel time from residential location \( i \) to the CBD \((d_1 < d_2 < \cdots < d_Q)\).

Even though VOTs are differentiated with respect to job, all groups share exactly the same schedule delay preference. We will discuss relaxation of the apparently restrictive assumptions at the end of Section 4.

In the long term, in addition to their arrival time at the CBD, commuters can choose their jobs and locations to maximize their own utility. For simplicity, we further assume that land consumption of each consumer is unity. Then, the maximum number of commuters that can reside at location \( i \) coincide with the total land endowment \( A_i \). Let \( r_i \) be the market land rent at \( i \) and \( w_j \) the market wage of job \( j \) which are endogenously determined in the markets. We assume that the land and job markets are perfectly competitive. Then, an \((i, j)\)-commuter’s indirect utility is expressed as the following function of arrival time \( t \in \mathcal{T} \):

\[
v_{i,j}(t) = w_j - r_i - C_{i,j}(t) \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T}.
\]

Commuters choose \( i, j, \) and \( t \) to maximize their own utility: \( \max_{i, j, t} v_{i,j}(t) = \max_{i, j} \{ w_j - r_i - \min_{t \in \mathcal{T}} C_{i,j}(t) \} \).

Let \( q_{i,j}(t) \) be the arrival-flow rate at the CBD of \((i, j)\)-commuters at \( t \). At an equilibrium, there is no incentive for any commuter to switch his/her choice of \((i, j, t)\) unilaterally; that is, the following condition should hold true:

\[
\begin{align*}
V - v_{i,j}(t) &= 0 & \text{if } q_{i,j}(t) > 0 \\
V - v_{i,j}(t) \geq 0 & \text{if } q_{i,j}(t) = 0 \\
\forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \mathcal{T},
\end{align*}
\]

where \( V \) is the equilibrium utility level. Also, the following three demand–supply equilibrium condition at each market should be satisfied:

\[
\begin{align*}
\mu_i - \sum_{m=1}^{d} \sum_{j=1}^{Q} q_{m,j}(t) &= 0 & \text{if } p_m(t) > 0 \\
\mu_i - \sum_{m=1}^{d} \sum_{j=1}^{Q} q_{m,j}(t) &\geq 0 & \text{if } p_m(t) = 0 \\
\forall i \in \mathcal{I}, t \in \mathcal{T},
\end{align*}
\]

---

1. For the time index, we take the arrival time \( t \) at the CBD rather than the absolute departure time at each bottleneck to simplify the analysis. For discussions on the merits of using the arrival time \( t \) at the CBD (i.e., “time-based Lagrangian” coordinate system that focuses on each commuter) rather than departure time at the bottlenecks (i.e., “Eulerian” absolute coordinate system), see Akamatsu et al. (2015).
2. For the standard definition of quasilinear preference, see, for example, Mas-Colell et al. (1995) (Definition 3.B.7). In effect, the indirect utility function is expressed by (2).
3. Since we assume quasilinear preference, the cost minimization behavior is consistent with long-term utility maximization discussed below.
4. One may think that the formulation is too simplified because in reality the choice of \( t \) and those of \( i, j \) are the problems with very different time scales. Actually, in the following, we do decompose the integrated equilibrium problem into two interdependent components of the short- and the long-run and analyze them sequentially.
5. To keep the model compact, we assume that the wage is dependent solely on demand–supply condition and abstract from the effects from productivity of firms or other realistic determinants of the wage. The job–location choice modeling in our formulation corresponds to the celebrated Herbert–Stevens (H–S) model (Herbert and Stevens, 1960; Fujita, 1989) in the urban economics literature, as per revisited by Wheaton (1974). The H–S model has long provided rich insights into the internal structure of cities; applications include the seminal work of Fujita and Ogawa (1982).
\begin{align}
\text{(land market)} & \quad \begin{cases}
    A_i - \sum_j \int_T q_{i,j}(t) \, dt = 0 & \text{if } r_i > 0 \\
    A_i - \sum_j \int_T q_{i,j}(t) \, dt \geq 0 & \text{if } r_i = 0
\end{cases} \quad \forall i \in I, \\
\text{(labor market)} & \quad \begin{cases}
    L_j - \sum_i \int_T q_{i,j}(t) \, dt = 0 & \text{if } w_j > 0 \\
    L_j - \sum_i \int_T q_{i,j}(t) \, dt \leq 0 & \text{if } w_j = 0
\end{cases} \quad \forall j \in J.
\end{align}

Lastly, the number of commuters should not change:
\begin{equation}
\sum_{i \in I} \sum_{j \in J} \int_T q_{i,j}(t) \, dt = Q. \tag{7}
\end{equation}

An integrated equilibrium, which is a long-run equilibrium that respects short-run effects, is defined as follows:

**Definition 2.1.** An integrated equilibrium is a collection of variables \((V^*, q^* = \{q^*_{i,j}(t)\}), p^* = \{p^*_i(t)\}, r^* = \{r^*_i\}, w^* = \{w^*_j\})\) that satisfies (3), (4), (5), (6), and (7).\(^7\)

In the following, we decompose the model into two interdependent components of the short- and the long-term.

3. Equivalent optimization problems

First, by formulating an equivalent optimization problem (EOP) for our integrated equilibrium problem, the present section shows that any integrated equilibrium is efficient. Then, we decompose the EOP into two natural components: the short- and the long-term. According to the decomposition, the short- and the long-term equilibria are defined.

3.1. Efficiency of integrated equilibria

We expect that our equilibrium problem is equivalent to a socially optimal (or efficient) allocation problem since there are no externalities: the land and labor markets are perfectly competitive, whereas congestion externalities due to queueing at the bottlenecks are internalized by the short-term TDM policy (i.e., TNP scheme). In fact, we have the following proposition and its corollary, which confirm the intuition.

**Proposition 3.1.** An integrated equilibrium with any utility level \(V\) is obtained as a solution of the following linear programming problem (LP):
\begin{equation}
\begin{align}
\text{[Int]} \quad \min_{q \geq 0} & \quad z(q) = \sum_{i \in I} \sum_{j \in J} \int_T c_{i,j}(t) \, q_{i,j}(t) \, dt, \\
\text{s.t.} & \quad \sum_{m \in I} \int_T q_{m,j}(t) \, dt \leq \mu_i \quad (p_i(t)) \quad \forall i \in I, \quad t \in T, \\
& \quad \sum_{j \in J} \int_T q_{i,j}(t) \, dt \leq A_i \quad (r_i) \quad \forall i \in I, \\
& \quad \sum_{i \in I} \int_T q_{i,j}(t) \, dt \geq L_j \quad (w_j) \quad \forall j \in J,
\end{align}
\end{equation}
with \(\sum_i A_i = \sum_j L_j = Q\), where the price variables \(p, r,\) and \(w\) are determined as the Lagrange multipliers for the constraints (9), (10), and (11), respectively.

**Proof.** Comparing the first-order conditions with the equilibrium conditions, one shows the assertion. \(\square\)

**Corollary 3.1.** An integrated equilibrium (Definition 2.1) achieves a socially optimal (efficient) allocation in the sense that it minimizes the total social transportation cost in the network.

Note that, since at any equilibrium the total permit costs paid by the commuters coincides with the total revenue of the road manager from the TNP system, the permit costs cancels out if one consider the total social cost; one can actually regard \(z(q^*)\) as the total social cost incurred by an (integrated equilibrium) assignment \(q^*\). We thus conclude any integrated equilibrium under the TNP scheme is regarded as a social optimal assignment.

---

\(^7\) Observe that our integrated equilibrium is formulated as an equilibrium under the TNP scheme. If, instead, one assumes a top-down congestion-tolling scheme, the equilibrium conditions should include the queueing delays and their dynamics. The value of optimal tolling should be also determined through a mathematical programming with equilibrium constraints (MPEC).
3.2. Decomposition into the short- and the long-term components

We study the properties of integrated equilibria, i.e., those of solutions for the EOP [Int]. The EOP admits a natural decomposition into the short- and the long-term components. To see this, first define an intermediate variable $Q_{i,j}$ by

$$Q_{i,j} = \int_{T} q_{i,j}(t) \, dt \quad \forall i \in \mathcal{I}, j \in \mathcal{J}. \tag{12}$$

Then, $Q = \{Q_{i,j}\}$ is interpreted as a job–location choice pattern of commuters, which reflects the long-term decision of commuters. Observe that $z(q)$ is decomposed as $z(q) = z_1(q) + z_2(Q)$, where

$$z_1(q) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \int_{0}^{T} (\alpha_j s(t)) q_{i,j}(t) \, dt \quad \text{and} \quad z_2(Q) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} (\alpha_i d_i) Q_{i,j}. \tag{13}$$

Also, (10) and (11) can be rewritten by $Q$. In effect, [Int] is decomposed into the long- and short-term components.\(^8\)

3.2.1. Short-term problem

The short-term problem for a given $Q$ is formulated as the following LP:

$$\begin{align*}
\text{[ST–Primal]} \quad \min_{q \geq 0} & \quad z_1(q) \quad \text{s.t.} \quad (9) \text{ and } (12),
\end{align*}$$

where $Q$ constrains $q$ via (12). Its associated dual is

$$\begin{align*}
\text{[ST–Dual]} \quad \max_{\rho \geq 0} & \quad \bar{z}_1(\rho, p) = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \rho_{i,j} Q_{i,j} - \sum_{i \in \mathcal{I}} \int_{T} p_i(t) \, dt,
\end{align*}$$

s.t. \( \alpha_j s(t) + \sum_{m \in \Sigma} p_m(t) \geq \rho_{i,j} \quad (q_{i,j}(t)) \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in T. \tag{16} \)

where $\rho \equiv \{\rho_{i,j}\}$ and $p = \{p_i(t)\}$ are the Lagrange multipliers associated to the constraints (9) and (12), respectively. Via strong duality, it follows that the optimal values of [ST–Dual] and [ST–Primal] coincide:

$$\begin{align*}
\bar{z}_1(Q) = z_1(\rho^*, p^* | Q) = z_1(q^* | Q),
\end{align*}$$

where the superscript ”*” denotes that the associated variable is evaluated at the optimal solution; $\{\rho^*, p^*\}$ and $q^*$ are the dual and primal optimal points, respectively. It is also used to express equilibrium variables, as in Definition 2.1.

The first-order optimality conditions for [ST–(Primal / Dual)] are

$$\begin{align*}
\rho_{i,j} = \alpha_j s_i(t) + \sum_{m \in \Sigma} p_m(t) \quad \text{if} \quad q_{i,j}(t) > 0 \quad \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in T.
\end{align*}$$

along with (4) and (12). Observe that the condition (18) is interpreted as a condition for optimal arrival-time choice behavior of commuters under fixed $Q$ (i.e., a fixed job–location choice pattern) if we interpret $\{p_i(t)\}$ as permit prices. In this context, $\rho_{i,j}$ is regarded as the equilibrium commuting cost of $(i, j)$-commuters, whereas $\sum_{m \in \Sigma} p_m(t)$ is the total permit cost a commuter from the residential location $i$ has to pay to arrive at the CBD at time $t$. Conversely, we shall regard the problems [ST–(Primal / Dual)] as EOPs for the short-term (arrival-time-choice) equilibrium problem. We define a short-term equilibrium as follows.

**Definition 3.1** (Short-term equilibrium). A short-term equilibrium is a collection of variables $\{q^*, \rho^*, p^*\}$ that satisfies (4), (12), and (18).

Note that the right hand sides of (18) differs from $C_{i,j}(t)$ defined by (1) in a sense that the time-independent fixed commuting cost $\alpha_j d_i$ is absent. It is because $\alpha_j d_i$ is irrelevant for commuters’ choices if we fix the pattern of $Q$. Instead, $\{\alpha_i d_i\}$ appears in $z_2(Q)$ defined by (13), the objective function of the long-term component discussed below.

3.2.2. Long-term problem

Using the (dual) optimal value function $\bar{z}_1(Q)$, the long-term problem is formulated as follows, where we note that (20) and (21) are reformulations of (10) and (11) in terms of $Q$, respectively:

$$\begin{align*}
\text{[LT]} \quad \min_{Q \geq 0} & \quad \bar{z}_1(Q) + z_2(Q),
\end{align*}$$

s.t. \( \sum_{j \in \mathcal{J}} Q_{i,j} \leq A_i \quad (r_i) \quad \forall i \in \mathcal{I}, \tag{20} \)

\(8\) The decomposition corresponds to the so-called Benders decomposition.
Applying the envelope theorem to [ST–Dual], we see that \( \rho^\ast \) satisfies \( \nabla \tilde{z}_1(Q) = \rho^\ast(Q) \). Note that \( \rho^\ast_{*} \) is interpreted as a function of \( Q \) (i.e., the dual solution for [ST–Dual] for given \( Q \)). The first-order optimality conditions for [LT] are

\[
\begin{align*}
\sum_{i \in I} Q_{i,j} & \geq L_j & (w_j) & \forall j \in J. \\
\end{align*}
\]

(21)

and

\[
\begin{align*}
\begin{cases}
\begin{align*}
& w_j - r_i - \rho^\ast_{*}(Q) - \alpha_i d_i = 0 & \text{if } Q_{i,j} > 0 \\
& w_j - r_i - \rho^\ast_{*}(Q) - \alpha_i d_i \leq 0 & \text{if } Q_{i,j} = 0
\end{align*}
\end{cases} & \forall i \in I, j \in J.
\end{align*}
\]

(22)

Here, \( \{r_i\} \) and \( \{w_j\} \) are the Lagrange multipliers associated to (20) and (21), respectively.

As in [ST–(Primal / Dual)], we interpret (22) as the condition for optimal job–location choice behavior of commuters, where \( \{r_i\} \), \( \{w_j\} \), and \( \{\rho^\ast_{*}(Q) + \alpha_i d_i\} \) are each interpreted as land rent, wage, and commuting cost. The conditions (23) and (24) express market clearing. For instance, the condition (23) determines land rent \( r_i \) via demand–supply balance. It states that, if the total land demand \( \Sigma Q_{i,j} \) at a location \( i \in I \) is smaller than its land supply \( A_i \), i.e., if there is excess supply, \( r_i \) should be zero; if \( r_i \) is positive, then demand equals supply. In a similar manner, (24) determines wage \( w_j \). Analogous to the short-term equilibrium, we define a long-term equilibrium as follows.

**Definition 3.2 (Long-term equilibrium).** A long-term equilibrium is a collection of variables \( (Q^\ast, r^\ast, w^\ast) \) that satisfies (22), (23), and (24).

In [LT], the short-run (dual) optimal value function \( \tilde{z}_1(Q) \) represents the feedback effect from the short-run component. Even though \( \tilde{z}_1(Q) \) is, conceptually, a function of \( Q \), we do not yet have analytical formula of it. Thus, usually we cannot solve [LT] independently of [ST–(Primal / Dual)] and an iteration between [ST–(Primal / Dual)] and [LT] is required. Fortunately, Section 4 shows that [ST–(Primal / Dual)] is analytically solvable; i.e., the short-term flow pattern \( q^\ast(Q) \), as well as the associated Lagrange multipliers \( \rho^\ast(Q) \) and \( p^\ast(Q) \), are determined as functions of \( Q \). Using the result, Section 5 demonstrates that \( \tilde{z}_1(Q) \) is also analytically evaluated in terms of \( Q \).

4. Analysis of the short-term equilibrium

This section analyzes the properties of the problems [ST–(Primal / Dual)], which are equivalent to the short-term arrival-time-choice equilibrium under TNP (Definition 3.1), or to a dynamic system optimal assignment (in the sense of Corollary 3.1). We first review known results for two special cases of the problem: (i) the single-bottleneck case (i.e., \( I = 1 \)) (Akamatsu et al., 2016), and (ii) the homogeneous commuter case (i.e., \( J = 1 \)) (Yodoshi and Akamatsu, 2008; Fu, 2015; Fu et al., 2016). We show that analytical solutions for general setting (i.e., \( I > 1 \) and \( J > 1 \)) is naturally obtained by combining the properties of (i) and (ii).

To simplify the presentation throughout this section, we introduce some notations. First, we define \( t^- \), \( t^+ \), and \( \hat{s} \) based on the common schedule delay function \( s(t) \). \( t^- (T) \) is defined as the unique solution for the equation \( s(t^-) = s(t^+ + T) \) with \( t \leq t_0, T \geq 0 \) (see Fig. 2). Using \( t^- \), we define \( t^+ \) and \( \hat{s} \) by \( t^+(T) = t^- (T) + T \) and \( \hat{s}(T) = s(t^- (T)) = s(t^+(T)) \).

Next, we define \( \{\hat{a}_i\} \) and \( \{\hat{\mu}_i\} \) by \( \hat{a}_i \equiv \alpha_i - \alpha_{i+1} \) and \( \hat{\mu}_i \equiv \mu_i - \mu_{i+1} \), respectively, where \( \alpha_{J+1} \equiv 0, \mu_{I+1} \equiv 0 \).

In addition, \( t^-_{ij} \) and \( t^+_{ij} \) denote the earliest and latest arrival time of \((i, j)\)-commuters, respectively. We denote the arrival-time window \( \mathcal{T}_{ij} \) of \((i, j)\)-commuters and its length \( T_{ij} \) by \( \mathcal{T}_{ij} \equiv [t^-_{ij}, t^+_{ij}] \) and \( T_{ij} \equiv t^+_{ij} - t^-_{ij} \). Note that \( \mathcal{T}_{ij} \) is the convex hull of the support of \((i, j)\)-specific arrival flow. Finally, we denote \( \mathcal{T}_T \equiv [t_0, T] \).
4.1. Special case (i): Analytical solutions for a single bottleneck problem with heterogeneous commuters

The single-bottleneck case $(l = 1)$ is well-studied in the discipline; for this case, [ST–Primal] reduces to:

$$\text{[ST} - I_1]\quad \min_{q_i} \sum_{j \in J} \int_0^t \alpha_j s(t) q_{i,j}(t) \, dt,$$

\hspace{1cm} \text{s.t. } \sum_{j \in J} q_{i,j}(t) \leq \mu_1 \quad (p_1(t)) \quad \forall t \in T, \tag{25}

$$\int_0^t q_{i,j}(t) \, dt = L_j \quad (\rho_{i,j}) \quad \forall j \in J. \tag{26}$$

Note that, as we abstract from location choice decision of commuters in this case, we interpret $L_j = Q_{1,j}$.

Akamatsu et al. (2016) showed that the problem can be treated as an instance of mass transportation problem (Rachev and Rüschendorf, 1998). In particular, the coefficients $\{\alpha_j(s)\}$ is shown to satisfy Monge property.\(^\text{10}\) Monge property of $\{\alpha_j(s)\}$ yields that the equilibrium is unique, and commuters of higher VOT, $\alpha_j$ arrive at times closer to the unique desired arrival-time $t_0$ in equilibrium.\(^\text{11}\) In concrete terms, the arrival time windows exhibit the following job-based temporal sorting property:

**Lemma 4.1** (Job-based sorting properties in [ST–I1]). Consider an equilibrium solution for [ST–I1]. Under Assumption 2.2, the arrival-time windows $\{T_{1,j}\}$ satisfy $t_0 \subset T_{1,j} \subset T_{1,j+1}$ for all $j \in J \setminus \{\}$, or equivalently

$$t_{1,j} < \cdots < t_{1,2} < t_{1,1} < t_0 < t_{1,1}^+ < t_{1,2}^+ < \cdots < t_{1,j}^+. \tag{27}$$

The equilibrium values of $t_{1,j}^\pm$ are determined by $t_{1,j}^- = t^-(T_{1,j})$ and $t_{1,j}^+ = t^+(T_{1,j})$, where $T_{1,j} = \sum_{n \leq j} L_n / \mu_1$.

The above sorting regularity yields the following analytical solution for [ST–I1].\(^\text{12}\)

**Proposition 4.1.** The solution for [ST–I1] $(q^*, \rho^*, \mathbf{p}^*)$ is obtained as follows:

- arrival flow rates: $q_{i,j}^*(t) = \begin{cases} \mu_1 & \text{if } t \in T_{1,j} \setminus T_{1,j-1} \\ 0 & \text{otherwise} \end{cases} \quad \forall j \in J. \tag{29}$

- commuting cost: $\rho_{i,j}^* = \sum_{n \leq j} \alpha_n s(T_{1,n}) \quad \forall j \in J. \tag{30}$

- permit price: $p_i^*(t) = \begin{cases} \rho_{i,j}^* - \alpha_j s(t) & \text{if } t \in T_{1,j} \setminus T_{1,j-1} \\ 0 & \text{otherwise} \end{cases} \quad \forall j \in J. \tag{31}$

Fig. 3 illustrates Lemma 4.1 and Proposition 4.1 for the case of $J = 2$. Fig. 3a depicts equilibrium flow rates, which always coincide with the bottleneck capacity $\mu_1$ during the peak periods. Note that the length of arrival time windows are simply determined via exogenous constants $\{L_j\}$ by $T_{1,1} = |T_{1,1}| = L_1 / \mu_1$ and $T_{1,2} = |T_{1,2}| = (L_1 + L_2) / \mu_1$ (Lemma 4.1). From (29), it follows that the area of the red region equals $L_1$, whereas the total area of the blue regions $L_2$. Also note that $L_1 + L_2 = Q$. Using $\{T_{1,j}\}$, the actual values of $\{t_{1,j}^\pm\}$ is determined via the functions $t^\pm$ illustrated in Fig. 2.

On the other hand, Fig. 3b depicts permit price $p_i^*(t)$ at the bottleneck as a function of arrival time (in relation to equilibrium communting costs $\{\rho_{i,j}^*\}$). It is noted that the black curve is closely related to the so-called isocost curve (Cohen, 1987; Lindsey, 2004). Yet, there is a subtle but significant difference in that usual isocost curves express time-dependent queueing delay—instead of permit prices. Nonetheless, we shall call it the isocost curve (in terms of money), as the role it plays for our analysis is quite similar to usual ones.

Fig. 3b also graphically illustrates how to recursively construct the isocost curve (or graph of time-dependent permit prices). We first define the following time-varying “job-specific willingness to pay” for permit costs conditional on a given level of the commuting cost $\rho$:

$$p_j(t \mid \rho) = \rho - \alpha_j s(t). \quad \forall j \in J. \tag{32}$$

\(^9\) The EOP [ST–I1] is the one proposed by Iryo and Yoshii (2007). In light of this, our model [ST–Primal] is a corridor generalization of it.

\(^{10}\) An array $\{M_{i,j}\}$ is said to satisfy the (strict) Monge property if $M_{i,j} + M_{i,j+1} > M_{i,j+1} + M_{i+1,j}$ for all $i, j$. For reviews on Monge property and its applications, see for example Bein et al. (1995) and Burkard (2007).

\(^{11}\) Some previous studies also discuss temporal sorting properties in departure-time choice equilibrium problems with user heterogeneity (Newell, 1987; Arnott et al., 1992; 1994; Liu et al., 2015). In contrast to the results presented in this section, the previous studies assume piecewise linear schedule delay to obtain concrete results on sorting phenomena.

\(^{12}\) Proofs are provided in Appendix.
Then, in equilibrium it must be that \( p_1^*(t) = \max_j \{ P_j(t \mid \rho_{1,j}^*) \} \). As we already know the arrival time windows \( \mathcal{T}_{1,1} \) and \( \mathcal{T}_{1,2} \), we know that \( P_2(t_{1,2}^+ \mid \rho_{2,1}^*) = P_2(t_{1,2}^- \mid \rho_{2,1}^*) = 0 \). It yields that \( \rho_{2,1}^* = \alpha_2 s(T_{1,2}) \), so that the equilibrium curve \( P_2(t \mid \rho_{2,1}^*) \) in Fig. 3b is determined. Then, since \( t_{1,j}^+ \) are the crossing times of \( P_2(t \mid \rho_{2,1}^*) \) and \( P_1(t \mid \rho_{1,j}^*) \), the equilibrium cost \( \rho_{1,1}^* \) for job \( j = 1 \) is determined by the equation \( P_1(t_{1,1}^+ \mid \rho_{1,1}^*) = P_2(t_{1,1}^- \mid \rho_{2,1}^*) \). One can observe that no commuter can improve his/her commuting cost by changing the arrival time; e.g., no commuter with job \( j = 2 \) can afford to buy permit to arrive at \( t \in \mathcal{T}_{1,1} \) since \( p_1^*(t) = P_1(t \mid \rho_{1,1}^*) \geq P_2(t \mid \rho_{2,1}^*) \) for \( t \in \mathcal{T}_{1,1} \).

4.2. Special case (ii): Analytical solutions for a corridor problem with homogeneous commuters

Next, we consider another special case where commuters are identical with respect to their jobs (i.e., \( J = 1 \)) but the network structure is the original corridor in Fig. 1. This homogeneous case also admits analytical solutions and exhibits another temporal sorting regularities in equilibrium ([Fu, 2015; Fu et al., 2016]).

We set \( J = 1, \) \( \alpha_1 = 1 \) and \([ST–Primal]\) reduces to the following problem:

\[
[ST - J_1] \quad \min_{\theta \in \Theta} \sum_{i \in \mathcal{I}} \int_{\mathcal{T}} s(t) q_i(t) \, dt, \tag{33}
\]

\[
\text{s.t.} \quad \sum_{m \in \mathcal{J}} q_m(t) \leq \mu_i \quad (p_i(t)) \quad \forall i \in \mathcal{I}, \quad t \in \mathcal{T}, \tag{34}
\]

\[
\text{s.t.} \quad \int_{\mathcal{T}} q_i(t) \, dt = A_i \quad (\rho_i) \quad \forall i \in \mathcal{I}, \tag{35}
\]

where we omit job subscript \( j = 1 \) and denote \( q_i(t) = q_{1,i}(t), \rho_i = \rho_{1,i} \) for expositional simplicity. Note that, as we abstract from job choice of commuters, we interpret \( A_i = Q_{i,1} \) in this case.

First, two concepts of false bottleneck and reduced network are introduced to simplify the presentation.

**Definition 4.1** (False bottleneck). A false bottleneck is a bottleneck \( i \) whose permit price is always zero in equilibrium.

**Definition 4.2** (Reduced network). A reduced network is a network with no false bottlenecks.

In a corridor network, a false bottleneck does not affect essential properties of the solution. In equilibrium, passing such a bottleneck at any time does not affect commuting cost for any commuter since permit price always equals zero. It causes some irrelevant freeness in analytical solution. To avoid such unnecessary complications in presentation, we shall consider a reduced corridor network. Fortunately, we have a simple criterion to detect false bottlenecks. First, define the normalized demand \( D_i \), or the demand/supply ratio, at each bottleneck \( i \), by \( D_i = \Sigma_{m \geq i} |\rho_m|/\mu_i \). Note that the numerator of \( D_i \) is the total upstream demand which is to pass through the bottleneck \( i \). Using the normalized demand \( D_i \), we have the following lemma that provides the means to detect a false bottleneck.

**Lemma 4.2** (Detection of a false bottleneck). For the problem \([ST–J_1]\), a bottleneck \( i \in \mathcal{I} \) is a false bottleneck if and only if \( D_m \geq D_i \) for some \( m < i \).

From any setting of \( \{A_i\} \) and \( \{\mu_i\} \), one can obtain a reduced corridor network by sequentially removing false bottlenecks. We thus justify proceeding with a reduced corridor network assumption.

The following lemma and proposition summarizes the analytical results for \([ST–J_1]\) in the reduced corridor network. First, arrival time windows \( \mathcal{T}_i = \mathcal{T}_{i,1} \) satisfy the following location-based temporal sorting property.
Lemma 4.3 [Location-based sorting properties in [ST–J1]]. Consider a equilibrium solution for [ST–J1]. Under Assumption 2.2, the arrival-time windows \(\{T_i\}\) satisfy \(\{t_0\} \subset T_i \subset T_{i+1}\) for all \(i \in \mathcal{I} \setminus \{l\}\), or equivalently
\[
t_0 < \cdots < t_2 < t_1 < t_0 < t_1^+ < t_2^+ < \cdots < t_l^+.
\] (36)
The equilibrium values of \(t_i^\pm\) are determined according to \(T_i \equiv |T_i|\) as \(t_i^- = t^- (T_i)\) and \(t_i^+ = t^+ (T_i)\), where \(T_i = A_i / \mu_i\).

In other words, commuters with a smaller location index always arrive at the CBD nearer to the desired arrival time.

On the basis of Lemma 4.3, we obtain the following proposition which summarize the equilibrium solution in a reduced corridor network with \(J = 1\).

Proposition 4.2. In a reduced corridor network, the solution for [ST–J1] \((q^*, \rho^*, P^*)\) is obtained as:

- arrival flow rates: \(q_i^* (t) = \begin{cases} \mu_i & \text{if } t \in T_i, \\ 0 & \text{otherwise} \end{cases}\) \(\forall i \in \mathcal{I}\) (37)
- commuting cost: \(\rho^*_i = \tilde{s} (T_i)\) \(\forall i \in \mathcal{I}\) (38)
- permit price: \(p_i^* (t) = \begin{cases} \rho_i^* - s(t) & \text{if } t \in T_i \setminus T_{i-1}, \\ \rho_i^* - s(t) - \sum_{m<i} P_m^* (t) & \text{if } t \in T_{i-1}, \\ 0 & \text{otherwise} \end{cases}\) \(\forall i \in \mathcal{I}\). (39)

Fig. 4 illustrates the analytical solution in this proposition for the case of \(l = 3\). Fig. 4a depicts equilibrium arrival flow rates at the CBD. One can reconstruct disaggregated, i-specific arrival flow rates \(q_i^* (t)\) from the figure since \(q_i^* (t) = \mu_i\) for all \(t \in T_i\) and zero otherwise. Because \(T_i = A_i / \mu_i\), the areas of the gray regions equal \(\{A_i\}\), where the darker color corresponds to that of the closer residential location.

Fig. 4b depicts the time-varying permit cost \(\sum_{m<i} P_m^* (t)\) that a commuter from residential location \(i\) has to pay to arrive the CBD at \(t \in T\) (in relation to \(\{\rho_i\}^\star\)). The top curve corresponds to \(i = 3\), the middle \(i = 2\), and the bottom \(i = 1\). As in [ST–J1], the curves can be interpreted as isocost curves for commuters from each residential location. At each location, the isocost curve is the (homogeneous) willingness-to-pay function \(P(t|\rho_i)\) of its residents, subject to a given level \(\rho_i\) of commuting cost at location \(i\):

\[P(t | \rho_i) = \rho_i - s(t), \quad \forall i \in \mathcal{I}\] (40)

It is interesting to compare (40) with (32), where commuters are heterogeneous. In equilibrium, it must be that

\[P(t | \rho_i^*) = \rho_i^* - s(t) = \sum_{m<i} P_m^* (t) \quad \forall t \in T_i, \ i \in \mathcal{I},\] (41)

which can be readily verified from Fig. 4b.

The false-bottleneck condition in Lemma 4.2 is understood using Fig. 4b. For instance, if \(D_1 > D_2\), it follows that \(T_1 > T_2\). It in turn implies that \(\rho_1^* > \rho_2^*\), but this contradicts to the fact that \(p_i^* (t) \geq 0\) for all \(t \in T_i\); when \(D_1 = D_2\), it implies \(\rho_1^* = \rho_2^*\); but then the bottleneck 2 becomes irrelevant for analyzing the properties of the equilibrium.

4.3. Analytical solutions for the general case

Equipped with intuitions from the two special cases, we turn our attention to the general case of [ST–(Primal / Dual)]. In essence, the general case is understood as a combination of the two special cases presented in the above.
For a given pattern of $Q$, we introduce the location-specific job set $J_i$, which is the set of jobs that have employees residing at location $i \in I$, by $J_i = \{ j \in J | Q_{ij} > 0 \}$ for all $i \in I$. Its maximal and minimal elements are defined by $j_-(i) \equiv \min J_i$ and $j_+(i) \equiv \max J_i$.

Analogous to [ST–J1], we consider a reduced network for the general setting to simplify the presentation. In doing so, we define a false bottleneck for the problem [ST–(Primal / Dual)] as follows:

**Definition 4.3** (False bottleneck in [ST–(Primal / Dual)]). A false bottleneck in [ST–(Primal / Dual)] is a bottleneck $i$ such that $p_i^*(t) = 0$ for all $t \in \text{supp}(q_i(t))$ for some job index $j_\kappa \in J_i$ in equilibrium.

Suppose that a bottleneck $i$ is a false bottleneck in the sense of Definition 4.3 and that $j = j_\kappa$. Then, any $(i, j)$-commuter can pass through the bottleneck for free of charge, thereby causing unnecessary degeneracy in the equilibrium flow rates; we will discuss this point later using Fig. 6. To screen out false bottlenecks, we define the normalized demand for the problem [ST–(Primal / Dual)] by $D_{ij} = \sum_{k \geq 1} \hat{\alpha}_{ij} \sum_{n \leq 1} Q_{k,n,i}/\mu_i$, where $\hat{\alpha}_{ij} = \hat{\alpha}_j$ for $j \in J_i \setminus \{ j_\kappa (i) \}$ and $\hat{\alpha}_{ij_\kappa} = \alpha_{j_\kappa, i}$. Analogous to Lemma 4.2, we conclude as follows:

**Lemma 4.4** (Detection of a false bottleneck in [ST–(Primal / Dual)]). For the problem [ST–(Primal / Dual)], a bottleneck $i \in I$ is a false bottleneck if and only if $D_{mj} \geq D_i$ for some $m < i$ and some $j \in J_i$.

We can always construct a reduced corridor using Lemma 4.4; we thus assume a reduced corridor network in the following presentation: $D_{ij} < D_{i+1,j}$ for all $i \in I \setminus \{ 0 \}$ and $j \in J_i$.

We first show that, at each bottleneck in the corridor network, job-based temporal sorting property is present as it is in the single bottleneck case [ST–I1] (Lemma 4.1). Assuming a reduced corridor, we show that the arrival time windows $T_{ij}$ exhibit the following sorting property.

**Lemma 4.5** (Job–location-based sorting property in [ST–(Primal / Dual)]). Consider an equilibrium solution for [ST–(Primal / Dual)]. Under Assumption 2.2, the arrival-time windows $\{T_{ij}\}$ satisfy $t_0 \subset T_{ij} \subset T_{ij+1}$ for all $j \in J_i \setminus \{ j_\kappa (i) \}$ and for all $i \in I$, or equivalently

$$
t_{ij+1}^N < t_{ij}^N < t_0 < t_{ij} < t_{ij+1}^N \quad \forall j \in J_i \setminus \{ j_\kappa (i) \}, i \in I
$$

*The equilibrium values of $\{t_{ij}^N\}$ are determined according to $T_{ij} = t^-(T_{ij})$ and $t_{ij+1}^N = t^+(T_{ij})$, where $T_{ij}$ is obtained as a function of $Q_i \equiv \{Q_{ij} | j \in J\}$ such that $T_{ij}(Q_i) = \sum_{n \leq 1} Q_{i,n}/\mu_i$.

Observe that the length of the arrival time window $T_{ij}(Q_i)$ is interpreted as a natural generalization of its counterpart for [ST–I1] and [ST–J1]; if we set $I = 1$ or $J = 1$, it reduces to $T_{ij}$ in Lemma 4.1 or $T_i$ in Lemma 4.3, respectively. It is also noted that for $j > j_\kappa (i)$ and $j > j_\kappa (i)$, we define $T_{ij} = 0$ and $T_{ij} = T_{ij}(Q_i)$, respectively.

The following proposition summarizes the analytical solution for [ST–(Primal / Dual)]. Note that $T_i = \bigcup_j T_{ij}$.

**Proposition 4.3**. In a reduced corridor network, the solution for [ST–(Primal / Dual)] $(q^*, \rho^*, p^*)$ is obtained as:

- **Arrival flow rates:**
  $$q^*_{ij}(t) = \begin{cases} \hat{\mu}_i & \text{if } t \in T_{ij} \setminus T_{ij-1} \\ 0 & \text{otherwise} \end{cases} \quad \forall j \in J_i, i \in I$$

- **Commuting cost:**
  $$\rho^*_{ij}(Q_i) = \sum_{n \geq 1} \hat{\alpha}_{i,n} s(T_{in}(Q_i)) \quad \forall j \in J_i, i \in I$$

- **Permit price:**
  $$p^*_{ij}(t) = \begin{cases} \rho^*_{ij} - \alpha_j s(t) - \sum_{m \leq 1} p^*_m(t) & \text{if } t \in (T_{ij} \setminus T_{ij-1}) \setminus T_{ij-1} \\ \rho^*_{ij} - \alpha_j s(t) & \text{if } t \in (T_{ij} \setminus T_{ij-1}) \cap T_{ij-1} \\ 0 & \text{otherwise} \end{cases} \quad \forall j \in J_i, i \in I$$

Fig. 5 illustrates Lemma 4.5 and Proposition 4.3 for the case of $I = 2, J = 2$. The figure can be understood as a combination of Fig. 3 and Fig. 4. Note that for simplicity we assume $Q > 0$ here, so that $J_1 = J_2 = J = \{ 1, 2 \}$. In Fig. 5, from Lemma 4.5, we see that

$$T_{1,1} = |T_{1,1}| = Q_{1,1}/\hat{\mu}_1, \quad T_{1,2} = |T_{1,2}| = (Q_{1,1} + Q_{1,2})/\hat{\mu}_1 = A_1/\hat{\mu}_1$$

$$T_{2,1} = |T_{2,1}| = Q_{2,1}/\hat{\mu}_2, \quad T_{2,2} = |T_{2,2}| = (Q_{1,1} + Q_{2,2})/\hat{\mu}_2 = A_2/\hat{\mu}_2.$$
Fig. 5b illustrates equilibrium isocost curves for [ST–(Primal / Dual)] in relation to equilibrium permit prices and commuting costs. We immediately notice that Fig. 5b is a combination of Figs. 3b and 4b. The isocost curve at each bottleneck is constructed as in [ST–I]. Specifically, the isocost curve at the bottleneck $i$ is

$$\sum_{m \leq i} p_m^*(t) = \max_j \{ P_j(t \mid \rho_{i,j}) \} \quad \forall t \in T_i, \ j \in J_i, \ i \in I,$$

(48)

where $P_j(t \mid \rho_{i,j}) = \rho_{i,j} - \alpha_j s(t)$ is the $(i, j)$-specific willingness-to-pay for permit costs; special cases are given by (32) and (40). Then, as in [ST–I], the isocost curves are vertically sorted according to the location index, as long as we consider a reduced corridor network. Observe that the isocost curve at $i \in I$ depends only on $Q_i = \{ Q_{i,j} \mid j \in J_i \}$.

Fig. 6 illustrates a situation where a false bottleneck in Definition 4.3 appear. In the figure, the bottleneck $i = 2$ is a false bottleneck, where $p_2^*(t) = 0$ for $T_{2,1}$. Under such situation, the equilibrium arrival flow rates for commuter with job $1$ (i.e., $q_{1,1}^*(t)$ and $q_{2,1}^*(t)$) are not uniquely determined in $T_{2,1}$, since equilibrium costs are the same for both location 1 and 2 during the time window. In contrast to the homogeneous case [ST–I], however, such a situation does not necessarily mean that the equilibrium costs are not affected by the existence of the bottleneck. Even though the bottleneck 2 is irrelevant for commuters with job 1, it does affect commuters with job 2 since $p_2^*(t) > 0$ for $T_{2,2} \setminus T_{2,1}$. As Fig. 6 illustrates, a false bottleneck appears as a result of some "overlaps" between isocost curves for consecutive bottlenecks; Lemma 4.4 allows one to detect and sort out such situations.

Since our analysis is conducted under specific assumptions on $c_{i,j}(t)$, it would be beneficial to discuss robustness of the obtained short-run results against relaxations of our simplifying assumptions (made in Section 2) before proceeding to the analysis of the long-run equilibria. The most important point is that $s(t)$ is job-independent. It says that all groups share exactly the same schedule delay preference. Though seemingly restrictive, we can generalize our analytical results to “single-crossing” cases, provided that the desired arrival time is unique. Consider a case where $s(t)$ is differentiated by job as $s_j(t)$. With job-dependent schedule delay functions $s_j(t)$, define the willingness-to-pay functions by $P_j(t \mid \rho_j) = \rho_j - s_j(t)$.

A single-crossing case is where for any given pair of jobs $j, k \in J$, for any given levels of $\rho_j$ and $\rho_k$, the equation $P_j(t \mid \rho_j) = P_k(t \mid \rho_k)$ has at most a single solution in each of $t < t_0$ and $t > t_0$. It is noted that the single-crossing property is actually general, as majority of previous studies were conducted under the property. To simplify the presentation, however, we employ job-independent $s(t)$. 
5. Analysis of the long-term equilibrium and comparison with a “naïve” long-term equilibrium model

Having established the analytical formula (44) of the short-term equilibrium commuting cost function \( \rho^*(Q) \), we are now in a position to analyze the properties of the long-term problem.

5.1. Properties of the long-term equilibrium

5.1.1. Qualitative property

The properties of \( \rho^*(Q) \), which in turn determine those of long-term equilibria, are summarized as follows:\(^{13}\)

**Lemma 5.1** (Properties of \( \rho^*(Q) \)). The commuting cost function \( \rho^*(Q) \) satisfy the following properties:

(a) \( \rho^*(Q) \) is separable with respect to location. That is, \( \rho_i^* \equiv \{ \rho_j^* \mid j \in J \} \) depends only on \( Q_i \), \( \{Q_i \mid j \in J \} \).

(b1) Each \( \rho_i^*(Q_i) \) is strictly monotone.

(b2) Each \( \rho_i^*(Q_i) \) is differentiable and \( \nabla \rho_i^*(Q_i) \) is positive, symmetric, and positive definite, if \( s(t) \) is differentiable.

(c1) \( \rho_i^*(Q) \) is monotonically increasing in \( i \).

(c2) \( \rho_i^*(Q) \) is monotonically decreasing in \( j \).

Combining the definition of the long-term equilibrium (Definition 3.2) and Lemma 5.1 (b2), we show:

**Proposition 5.1.** The long-term equilibrium is obtained as a solution for the following convex optimization problem.

\[
\begin{align*}
\text{[LT']} & \quad \min_{Q \geq 0} z_1(Q) + z_2(Q) \quad \text{s.t.} \quad (20) \text{ and } (21),
\end{align*}
\]

where

\[
z_1(Q) = \sum_{i \in \mathcal{I}} \int_{0}^{\rho_i^*(Q)} \rho_i^*(\omega) \, d\omega = \sum_{i \in \mathcal{I}} \left( \rho_i^*(Q_i) Q_i - \mu_i \int_{T} p_i^* (t | Q_i) \, dt \right) = \tilde{z}_1(Q)
\]

Since strict monotonicity of \( \rho^*(Q) \) (Lemma 5.1 (b1)) implies strict convexity of \( \tilde{z}_1(Q) \) (and hence \( z_1(Q) + z_2(Q) \)), we have the following corollary regarding uniqueness of the equilibrium.

**Corollary 5.1.** The solution for [LT'], or the long-term equilibrium, is unique.

The above problem can also be regarded as a part of integrated equilibrium problem which includes feedbacks from the short-term component. As the short-term equilibrium is uniquely determined in a reduced corridor, we can sequentially solve [LT'] and [ST–Primal] to obtain integrated equilibrium pattern (Definition 2.1).

5.1.2. Equivalence and graphical interpretations

As [LT'] in the above and [LT] in Section 3 should be equivalent to each other by construction, it must be that \( \tilde{z}_1(Q) = \tilde{z}_1(Q) \). In other words, recalling the definition of \( \tilde{z}_1(Q) \) in (17), we must have the following identity in short-term equilibrium for any given values of \( Q_i \):}

\[
\tilde{z}_1(Q) = \sum_{i \in \mathcal{I}} \int_{0}^{\rho_i^*(Q)} \rho_i^*(\omega) \, d\omega = \sum_{i \in \mathcal{I}} \left( \rho_i^*(Q_i) Q_i - \mu_i \int_{T} p_i^* (t | Q_i) \, dt \right) = \tilde{z}_1(Q)
\]

where \( \{ p^*(t|Q) \} \) is permit price at the short-term equilibrium. Actually, we conclude as follows.

**Lemma 5.2.** \( \tilde{z}_1(Q) = \tilde{z}_1(Q) \) holds true. Furthermore, the value is explicitly computed as

\[
\tilde{z}_1(Q) = \sum_{i \in \mathcal{I}} \int_{0}^{\rho_i^*(Q)} \rho_i^*(\omega) \, d\omega = \sum_{i \in \mathcal{I}} \left( \rho_i^*(Q_i) Q_i - \mu_i \int_{T} p_i^* (t | Q_i) \, dt \right) = \tilde{z}_1(Q)
\]

Observe that \( \tilde{z}_1(Q) \) is a weighted sum of integrals of job-specific schedule delay functions.

The last equality of (51) has intuitive graphical interpretation; recalling that \( Q_{i,j} = \tilde{\mu_i} (T_{i,j} - T_{i,j-1}) \) in equilibrium (Proposition 4.3), (51) implies the following relation which is satisfied at each bottleneck \( i \in \mathcal{I} \):

\[
\tilde{\mu_i} \sum_{j \in J} \tilde{c}_{i,j} - \tilde{\mu_i} \tilde{R}_i = \tilde{\mu_i} \sum_{j \in J} \tilde{S}_{i,j} \quad \text{where} \quad \tilde{c}_{i,j} = \rho_{i,j}^* (T_{i,j} - T_{i,j-1}) \text{ and } \tilde{R}_i = \int_{j \in J} \sum_{m \in J} p_m (t) \, dt.
\]

Notice that \( \tilde{\mu_i} \tilde{R}_i \) is the total permit cost paid by commuters residing at residential location \( i \). The above equality thus states that, in equilibrium, the total commuting cost net of the total permit cost coincides with the total schedule delay cost. In other words, it represents the strong duality between [ST–Primal] and [ST–Dual]. The relation is schematically illustrated by Fig. 7. The isocost curve approach in the literature, which determine the boundary in Fig. 7, is thus interpreted to have its basis on the primal–dual relation of the short-term equivalent optimization problem.

\(^{13}\)Lemma 5.1 can also be generalized to single-crossing cases. In effect, the results of Section 5 are robust against relaxation of assumptions on schedule delay preference.
5.1.3. Properties of equilibrium job–location choice pattern

Solving [LT], we obtain the equilibrium job–location choice pattern $Q^* = \{Q^*_j\}$, or travel demand for the short-term problem. Analogous to $J_j$, we define job-specific location sets by $I_j = \{i \in I \mid Q_{i,j} > 0\}$ for all $j \in J$. We show the following regularity in equilibrium job–location choice pattern $Q^*$:

**Proposition 5.2.** $j_-(i) = \min J_i$, $j_+(i) = \max J_i$. $l_-(j) = \min I_j$, and $l_+(j) = \max I_j$ satisfy

$$\begin{align*}
\{j_-(i) \leq j-(i+1), j_+(i) \leq j_+(i+1) & \} \quad \forall i \in I \setminus \{l\} \quad \text{and} \quad \{l_-(j) \leq l_-(j+1), l_+(j) \leq l_+(j+1) & \} \quad \forall j \in J \setminus \{j\}. 
\end{align*}$$

(54)

Proposition 5.2 states that commuters with larger job index (i.e., lower VOT $\alpha_j$) are located at residential locations distant from the CBD, and vice versa. Or, we have a spatial sorting property in $Q^*$. An example of equilibrium pattern $Q^*$ is illustrated by Fig. 8a, in which one can observe both of the sorting property (54).

5.1.4. Properties of equilibrium wage and land rent

In the long-term equilibrium, by (22), the following equality should hold true in the support of $Q^*$:

$$w_j^* - r_i^* - \rho^*_j(Q_{i}) - \alpha_j d_i = 0 \quad \forall j \in J_i, \ i \in I. \quad (55)$$

Recall that $\rho_{i,j}(Q_{i})$ is monotonically decreasing in the job index $j$ and monotonically increasing in the location index $i$. Then, (55) yields the following proposition.

**Proposition 5.3.** $w_j^*$ is monotonically decreasing in $j$ and $r_i^*$ is monotonically decreasing in $i$.

Without loss of generality, suppose that $\hat{\alpha}_j$ is a $j$-independent constant. We can interpret this case as the continuous limit of the job set where there are large number of jobs. In this case, $\{l\}$ should be understood as the probability distribution of VOT $\{\alpha\}$. We show the following proposition.

**Proposition 5.4.** Assume that $\hat{\alpha}_j$ is a $j$-independent constant. Then, $\{w_j^*\}$ is strictly concave on $J_j$ at every residential location $i \in I$, in the sense that the second-order difference is negative. Overall, $\{w_j^*\}$ is piecewise concave in $j \in J$.

5.2. The “naïve” model and its analytical solutions

To investigate the role of interactions between the long- and the short-term components, we consider a special case in which we completely ignore interactions between the two time scales. We call it the “naïve” model.

---

14 For Figs. 8 and 9, exogenous parameters and functions are set as follows: $l = 10, j = 20, A_l = 18 + 4i$ and $\mu_i = 11 - i$ for all $i \in I$. $t_j = 20$ and $\alpha_j = 2.1 - 0.1j$ for all $j \in J$. and $s(t) = \max[-0.2(t - t_0), 0.4(t - t_0)]$ for all $t \in T$. 

Definition 5.1 (Naïve model). The long-term component of the naïve model is defined by

\[ \text{[LT–N]} \min_{Q \geq 0} z_2(Q) \quad \text{s.t.} \quad (20) \text{ and } (21), \]  

along with \( \sum_{i \in J} A_i = \sum_{j \in J} L_j = Q. \) In the model, we first solve [LT–N] to obtain the long-term solution \( Q^N. \) Then, we solve [ST–Primal] to determine the short-term dynamic equilibrium flow pattern.

The long-term component of the naïve model, [LT–N], is obtained by ignoring the feedback from the short-term component, i.e., the first term \( z_1(Q) \) of the problem [LT']. Even though we sequentially solve the long-term and the short-term problems also in the naïve model, there is no consideration of the short-term effect in the first step; even though both components each seek to achieve efficient allocation, the two components are completely separated in the naïve model. We note that the naïve model is equivalent to the Herbert–Stevens model (as revisited by Wheaton, 1974), which is an important model in theoretical literature in urban land use forecasting.

The problem [LT–N] admits a unique analytical solution \( Q^N, \) thanks to Monge property of the coefficient array \( (a_i d_i) \) of \( z_2(Q). \) It is noted that the analytical solution \( Q^N \) corresponds to the so-called northwest corner rule.

Lemma 5.3 (Analytical solution for [LT–N]). Define the cumulative distributions for \( (A_i), (L_i), \) and \( (Q^N_{ij}) \) by \( \hat{A}_i = \sum_{m \leq i} A_m, \hat{L}_j = \sum_{n \leq j} L_n, \) and \( \hat{Q}^N_{ij} = \sum_{k \leq i} \sum_{l \leq j} Q^N_{kl}, \) respectively. Then, \( \hat{Q}^N_{ij} = \min \{ \hat{A}_i, \hat{L}_j \}. \)

Fig. 8b illustrates the analytical solution \( Q^N \) in marginal form. Observe that the commuters with higher VOT are prioritized and sequentially assigned to residential locations closer to the CBD. Equilibrium land rent and wage are shown to satisfy similar properties as Proposition 5.3.

5.3. Comparisons and policy implications

5.3.1. Equilibrium job–location choice patterns

Define location-specific job set \( J^N_i \) and job-specific location set \( I^N_j \) for the naïve model by \( J^N_i = \{ j \in J \mid Q^N_{ij} > 0 \} \) and \( I^N_j = \{ i \in I \mid Q^N_{ij} > 0 \} \), respectively. Then, we have the following regularity:

Proposition 5.5. The location-specific job set for the naïve model is contained by that of integrated model: \( J^N_i \subset J_i \) for all \( i \in I. \) Similarly, \( I^N_j \subset I_j \) holds true.

Fig. 8 illustrates the proposition. Even though the total number of commuters at location \( i \) always equals to \( A_i, \) the integrated equilibrium has a greater job variety at every location than the equilibrium in the naïve model. In other words, the extra term \( z_1(Q) \) acts as—due to its strict convexity—a “regularization” which prefers interior solutions.

5.3.2. Equilibrium land rents and wages

In equilibrium of [LT–N], we have \( w^N_j - r^N_i - \alpha d_i = 0 \) for all \( i \in I, j \in J^N, \) where \( \{ w^N_j \} \) and \( \{ r^N_i \} \) are the equilibrium wage and land rent of the naïve model, respectively. Compared with (55), this condition misses the short-term equilibrium cost \( \rho^*_i(Q). \) It leads to some “bias” in the price variables in the naïve model:

Proposition 5.6. Let \( r^N_i = r^*_i = 0 \) for normalization. Then, \( r^N_i < r^*_i \) for all \( i \in I \setminus \{ l \} \) and \( w^N_j < w^*_j \) for all \( j \in J. \)

That is, equilibrium land rent and wage are underestimated in the naïve model. Fig. 9 illustrates the above Proposition 5.6, as well as Propositions 5.3 and 5.4. First, Fig. 9a illustrates the land rent curves. One readily observe the

---

15 Since \( \{ a_j \} \) is decreasing in \( j \) and \( \{ d_i \} \) is increasing in \( i \), one can easily verify that \( \alpha d_i + \alpha_{i+1} d_{i+1} > \alpha_{i+1} d_{i+1} + \alpha_{i+2} d_{i+2}. \)

16 In marginal form, we have \( J^N_i = \{ j \in J \mid Q^N_{ij} > 0 \} = \{ j \in J \mid \hat{A}_{i+1} < \hat{L}_j < \hat{A}_i \} \) and \( Q^N_{ij} = \min(\hat{A}_i, \hat{L}_j) - \max(\hat{A}_{i+1}, \hat{L}_{j+1}) \) for all \( i \in I, j \in J^N. \)
relation $r^N_i < r^*_i$ except for $i = l$. By Henry George Theorem, the optimal land investment at residential location $i$ coincides with the total land rent at $i$. Recall that the total number of commuters at location $i$ is fixed to $A_i$, so that the land investment also follows a similar inequality $(r^N_i A_i < r^*_i A_i)$. The proposition thus implies that if we employ $[LT^*-N]$ to determine the long-term land investment policy, it results in a insufficient development of residential areas.

On the other hand, Fig. 9b compares wages. Observe that $w^N_j < w^*_j$. Analogous to land development, if one consider industry zoning at the CBD, then the total wage of each job $j$ is the measure to determine which industry to lure. The relation $w^N_j L_j < w^*_j L_j$ then implies that high VOT industries are undervalued in the naïve model.

5.3.3. **Comparison of the total revenues of the road manager**

The total equilibrium revenue $R(Q)$ of the road manager from the TNP scheme is of interest for considering long-term road capacity investments. $R(Q)$ is the total of the equilibrium permit revenues $R_i(Q)$ at every bottleneck $i \in I$:

$$R(Q) = \sum_{i \in I} R_i(Q) \quad \text{where} \quad R_i(Q) = \mu_i \int_{T_i} p_i(t) \, dt.$$  \hspace{1cm} (57)

The analytical solution $Q^N$ (Lemma 5.3) and the relation $J^N_i \subseteq J_i$ (Proposition 5.5) yields the following proposition:

**Proposition 5.7.** In equilibrium, $R_1(Q^N) \geq R_1(Q^\ast)$.

In words, the aggregate permit revenue $R_1(Q)$ at the first bottleneck is overestimated by the naïve model. The property can be understood using intuitions from Section 4. Recall that commuters with higher VOT are concentrated around the CBD in the naïve model (see Fig. 8b). This causes higher permit prices at bottlenecks which are closer to the CBD, in particular at the first one, resulting in a higher commuting costs as well as a greater TNP revenue on the side of the road manager for such bottlenecks. In addition, in Appendix we numerically confirm that the total investment over the network, $R(Q)$, is overestimated by the naïve model; that is, $R(Q^N) > R(Q^\ast)$. It is intuistic since a concentration of high-VOT users nearby the CBD shifts all the isocost curves for the upstream bottlenecks upward.

These results have an important policy implication for long-term road construction decisions. The TNP scheme is known to satisfy a self-financing property (Akamatsu, 2007; Wada and Akamatsu, 2013; Akamatsu and Wada, 2017); that is, the total TNP revenue of the road manager at a bottleneck $i \in I$ coincides with the optimal investment on the capacity of the bottleneck $i$, so long as the cost function for road construction is homogeneous of degree one. If we employ the self-financing property of the TNP scheme to determine the amount of investment on bottleneck capacity expansion, the optimal value of the total investment also coincides with $R(Q)$. Then, our results imply that the naïve model overestimates the total required investment on capacity expansion. Without consideration of the short-term dynamics in the long-term component, one ends up in excessive construction of roads. A naïve approach to separately analyze the short- and long-term problems thus falls short. The result should be regarded as an illustrative example of definite need to synthesize the short-term TDM policy with the long-term evolution of travel demand patterns.

6. **Concluding remarks**

Using a minimalist model, this study theoretically demonstrated the fundamental need to synthesize the short-term TDM policy with the long-term demand dynamics and, conversely, to include the effect of short-term traffic dynamics in the long-term analysis. In doing so, our analysis involved many choices and compromises which call for further follow-up studies. We discuss three important directions below.

First, we assumed not only that the long-term preference of commuters is quasilinear but also that land consumption is fixed to unity. Introducing elastic demand for land would be a natural next step. Mathematically, this corresponds to introducing another entropy-like strictly convex term into the long-term problem $[LT^*-N]$ in addition to the short-term objective $z_1(Q)$. Even though we expect that basic intuitions of Section 5.3.2 and, in particular, Section 5.3.3 are kept unchanged, this is a claim that we should verify.

Second, job-choice in our model is a simplified one. The labor demands are inelastic. As a result, the only determinant of the wage in our model is commuting cost. Introducing profit-maximizing firms along with other appropriate assumptions such as productivity will provide richer implications (cf. the single bottleneck model of Takayama, 2015).

Third, as the short-term problem, we focused solely on a DSO assignment achieved as the equilibrium under an optimal short-term TDM policy. Comparison with the dynamic user equilibrium (DUE) assignment without any tolling (Friesz and Mookherjee, 2006; Ma et al., 2015; Akamatsu et al., 2015) is also a fascinating direction, since analytical solutions are unavailable for DUE assignment even in a corridor network—in contrast to the clear-cut results for the DSO assignment presented in Section 4. However, due to the complicated dynamics of the queuing delays, this would involve much more efforts.

**Acknowledgments**

We are grateful for the constructive comments and criticisms made by the four anonymous reviewers of the 22nd International Symposium on Transportation and Traffic Theory. This research is partially supported by JSPS KAKENHI (17K14735, 15H04053, 14J02901).
Appendix A

A1. Proof of Lemmas 4.1, 4.3, and 4.5

By contradiction. Consider two job indices $j, k \in J$ such that $j < k$ (i.e., $\alpha_j > \alpha_k$ by Assumption 2.2). Assume that either or both of $t_{1,j}^- \leq t_{1,k}^-$ and $t_{1,j}^+ \leq t_{1,k}^+$ holds true, which is the negation of the assertion of the lemma ($T_j \subset T_k$ for all $j < k$). Then, the first-order optimality condition (18) for [ST-Primodal / Dual] facilitates the following evaluation:

$$
\begin{align*}
\frac{q_{1,j}}{q_{1,k}(t_{1,j}^-)} > 0 & \iff \rho_{1,j} = \alpha_j s(t_{1,j}^-) + p_1(t_{1,j}^-) \\
\frac{q_{1,k}}{q_{1,j}(t_{1,j}^-)} = 0 & \iff \rho_{1,k} = \alpha_k s(t_{1,j}^-) + p_1(t_{1,j}^-) \iff \rho_{1,j} - \rho_{1,k} \geq (\alpha_j - \alpha_k) \cdot s(t_{1,j}^-), \\
\frac{q_{1,j}}{q_{1,k}(t_{1,j}^+)} > 0 & \iff \rho_{1,j} = \alpha_j s(t_{1,j}^+) + p_1(t_{1,j}^+) \\
\frac{q_{1,k}}{q_{1,j}(t_{1,j}^+)} = 0 & \iff \rho_{1,k} = \alpha_k s(t_{1,j}^+) + p_1(t_{1,j}^+) \iff \rho_{1,j} - \rho_{1,k} \geq (\alpha_j - \alpha_k) \cdot s(t_{1,j}^+).
\end{align*}
$$

where either or both of (A.1) or (A.2) holds true by assumption. First, assume (A.1). Regarding the last term of (A.1), since $s(t)$ is strictly quasiconvex and minimized at $t_0$ (Assumption 2.2), the relation $t_{1,j}^- > t_{1,k}^- < t_0$ yields

$$
(\alpha_j - \alpha_k) \cdot s(t_{1,j}^-) = (\alpha_j - \alpha_k) \cdot \max\left\{s(t_{1,j}^-), s(t_{1,k}^-)\right\} > (\alpha_j - \alpha_k) \cdot s(t) \quad \forall t \in \hat{T}_{i,j} = \left(t_{1,j}^-, t_{1,j}^+\right)
$$

where we determine $t_{1,j}^- > t_0$ by $s(t_{1,j}^-) = s(t_{1,j}^+)$. From (A.1) and (A.3), we have $\rho_{1,j} - \rho_{1,k} > (\alpha_j - \alpha_k) \cdot s(t) \iff \rho_{1,j} - \rho_{1,k} \cdot s(t) < \rho_{1,j} - \alpha_j s(t)$ for all $t \in \hat{T}_{i,j}$. But since $\rho_{1,j} < \rho_1(t) + \alpha_j s(t)$ for all $t \in T$, it must be that $\rho_{1,j} - \alpha_j s(t) < \rho_1(t) \iff \rho_{1,k} < \rho_{1,k} + \alpha_k s(t)$ for all $t \in \hat{T}_{i,j}$. From the short-term equilibrium condition (18), the above inequality yields that $q_{1,k}(t) = 0$ for all $t \in \hat{T}_{i,j}$. Since we assumed that $t_{1,j}^- < t_{1,k}^-$ by (A.1), we conclude that $t_{1,j}^+ \leq t_{1,k}^-$ should hold true. Via similar discussion, assuming (A.2) yields $t_{1,k}^+ < t_{1,j}^-$, where $t_{1,j}^-$ is determined by $s(t_{1,j}^-) = s(t_{1,j}^+)$. Then, assuming both (A.1) and (A.2) leads to a contradiction: $t_{1,k}^- < t_{1,j}^- < t_{1,j}^+ < t_{1,k}^+$, which cannot hold true by definition of $t_{1,k}^+$ and $t_{1,j}^+$. Therefore, only one of (A.1) and (A.2) should hold true, but this leads to a contradiction, as we discuss in the following.

First, assume (A.1). Then we have $(\alpha_j - \alpha_k) \cdot s(t_{1,j}^-) \leq (\alpha_j - \alpha_k) \cdot s(t_{1,j}^-)$. The second inequality is the negation of (A.2). It then yields $s(t_{1,j}^-) < s(t_{1,j}^+)$, whence we conclude that $p_1(t_{1,j}^-) > p(t_{1,j}^+)$. Without loss of generality, assume that $j = \arg \min_i \{t_{1,j}^+\}$ so that $p_1(t_{1,j}^-) = 0$. It then implies $0 > p(t_{1,j}^-)$, but this violates to nonnegativity of permit prices. We thus conclude that the first assumption is false, showing the assertion: for all $j < k$, it must be that $t_{1,k}^- < t_{1,j}^- < t_{1,j}^+ < t_{1,k}^+$. It then yields $j = \arg \min_i \{t_{1,j}^+\} = \arg \max_i \{t_{1,j}^+\}$, which in turn determine the concrete values of $t_{1,j}^+$. Lemmas 4.3 and 4.5 can be proven by a quite similar logic.

A2. Proof of Lemma 4.2

We show the following equivalent lemma.

**Lemma A2.1.** There is no false bottleneck if and only if $D_i < D_{i+1}$ for all $i \in I$

Since sufficiency is immediate, we show only necessity by recursion over $i \in I$ from $i = 1$. (i) Assume that $i = 1$. We note that it cannot be false bottleneck (i.e., $p_1(t) = 0$ for all $t \in T$) so long as the bottleneck capacity $\mu_{1}$ is finite. (ii) Suppose that a bottleneck i is not a false bottleneck. Then, we can show that $\sum_{m=1}^{\infty} q_m(t) = \mu_{1}$ for all $t \in T_{i-1}$. Similarly, iff $i+1$ is also not a false bottleneck, we can show that $\sum_{m=1}^{\infty} q_m(t) = \mu_{i+1}$ for all $t \in T_{i+1}$. Noting $T_i \subset T_{i+1}$, we then see that $q_1(t) = \mu_{i} - \mu_{i+1}$ for all $t \in T_i$ and $q_1(t) = 0$ otherwise. Integrating over $T_i$, we have $\int_{T_i} T_i = (\mu_{i} - \mu_{i+1})T_i = \mu_i \Rightarrow T_i = A_i/\mu_i$. On the other hand, from Lemma 4.1, we must have $T_i < T_{i+1}$. Requiring this, we conclude that $\{A_i\}$ and $\{\mu_i\}$ should satisfy $\sum_{i=1}^{\infty} A_i/\mu_i = \sum_{m=1}^{\infty} A_m/\mu_{i+1}$. This means that $D_i \leq D_{i+1}$. By (i) and (ii), it is shown that $D_i < D_{i+1}$ must hold true at every $i \in I \setminus \{i\}$. We observe that, if we use the relationship $\rho_i^* = \tilde{s}(T_i)$, the same conclusion is obtained by requiring $\rho_i^* < \rho_{i+1}^*$ for all $i \in I$.  

A3. Proof of Lemma 4.4

We show the following equivalent lemma.

**Lemma A3.1.** There is no false bottleneck if and only if $D_{i,j} < D_{i+1,j}$ for all $i \in I$ and $j \in J_i$.

To eliminate false bottlenecks, one should require $\rho_{i,j}^*(Q_j) < \rho_{i+1,j}^*(Q_{i+1})$ for all $j \in J_{i+1}$. Therefore, we must have
\[ \rho^*_i(Q_j) < \rho^*_{i+1,j}(Q_{i+1}) \iff \sum_{l \leq j} \hat{a}_{i,l} \cdot \tilde{s}(T_{i,l}(Q_j)) < \sum_{l \leq j} \hat{a}_{i+1,l} \cdot \tilde{s}(T_{i+1,l}(Q_{i+1})) \]

\[ = \sum_{l \leq j} \hat{a}_{i,l} \cdot T_l(Q_j) < \sum_{l \leq j} \hat{a}_{i+1,l} \cdot T_{i+1,l}(Q_{i+1}) \]

\[ \iff \frac{1}{\mu_{i+1}} \sum_{l \leq j} \hat{a}_{i,l} \sum_{n \geq j} Q_{l,n} < \frac{1}{\mu_i} \sum_{l \leq j} \hat{a}_{i+1,l} \sum_{n \geq j} Q_{l,n} \]


for all \( j \in J_{i+1} \), where the second equivalence follows from the fact that \( \tilde{s}(T) \) is monotone and increasing w.r.t. \( T \). Define \( A_{k,j} \equiv \sum_{l \leq j} \hat{a}_{i,k} \sum_{n \geq j} Q_{l,n} \). Then, it can be seen that one should require \( A_{k,j}/\mu_i < A_{k,j}/\mu_{i+1} \) for all \( i \in I \) and \( j \in J_i \), which is in turn equivalent to \( \sum_{k \geq i} A_{k,j}/\mu_i < \sum_{k \geq i+1} A_{k,j}/\mu_{i+1} \). Because \( D_{i,j} = \sum_{k \geq i} A_{k,j}/\mu_i \), this shows the assertion.

A4. Proof of Lemma 5.1

(a) Immediate from the analytical formula (44). Note that \( \rho^*_i(Q) \) is a \( I \)-dimensional vector-valued function.

(b1) First, note that the length of arrival time windows \( \{ T_{i,j} \} \) in Lemma 4.5 is expressed in vector-matrix form, as \( T_i(Q) = \{ T_{i,j}(Q) \} = (1/\mu_i) \cdot LQ_i \), where \( L \) is a \( I \)-dimensional lower triangular matrix whose upper off-diagonals are zero and other elements are one (i.e., \( L_{i,k} = 1 \)). Note also that the analytical formula (44) can be written as \( \rho^*_i(Q) = L' \tilde{s}(T_i(Q)) \) in vector-matrix form, where we define \( \tilde{s}(T) = [\tilde{s}(T_j)] \). We thus have, for any two nonnegative vectors \( Q_i \) and \( Q_j \), by defining \( T_i = (1/\mu_i) \cdot LQ_i \) and \( T'_j = (1/\mu_j) \cdot LQ'_j \),

\[ (\rho^*_i(Q_i) - \rho^*_i(Q_j))' (Q_j - Q_i)' = (L' (\tilde{s}(T_i) - \tilde{s}(T'_i)))' (Q_j - Q_i)' \]

\[ = (\tilde{s}(T_i) - \tilde{s}(T'_i))' L (Q_i - Q_j) = (\tilde{s}(T_i) - \tilde{s}(T'_i))' (T_i - T'_j) \]

\[ = \sum_{j \in J} \alpha_i (T_{i,j} - T'_{i,j}) (\tilde{s}(T_{i,j}) - \tilde{s}(T'_{i,j})) > 0. \]

It shows the strict monotonicity of \( \rho_i(Q) \). Note that the last inequality follows from the fact that \( \tilde{s}(T) \) is a strictly monotone function of \( T \) thanks to strict quasiconvexity of \( s(t) \).

(b2) If \( s(t) \) is differentiable, \( \rho^*_i(Q) \) is also differentiable. Recall that the analytical formula (44) can be written as \( \rho^*_i(Q) = L' \tilde{s}((1/\mu_i) \cdot LQ_i) \) (Lemma 4.5). Then, we see that \( \nabla \rho^*_i(Q) = (1/\mu_i) \cdot L'D_i \), where \( D_i \) being a \( I \)-dimensional square diagonal matrix whose \( j \)-th entry is given by \( D_{i,j}(Q) = \alpha_j \tilde{s}'(T_{i,j}(Q)) > 0 \). We thus conclude that \( \nabla \rho^*_i(Q) \) is not only symmetric but also admits Cholesky decomposition, whence we conclude that \( \nabla \rho^*_i(Q) \) is positive definite.

(c1) This property is assured by assuming a reduced corridor network.

(c2) Immediate from the analytical formula (44).

A5. Proof of Lemma 5.2

Define \( f_{i,j} \) by \( \int_{\text{supp}([q_{i,j}(t)])} f = \int_{T_{i,j} \cap T_{i,j-1}} \). Then using Proposition 4.3, we have the following computation:

\[ \tilde{z}_1(Q) = \sum_{i \in I} \sum_{j \in J} \rho_{i,j} Q_{i,j} - \sum_{i \in I} \mu_i \int_T p_i(t) \, dt \]

\[ = \sum_{i \in I} \sum_{j \in J} \rho_{i,j} \int_{T_{i,j}} \hat{\mu}_i \, dt - \sum_{i \in I} \mu_i \sum_{j \in J} \int_{T_{i,j}} (\sum_{m \leq i} p_m(t) - \sum_{m \leq i-1} p_m(t)) \, dt \]

\[ = \sum_{i \in I} \sum_{j \in J} \hat{\mu}_i \int_{T_{i,j}} p_{i,j} \, dt - \sum_{i \in I} \mu_i \sum_{j \in J} \int_{T_{i,j}} \sum_{m \leq i} p_m(t) \, dt = \sum_{i \in I} \sum_{j \in J} \hat{\mu}_i \sum_{j \in J} \int_{T_{i,j}} \left( \rho_{i,j} - \sum_{m \leq i} p_m(t) \right) \, dt \]

\[ = \sum_{i \in I} \sum_{j \in J} \sum_{j \in J} \alpha_j s(t) \, dt = \sum_{i \in I} \mu_i \int_{T_{i,j}} \sum_{j \in J} \alpha_j \left( \int_{T_{i,j}} Q_{i,j} \right. \, dt + \int_{T_{i,j}} s(t) \, dt) = \sum_{i \in I} \mu_i \tilde{s}_{i,j}(Q_j) \quad \text{[A.10]} \]

One can easily verify from (A.10) that \( \nabla \tilde{z}_1(Q) = \rho^*(Q) \). It means that \( \tilde{z}_1(Q) \) is a scalar potential for \( \rho(Q) \), i.e., \( \tilde{z}_1(Q) = \int \rho(\omega) d\omega \), which is the definition of \( \tilde{z}_1(Q) \).
A6. Proof of Propositions 5.3 and 5.4

The first-order finite differences of \( w \) and \( r \) are:

\[
W_{j+1} - W_j = \rho_{i,j+1}(Q) - \rho_{i,j}(Q) - \alpha_j a_i, \quad R_{j+1} - R_j = \rho_{i,j}(Q) - \rho_{i,j+1}(Q) - \alpha_j (d_{i+1} - d_i). \tag{A.11}
\]

By properties (c1) and (c2) in Lemma 5.1 and the fact that \( d_{i+1} > d_i \), we see that \( W_{j+1} - W_j < 0 \) and \( R_{j+1} - R_j < 0 \). The equation (A.11) further yields the second-order finite differences of \( w \): \( (W_{j+1} - W_j) - (W_j - W_{j-1}) = \{\rho_{i,j+1}(Q) - \rho_{i,j}(Q)\} - \{\rho_{i,j}(Q) - \rho_{i,j-1}(Q)\} = \alpha_j s_{i,j-1}(Q) - s_{i,j}(Q) < 0 \), where we note that \( s_{i,j-1}(Q) \) is increasing in \( j \).

A7. Proof of Proposition 5.2

By contradiction. Suppose that \( j_+(i) > j_+(i+1) \). We have \( \rho_{i,j_+(i)} = s(t_{i,j_+(i)}) > s(t_{i,j_+(i+1)}) = \rho_{i,j_+(i+1)} \) which in turn yields \( p_{i,j_+(i)}(t) = 0 \) for all \( t \in T_{i,j_+(i+1)} \setminus T_{i,j_+(i+1)} \). But this cannot be true in a reduced corridor network. Therefore, \( j_+(i) < j_+(i+1) \). One can show that \( j_-(i) < j_-(i+1) \) in a similar manner.

A8. Proof of Proposition 5.5

By recursion. First, for location \( i = 1 \), it is obvious that \( j_N^N(1) = j_-(1) = 1 \) and \( j_N^N(1) \leq j_+(1) \), which satisfies the assertion. If \( j_N^N(1) = j_+(1) \), then by Proposition 5.2 we know that \( j_N^N(2) = j_-(2) \). For location 2, the conservation law of K2 provides that \( j_N^N(2) \leq j_N^N(2) \). If \( j_N^N(2) = j_+(2) \) holds, we go to location 3. Repeat this until we find a location \( i \) with \( j_N^N(i) < j_+(i) \). Then there exists a \( j \in j_N^N(i) \) such that the following relation is satisfied: \( \sum_{m \in i} Q_m < \sum_{m \in i} Q_m^N = L_j \). By Proposition 5.2, we know that \( j_+(i) \leq j_+(i+1) \). Also, the solution of [LT–N] yields \( j_N^N(i) + 1 \geq j_N^N(i) \). These two inequalities yield \( j_N^N(i) + 1 \leq j_+(i+1) \), which satisfies the proposition. The conservation law of commuters requires that \( j_N^N(i) + 1 \leq j_+(i+1) \), otherwise \( \sum_{m \in i} K_m = \sum_{m \in i} \sum_{n \in j_N^N(i+1)} Q_{m,n}^N > \sum_{m \in i} \sum_{n \in j_N^N(i+1)} Q_{m,n}^N \). Then check if \( j_N^N(i+1) = j_+(i+1) \). Recursively, we conclude that \( j_N^N(i) \leq j_+(i) \) for any \( i \in \mathcal{I} \). This proves the first part of the assertion. Also, one can prove \( j_N^N(i) \leq j_+(i) \) by the same strategy, starting by considering for \( i = 1 \) instead of \( i = 1 \).

A9. Proof of Proposition 5.6

The optimality conditions of [LT’] and [LT–N] yield \( w_j - w_j^N = \rho_{i,j}(Q) \) for all \( i \in \mathcal{I} \cap \mathcal{I}^N \) and \( r_i - r_i^N = \rho_{i,j}(Q) \) for all \( j \in \mathcal{J}_i \cap \mathcal{J}_i^N \). From the properties (c1) and (c2) of \( \rho_{i,j}(Q) \) in Lemma 5.1, the assertion follows.

A10. Proof of Proposition 5.7

For simplicity, we focus on the piecewise linear schedule delay: \( s(t) = \max \{ \beta^-(t_0 - t), \beta^- (t - t_0) \} \). In the following, we denote the job-choice pattern at location \( i \) by the \( j \)-dimensional column vector \( Q_i \equiv [Q_i] \). We have

\[
R_i(Q) = \sum_{i \in \mathcal{I}} R_i(Q_i), \quad \text{where} \quad R_i(Q_i) = \frac{1}{2} Q_i^T \rho(Q_i) = \gamma_i Q_i^T M_i Q_i, \quad \tag{A.12}
\]

where \( \gamma_i = (1/\mu_i) \cdot (\beta/2) \) with \( \tilde{\beta} = \beta^- \beta^+/\beta^- + \beta^+ \), and \( M_i \) is a \( j \)-by-\( j \) symmetric matrix defined by

\[
M_i \equiv L_i^T \text{diag}[\tilde{\alpha}] L_i = \begin{bmatrix}
\tilde{\alpha}_1 & \tilde{\alpha}_2 & \cdots & \tilde{\alpha}_j \\
\tilde{\alpha}_2 & \tilde{\alpha}_1 & \cdots & \tilde{\alpha}_j \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{\alpha}_j & \tilde{\alpha}_j & \cdots & \tilde{\alpha}_1
\end{bmatrix} \quad \tag{A.13}
\]

using the \( j \)-dimensional lower triangular matrix \( L_i \equiv [1_{1 \geq j}] \). Note that \( R_i(Q) \) is half the total commuting cost of the commuters at location \( i \). We show that \( R(Q_i^N) \geq R(Q_i^N) \) where \( Q_i^N \) and \( Q_i^N \) are the solutions for the naïve model and the integrated model, respectively, with the equality only for \( i = 1 \) or \( j = 1 \). We note that \( R(Q_i^N) = R(Q_i^N) \) holds true for \( i = 1 \) or \( j = 1 \), because \( Q_i^N = Q_i^N \) in these cases. For the rest of this section, for notational simplicity we denote the solutions for the naïve and integrated model by \( Q_i \equiv [Q_i] \) and \( Q_i \equiv [Q_i] \), respectively.

Let \( Q_i \equiv [Q_i] \) and \( Q_i \equiv [Q_i] \). Because \( M_i \) is symmetric, we compute as follows:

\[
\delta R_i = R_i(Q_i - Q_i) = \gamma_i \cdot (Q_i^T M_i Q_i - Q_i^T M_i Q_i) = \gamma_i \cdot (Q_i - Q_i)^T M_i (Q_i + Q_i) \tag{A.14}
\]

We prove that \( \delta R_i > 0 \) for any \( i \geq 2 \) by showing that \( (Q_i - Q_i)^T M_i \) is a positive vector.

Some observations are in order. We denote the largest job index at location 1 for the naïve and integrated models by \( j_N^i \equiv j_N^i (1) \) and \( j_+^i \equiv j_+^i (1) \), respectively. For any location \( i \in \mathcal{I} \), the job set \( j_N^i \) of the naïve model is contained by that of the integrated model \( j_+^i \). That is, \( j_-(i) \leq j_N^i (i) \leq j_+(i) \) (Proposition 5.2). Noting that \( Q_{i,j} = L_j \) for all \( j < j_N^i \), \( Q_i \) and \( Q_i \) take the following form:

\[
Q_i = [Q_{1,j}, Q_{1,j}, \ldots, Q_{1,j}, Q_{1,j}, 0, 0, 0, \ldots] \quad \tag{A.15}
\]

where \( Q_{i,j} \) is the row vector of \( Q_i \) for location \( i \).
Then, by the conservation law at location \( \delta Q_1 \equiv \sum_{j=1}^{j^*} (\hat{Q}_{1,j} - Q_{1,j}) = \sum_{j=j^*+1}^{J} Q_{1,j} > 0 \). Employing the above observations, we see that if \( k \leq j^* \), every kth element \( \epsilon_k = [(Q_1 - \hat{Q}_1)^T M]_k \) is strictly positive:

\[
\epsilon_k = \sum_{j=1}^{k} (\hat{Q}_{1,j} - Q_{1,j}) \alpha_k + \sum_{j=k+1}^{j^*} (\hat{Q}_{1,j} - Q_{1,j}) \alpha_j = \sum_{j=1}^{k} (L_j - Q_{1,j}) (\alpha_k - \alpha_{j^*}) + \sum_{j=k+1}^{j^*} (L_j - Q_{1,j}) (\alpha_j - \alpha_{j^*}) + \alpha_{j^*} \sum_{j=1}^{j^*} (\hat{Q}_{1,j} - Q_{1,j}) - \sum_{j=j^*+1}^{J} Q_{1,j} \alpha_j
\]

\[
> \alpha_{j^*} \sum_{j=1}^{j^*} (\hat{Q}_{1,j} - Q_{1,j}) - \alpha_{j^*} \sum_{j=j^*+1}^{J} Q_{1,j} = (\alpha_{j^*} - \alpha_{j^*}) \delta Q_1 > 0.
\]

We note that \( \alpha_j \) is decreasing in \( j \) and that \( \delta L_j \geq Q_{1,j} \) for all \( j \in J \). Strict positivity for the other cases of \( k \) can be shown in a similar vein. Since \( \hat{Q}_1 \neq \hat{Q}_1 \) is also a strictly positive vector, we conclude that \( \delta R_1 = R_1(Q_1) - R_1(Q_1^*) > 0 \).

### A3. Numerical example for the total investment over the network

This section numerically confirms that the overestimation result on the total permit revenue from the first bottleneck can be generalized to the total investment over the network.

**Benchmark case.** First, consider the following 3 \( \times \) 3 setting as the benchmark. Schedule delay function is piecewise linear: \( s(t) = \max(\beta^-(t - t_0), \beta^+(t - t_0)) \). We assume \( I = 3 \), \( J = 3 \), \( Q = 3000 \), \( \alpha_1 = 1 \), \( \tilde{\alpha}_j = \alpha_{j+1} - \alpha_j = 1 \), \( \beta^- = 0.8 \), \( \beta^+ = 4.0 \), \( A = [500, 1000, 1500] \), \( L = [500, 1000, 1500] \). Observe that \( \{A_i\} \) and \( \{L_j\} \) are both linearly increasing. Throughout, the bottleneck capacities are given by the following formula:

\[
A_{i+1}/\tilde{\mu}_i = (I - i + 4)/2 \cdot A_i/\tilde{\mu}_i \quad \forall i \in I \setminus \{I\}
\]

with \( \tilde{\mu}_I = 8 \), which guarantees that there are no false bottlenecks in any of the numerical examples in this appendix. The overestimation (in percentage) of total investment in the naïve model comparing to that in the integrated model is

\[
\Delta R = \frac{R(Q_1^*) - R(Q_1^*)}{R(Q_1^*)} \times 100 = 7.94\%.
\]

We thus confirm that the total amount of investment over the network is also overestimated by the naïve model.

To see robustness, below we present the results of our numerical experiments regarding the following four factors: (1) VOT difference \( \{\tilde{\alpha}_j\} \) among jobs; (2) schedule delay penalties \( \beta^- \) and \( \beta^+ \); (3) the numbers of locations and jobs, i.e. \( I \) and \( J \); and (4) patterns of land supply \( \{A_i\} \) and labor demand \( \{L_j\} \). All results confirm that \( \Delta R > 0 \), i.e., the naïve model always leads to an overestimation of the total investment. Although only the results for the representative cases are provided here due to space limitation, there have been no single case in which \( \Delta R \leq 0 \).

**VOT difference.** The first factor that we will inspect is the VOT difference \( \{\tilde{\alpha}_j\} \) among jobs. We consider five settings of VOT difference here: \( \tilde{\alpha}_j \in [1, 2, 4, 8, 16] \) for all \( j \in J \) with \( \alpha_1 = 1 \). The overestimation ratio is constant for these five cases: \( \Delta R = 7.94\% \); that is, the VOT difference has no effect on the total investment so long as it is constant over \( j \). It is because the permit price of each bottleneck is a multiple of VOT difference (refer to (45)). Therefore, the total investment in both the models increase in the same proportion and \( \Delta R \) does not change.

**Schedule delay penalties.** Next, we investigate effects of schedule delay penalties \( \beta^- \) and \( \beta^+ \) on the total investment by numerical examples under the following settings: \( \beta^- \in [0.2, 0.4, 0.8, 1.6, 3.2] \) and \( \beta^+ \in [0.5, 1.0, 2.0, 4.0, 8.0] \). The total investment under each pair of early-arrival penalty \( \alpha \) and late-arrival penalty \( \beta \) are then calculated. Table A1 shows that the overestimation ratio \( \Delta R \) increases along with increasing schedule delay penalties. But this trend turns to be vanishing when penalties are large.

**Numbers of locations and jobs.** Here, we provide examples for different sizes of the problem, i.e., different numbers of locations and jobs. The patterns of \( \{A_i\} \) and \( \{L_j\} \) are linearly increasing as in the benchmark case:

\[
A_i = Q_1/(I + 1) \times I/2 \times i \quad \forall i \in I
\]
Table A2
Overestimation ratio $\Delta R$ (%) of the naive model for different numbers of jobs and locations.

<table>
<thead>
<tr>
<th>$J$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6.78</td>
<td>2.35</td>
<td>4.82</td>
<td>4.61</td>
<td>4.98</td>
</tr>
<tr>
<td>3</td>
<td>1.79</td>
<td>7.94</td>
<td>2.93</td>
<td>3.76</td>
<td>5.76</td>
</tr>
<tr>
<td>4</td>
<td>1.29</td>
<td>3.16</td>
<td>8.02</td>
<td>3.13</td>
<td>3.52</td>
</tr>
<tr>
<td>5</td>
<td>0.94</td>
<td>1.20</td>
<td>4.04</td>
<td>7.89</td>
<td>3.30</td>
</tr>
<tr>
<td>6</td>
<td>0.99</td>
<td>1.42</td>
<td>4.60</td>
<td>3.43</td>
<td>7.75</td>
</tr>
</tbody>
</table>

Table A3
Overestimation ratio $\Delta R$ (%) of the naive model for different labor demands and land supplies.

<table>
<thead>
<tr>
<th>$L$</th>
<th>(i)</th>
<th>(ii)</th>
<th>(iii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>7.63</td>
<td>2.18</td>
<td>1.36</td>
</tr>
<tr>
<td>(ii)</td>
<td>1.53</td>
<td>7.94</td>
<td>1.05</td>
</tr>
<tr>
<td>(iii)</td>
<td>1.79</td>
<td>1.90</td>
<td>5.87</td>
</tr>
</tbody>
</table>

$L_j = Q/((J+1) \times J/2) \times j \quad \forall j \in J$. \hfill (A.22)

The results shown in Table A.2 support our statement about total investment, though there is no clear trend of the overestimation along with increasing scales.

**Land supply and labor demand.** The last factor we concern about is the pattern of land supply and labor demand. In parameter settings, we consider three types of land supply and labor demand distributions: uniform, linearly increasing and exponentially increasing:

- **Uniform:** $A_i = Q/J$ 
- **Linear:** $A_i = Q/((1 + I) \times I/2) \times i$ 
- **Exponential:** $A_i = 4A_i-1$ 

L_{j} = \frac{Q}{(J+1) \times J/2} \times j \quad \forall j \in J. \hfill (A.23)

All the results (Table A.3) support our statement of total investment. They show that the overestimation varies a lot for different patterns of land supply and labor demand. It is implied that the overestimation is larger when land supply and labor demand have similar patterns.

**References**


